Journal of Applied Mechanics

Published Bimonthly by ASME

VOLUME 74 • NUMBER 3 • MAY 2007

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.

TECHNICAL PAPERS

(Contents continued on inside back cover)

Journal of Applied Mechanics Volume 74, Number 3 MAY 2007

TECHNICAL BRIEFS

ERRATUM

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

Thermoelastic Fields of a Functionally Graded Coated Inhomogeneity With Sliding/Perfect Interfaces

Hamed Hatami-Marbini
Department of Mechanical,
Aerospace and Nuclear Engineering,
Rensselaer Polytechnic Institute,
Troy, NY 12180

Hossein M. Shodja[1]
Department of Civil Engineering,
Center of Excellence in Structures and
Earthquake Engineering,
Sharif University of Technology,
P. O. Box 11365-9313,
Tehran, Iran
e-mail: shodja@sharif.edu

The determination of the thermo-mechanical stress field in and around a spherical/cylindrical inhomogeneity surrounded by a functionally graded (FG) coating, which in turn is embedded in an infinite medium, is of interest. The present work, in the frame work of Boussinesq/Papkovich-Neuber displacement potentials method, discovers the potential functions by which not only the relevant boundary value problems (BVPs) in the literature, but also the more complex problem of the coated inhomogeneities with FG coating and sliding interfaces can be treated in a unified manner. The thermo-elastic fields pertinent to the inhomogeneities with multiple homogeneous coatings and various combinations of perfect/sliding interfaces can be computed exactly. Moreover, when the coatings are inhomogeneous, as long as the spatial variation of the thermo-elastic properties of the transition layer is describable by a piecewise continuous function with a finite number of jumps, an accurate solution can be obtained. The influence of interface conditions, stiffness of the core, spatial distributions of thermal expansion coefficient and shear modulus of FG coating, and loading condition on the stress field will be examined.
[DOI: 10.1115/1.2200655]

1 Introduction

In this paper, an inhomogeneity system is referred to an infinite medium containing a sub-domain of different material properties. If the hoop displacement along the interface between the inhomogeneity and its surrounding medium is discontinuous, it is referred to as sliding inhomogeneity. The thermo-elastic fields in the neighborhood of a single spherical/cylindrical inhomogeneity system with different combination of perfect / sliding interfaces are of primary interest in composite materials. In fabrication of such materials, application of coating on reinforcement particles serves to improve a number of desired properties, e.g., mechanical strength; electrical conductivity; reduction of thermal residual stresses, and/or providing a buffer zone to prevent chemical attacks. It is due to their prominent wide engineering usage that increasingly more attentions have been turned to coated reinforced composites during the past two decades. The present work aims to study the thermo-mechanical stress field in the vicinity of a functionally graded coated spherical/cylindrical inhomogeneity system with different combination of perfect/sliding interfaces subjected to both a uniform temperature change and a uniform far-field mechanical loading. The high temperature performance of some alloys can be considerably enhanced by addition of certain reinforcements. The optimum behavior may be achievable when the thermal expansion mismatch between the matrix and the reinforcements is very small. It is a well-known phenomenon that, the large thermal expansion coefficient mismatch at the matrix-fiber boundary gives rise to high tensile residual stresses there; as a result the matrix-fiber interface is a preferred site for crack initiation in the matrix during the cool-down processes. Usage of coating made of an appropriate functionally graded material (FGM) serves to eliminate the problem of mismatch at the matrix-coating and coating-particle interfaces. During fabrication processes of FGMs, their microstructure and specific properties are tailored to meet certain spatial variations. In the context of the present work, the coating is made of FGM, so that its thermo-elastic properties match those of matrix and particle at the corresponding interfaces and vary smoothly in between. In some composites, such as concrete, a transition zone between the cement paste and aggregates may be created by chemical processes.

In the literature, most of the previous studies on the subject are concerned with the effective thermo-elastic behavior of composites rather than seeking the actual distribution of the field quantities in and around a coated inhomogeneity. A fairly thorough literature on the subject up to 1996 is available in the review papers by Mura [1] and Mura et al. [2]. Therefore, only the studies which are closely related to the present developments are noted herein. Some of the treatments pertinent to the cylindrical and spherical inhomogeneities with homogeneous transition zone follow the work of Christensen and Lo [3], whose solution is the extension of the work of Love [4] on the spherical inhomogeneity with the perfect inhomogeneity-matrix interface. The contributions which follow this school of thought consider homogeneous coating in the absence of sliding interfaces; see, for example, [5,6]. Herve and Zaoui [5] proposed an $(n+1)$-phase model (n-phase inclusion embedded in an infinite matrix) with perfect interfaces to estimate the effective bulk and shear moduli of an isotropic composite material. It should also be noted that some of the existing theories, in addition to the limitations of homogeneous coating and perfect interfaces, are valid for thin coatings only; see, for example, Benveniste et al. [6]. Benveniste et al. approximated the local fields in a coated cylindrical inhomogeneity system under the thermo-elastic loading and gave the effective thermo-mechanical properties of composite cylinder assemblage. On the other hand, the shortcomings of the existing treatments associated to the cases where the transition zone is inhomogeneous are: (1) The sliding interfaces are ruled out; (2) the composition of the inhomogeneous coating is limited to a very specific distribution; and (3) the

[1]Author to whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received June 2, 2005; final manuscript received February 1, 2006. Review conducted by S. Govindjee. 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.

solutions are commonly in the form of infinite series as opposed to exact closed-form. Using the Frobenius series solution, Lutz and Ferrari [7] and Lutz and Zimmerman [8] presented the solution to the problem of an infinite body containing a spherical particle with an inter-phase whose material properties vary smoothly with radius outside the particle. Furthermore, Zimmerman and Lutz [9] conducted a similar analysis for a cylindrical inhomogeneity with circular cross section subjected to a uniform heating. Mikata [10] modeled continuous fiber composites with a linearly varying inter-phase under axisymmetric thermo-mechanical loadings by four concentric long circular cylinders. It should be emphasized that the above referenced studies [7–10] address a special continuous variation of the elastic moduli within the interfacial zone so that an infinite series solution is possible. Moreover, these solutions have focused on the treatment of perfect interfaces only. Jayaraman and Reifsnider [11] used the three phase concentric cylinder assemblage model to study the effect of inter-phase Young's modulus variations on the local stress state. They considered perfect bonding between the interfaces and assumed that the Poisson's ratio and the coefficient of thermal expansion are constant within the interfacial zone. The extension of the above study to investigate the effect of an inhomogeneous interphase on the elastic constants of unidirectional fiber reinforced composites was done by Jasiuk and Kouider [12].

Following the work of Sadowsky and Sternberg [13], Mikata and Taya [14,15] determined the mechanical/thermal stress field in a coated short fiber embedded in an infinite body. They modeled the coated fiber by two confocal prolate spheroids, and used Boussinesq-Sadowsky potential functions to end up with an infinite series solution for the corresponding problem. Mura et al. [16] applied the same approach to find the stress field of a sliding prolate spheroidal inclusion with nonshear eigenstrain. In that paper, they also considered the case of a prolate spheroidal inhomogeneity under uniform far-field tension. In an earlier study by Ghahremani [17] who aimed to study the effect of grain boundary sliding on anelasticity of polycrystalline materials, he has obtained the stress fields of a sliding spherical inhomogeneity under uniform tension at infinity. In this, Ghahremani extended the work of Lur'e [18] on the spherical inclusion with a perfect inclusion-matrix interface, and did not consider coating.

Using the axisymmetric elasticity solution of spherical regions given by Lur'e [18], Hashin [19] studied the effect of imperfect interface conditions on the stress fields in and around an inclusion. Hashin [20] has pointed out that the imperfect interface bond is often used to model a very compliant thin elastic homogeneous transition layer. On this basis, Hashin [20] estimated the effective elastic moduli of a unidirectional coated fiber composite. In seeking the Eshelby's tensor associated with an inclusion having imperfect interface, the work of Gao [21] should be mentioned. Gao considered the simple two-dimensional circular inclusion for which the Airy's stress functions are readily available. In this, most of the effort has been expended in calculation of the pertinent Eshelby's tensor.

The present paper, does not seek the infinite series solution, but rather in the frame work of Boussinesq/Papkovich-Neuber displacement potentials method offers a finite number of thermo-mechanical potential functions appropriate for the exact closed-form solution of the boundary value problem (BVP) of an n-phase cylindrical/spherical inhomogeneity system. An interesting application of the findings is to determine the thermo-mechanical stress field in a cylindrical/spherical inhomogeneity with an FG coating of arbitrary thickness, in which any of the desired interfaces may experience sliding. The influence of matrix-coating/coating-fiber interface condition, stiffness of the core inhomogeneity; thermal/mechanical loading; and spatial variation of the properties of FGM on the stress distribution will be given. The problem of functionally graded pressure vessels considered by Tutuncu and Ozturk [22], which is a simple case encompassed by the present theory is re-examined, and on the present groundwork the limitations posed by their approach are relaxed. Also, the error encountered in the paper of Tutuncu and Ozturk [22] for the expression of hoop stress is pointed out in this work, and its correct form is given in Appendix C.

2 Theoretical Background

A brief formulation for the well-posed boundary value problem, seeking quasi-static elastic fields due to instantaneous temperature distribution $T(\mathbf{x})$ is given in the followings. In the presence of body forces ρb_i, the displacement equations of equilibrium for isotropic materials read

$$(\lambda+\mu)u_{k,ki}+\mu u_{i,kk}=-\rho b_i+\alpha(3\lambda+2\mu)T_{,i} \tag{1}$$

where $u_i's$, $i=1,2,3$ are the components of displacement filed, λ and μ are Lame' constants, and α is the coefficient of thermal expansion (CTE). Assuming constant thermoelastic properties, one may make use of the Galerkin vector, $G_i(\mathbf{x})$

$$2\mu u_i=2(1-\nu)G_{i,kk}-G_{k,ki}+2\mu u_i^T \tag{2}$$

where ν is the Poisson's ratio and u_i^T is the displacement field due to a uniform temperature change in the medium.

Upon substitution into (1) and introducing the functions $\psi_i(\mathbf{x})$ and $\varphi(\mathbf{x})$, (Papkovich (1932), Neuber (1934)), one obtains

$$2\mu u_i=(3-4\nu)\psi_i-x_k\psi_{k,i}-\phi_{,i}+2\mu u_i^T \tag{3}$$

where

$$\psi_i=\tfrac{1}{2}\nabla^2 G_i \tag{4}$$

$$\phi=G_{k,k}-x_k\psi_k \tag{5}$$

$$G_{i,lljj}=\frac{-\rho b_i+\alpha(3\lambda+2\mu)T_{,i}}{1-\nu} \tag{6}$$

It is observed that, for a uniform temperature change and in the absence of body forces, ψ_i and ϕ are harmonic functions. These conditions in conjunction with the Boussinesq's (1885) choice of functions

$$\phi=\Phi,\quad \psi_1=0,\quad \psi_2=0,\quad \psi_3=-\Psi \tag{7}$$

Eq. (3) becomes

$$2\mu u_i=\Phi_{,i}+x_3\Psi_{,i}-\delta_{3i}(3-4\nu)\Psi+2\mu u_i^T \tag{8}$$

As long as the distributions of the thermo-elastic properties of the material within the interfacial zone are describable by piece-wise continuous functions with a finite number of jumps, the present theory is applicable. For the sake of demonstration, and conveying the fundamental concepts, two different variations of thermo-elastic properties of the transition layer, namely the exponential and power law distributions, which are of great value in applications are considered. The effects of these functions on the elastic field are explored in the results and discussion section. Let $\gamma^c(r)$ denote the shear modulus, $\mu^c(r)$ or the coefficient of thermal expansion, $\alpha^c(r)$ within the transition zone; it is assumed that $\gamma^c(r)$ obeys either the exponential form

$$\gamma^c(r)=\gamma^f+\frac{r-r_f}{r_m-r_f}(\gamma^m-\gamma^f)e^{r-r_m},\quad r_f\le r\le r_m \tag{9}$$

or the power law

$$\gamma^c(r)=\gamma^m+(\gamma^f-\gamma^m)\left(\frac{r-r_m}{r_f-r_m}\right)^\eta,\quad r_f\le r\le r_m \tag{10}$$

where the superscripts f and m over γ mean that the quantity is pertinent to the fiber (core inhomogeneity) and matrix, respectively. The distance r measures from the center of the inhomogeneity. It is seen that at the inhomogeneity-coating interface $\gamma^c(r_f)=\gamma^f$ and at the coating-matrix interface, $\gamma^c(r_m)=\gamma^m$. In relation (10), the power η may take on any real numbers.

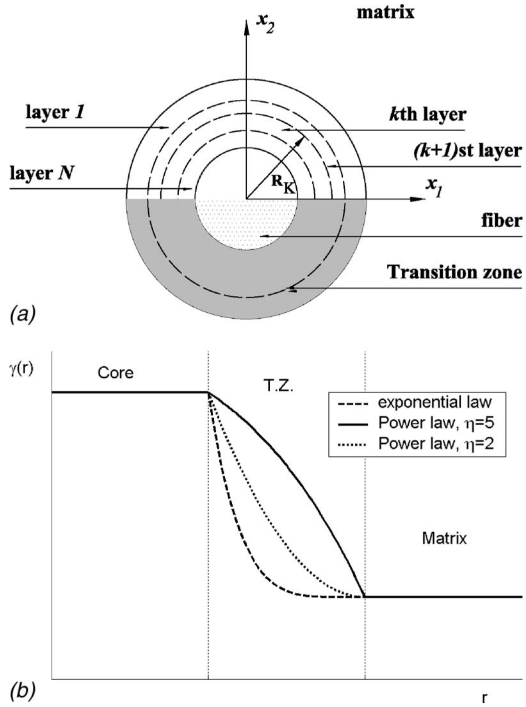


Fig. 1 (*a*) Topology of an FG coated cylindrical inhomogeneity system (*b*) A typical variation of thermal/mechanical properties of an FG coated inhomogeneity system

The exact solution of the governing equation (1) with variable coefficients stemming from expression (9) or (10) is unknown. It is shown in this paper that, under uniform thermo-mechanical loading the thermoelastic fields disturbed by FG coated cylindrical/spherical inhomogeneities with sliding interfaces can be solved accurately.

3 Formulation for a Two-Dimensional FG Coated Cylindrical Inhomogeneity

Suppose the FG coated cylindrical inhomogeneity system shown in Fig. 1 undergoes both a uniform change in temperature ΔT and a uniform tension, ΔT in x_1-direction. Having the uniaxial loading solution, the elastic field of the problem under a uniform hydrostatic or shear loading can be obtained from the superposition principle. Depending on the matrix-coating and coating-fiber interface conditions, four different configurations are possible: (1) Perfect bonding between matrix-coating and coating-fiber interfaces; (2) perfect bonding between the matrix-coating interface and free sliding between the coating-fiber interface; (3) free sliding between the matrix-coating interface and perfectly bonded coating-fiber interface; and (4) free sliding for both interfaces. For abbreviation, configurations (1)–(4) are referred to as PP, PS, SP, and SS in the subsequences. The approach to the determination of the elastic fields disturbed by such an inhomogeneity is to subdivide the coating into N thin enough layers. The continuity conditions at the Kth and $(K+1)$st interface, $r=R_k$ require that

$$\sigma_{rr}^{(K)}(R_k)=\sigma_{rr}^{(K+1)}(R_k),\quad \sigma_{r\theta}^{(K)}(R_k)=\sigma_{r\theta}^{(K+1)}(R_k)$$
$$u_r^{(K)}(R_k)=u_r^{(K+1)}(R_k),\quad u_\theta^{(K)}(R_k)=u_\theta^{(K+1)}(R_k)\qquad(11)$$

Moreover, within the Kth phase the displacement field is given in terms of the displacement potentials via

$$2\bar{\mu}_K u_m^{(K)}=(3-4\nu)\psi_m^{(K)}-x_l\psi_{l,m}^{(K)}-\phi_{,m}^{(K)}$$
$$+2\bar{\mu}_K\bar{\alpha}_K\Delta T x_m,\ \text{no sum on } K,$$
$$l,m=1,2,\ \text{and } K=1,2,\ldots,N\qquad(12)$$

where the superscript (K) indicates the Kth phase. Without loss of generality, the Poisson's ratio is assumed to be constant not only for the Kth phase but throughout the entire domain. In relation (12), $\bar{\mu}_K$ and $\bar{\alpha}_K$ are the respective values of the shear modulus and coefficient of thermal expansion at the midpoint of the Kth layer. It should be emphasized that, aside from this estimation the remainder of treatment is exact. As a result thinner layers lead to more accurate solution.

The layers are numbered in increasing order from 1 to N, phases 1 and N being the layers next to the matrix and fiber, respectively. The elastic field is obtained from the superposition of the elastic fields associated with disturbed and undisturbed fields. The functions given in (12) correspond to the disturbance caused by the presence of the inhomogeneity. The appropriate such potentials for the given boundary value problem are

$$\phi^{(m)}=a_m\log r+c_m\frac{1}{r^2}\cos 2\theta$$

$$\psi_1^{(m)}=f_m r\cos\theta,\quad \psi_2^{(m)}=f_m r\sin\theta$$

$$\left.\begin{array}{l}\phi^{(K)}=a_K\log r+c_K\dfrac{1}{r^2}\cos 2\theta\\ \psi_1^{(K)}=f_K r\cos\theta+g_K r^3\cos 3\theta,\\ \psi_2^{(K)}=f_K r\sin\theta+g_K r^3\sin 3\theta\end{array}\right\},\quad K=1,2,\ldots,N,\qquad(13a)$$

$$\psi_1^{(f)}=g_f r^3\cos 3\theta,\quad \psi_2^{(f)}=g_f r^3\sin 3\theta,$$

and the potentials for the undisturbed field are

$$V^{(m)}=d_m\frac{(1-\kappa)}{2}r^2+e_m r^2\cos 2\theta$$

$$V^{(K)}=d_K\frac{(1-\kappa)}{2}r^2+e_K r^2\cos 2\theta,\quad K=1,2,\ldots,N$$

$$V^{(f)}=d_f\frac{(1-\kappa)}{2}r^2+e_f r^2\cos 2\theta\qquad(13b)$$

where m and f in (13) denote matrix and fiber, respectively, and $\kappa=3-4\nu$ for plane strain, and $\kappa=(3-\nu)/(1+\nu)$ for plane stress. Hence,

$$2\mu_m u_r^{(m)}=-\frac{a_m}{r}+\frac{2c_m}{r^3}\cos 2\theta+d_m(\kappa-1)r-2e_m r\cos 2\theta$$
$$+f_m\frac{1+\kappa}{r}\cos 2\theta+2\mu_m\alpha_m r\Delta T$$

$$2\mu_m u_\theta^{(m)}=\frac{2c_m}{r^3}\sin 2\theta+2e_m r\sin 2\theta+f_m\frac{1-\kappa}{r}\sin 2\theta$$

$$\sigma_{rr}^{(m)}=\frac{a_m}{r^2}-\frac{6c_m}{r^4}\cos 2\theta+2d_m-2e_m\cos 2\theta-\frac{4f_m}{r^2}\cos 2\theta$$

$$\sigma_{r\theta}^{(m)}=-\frac{6c_m}{r^4}\sin 2\theta+2e_m\sin 2\theta-\frac{2f_m}{r^2}\sin 2\theta$$

$$2\bar{\mu}_K u_r^{(K)}=-\frac{a_K}{r}+\frac{2c_K}{r^3}\cos 2\theta+d_K(\kappa-1)r-2e_K r\cos 2\theta$$
$$+f_K\frac{1+\kappa}{r}\cos 2\theta+g_K(\kappa-3)r^3\cos 2\theta+2\bar{\mu}_K\bar{\alpha}_K r\Delta T$$

$$2\bar{\mu}_K u_\theta^{(K)}=\frac{2c_K}{r^3}\sin 2\theta+2e_K r\sin 2\theta+f_K\frac{1-\kappa}{r}\sin 2\theta$$
$$+g_K(\kappa+3)r^3\sin 2\theta$$

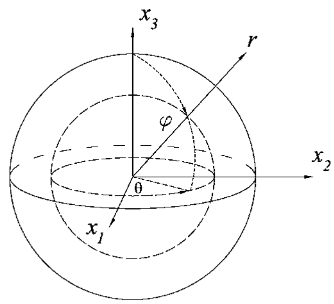


Fig. 2 Topology of a spherical inhomogeneity system

$$\sigma_{rr}^{(K)} = \frac{a_K}{r^2} - \frac{6c_K}{r^4}\cos 2\theta + 2d_K - 2e_k \cos 2\theta - \frac{4f_K}{r^2}\cos 2\theta$$

$$\sigma_{r\theta}^{(K)} = -\frac{6c_K}{r^4}\sin 2\theta + 2e_K \sin 2\theta - \frac{2f_K}{r^2}\sin 2\theta + 6g_K r^2 \sin 2\theta$$

$$2\mu_f u_r^{(f)} = d_f(\kappa-1)r - 2e_f r\cos 2\theta + g_f(\kappa-3)r^3\cos 2\theta + 2\mu_f \alpha_f r\Delta T$$

$$2\mu_f u_\theta^{(f)} = 2e_f r\sin 2\theta + g_f(\kappa+3)r^3 \sin 2\theta$$

$$\sigma_{rr}^{(f)} = 2d_f - 2e_f\cos 2\theta$$

$$\sigma_{r\theta}^{(f)} = 2e_f \sin 2\theta + 6g_f r^2 \sin 2\theta \qquad (14)$$

The boundary conditions at matrix-coating and coating-fiber interface can be written as

$$\sigma_{rr}^{(m)}(a) = \sigma_{rr}^{(1)}(a), \quad u_r^{(m)}(a) = u_r^{(1)}(a)$$

$$\sigma_{rr}^{(N)}(b) = \sigma_{rr}^{(f)}(b), \quad u_r^{(N)}(b) = u_r^{(f)}(b) \qquad (15)$$

and

$$\sigma_{r\theta}^{(m)}(a) = \sigma_{r\theta}^{(1)}(a), \quad u_\theta^{(m)}(a) = u_\theta^{(1)}(a)$$

$$\sigma_{r\theta}^{(N)}(b) = \sigma_{r\theta}^{(f)}(b), \quad u_\theta^{(N)}(b) = u_\theta^{(f)}(b) \qquad (16a)$$

if all interfaces remain perfect,

$$\sigma_{r\theta}^{(m)}(a) = \sigma_{r\theta}^{(1)}(a), \quad u_\theta^{(m)}(a) = u_\theta^{(1)}(a)$$

$$\sigma_{r\theta}^{(N)}(b) = \sigma_{r\theta}^{(f)}(b) = 0 \qquad (16b)$$

if coating-fiber interface slips,

$$\sigma_{r\theta}^{(m)}(a) = \sigma_{r\theta}^{(1)}(a) = 0$$

$$\sigma_{r\theta}^{(N)}(b) = \sigma_{r\theta}^{(f)}(b), \quad u_\theta^{(N)}(b) = u_\theta^{(f)}(b) \qquad (16c)$$

if matrix-coating interface slips,

$$\sigma_{r\theta}^{(m)}(a) = \sigma_{r\theta}^{(1)}(a) = 0$$

$$\sigma_{r\theta}^{(N)}(b) = \sigma_{r\theta}^{(f)}(b) = 0 \qquad (16d)$$

if both interfaces slip. As it is outlined in Appendix A, following a systematic procedure, the unknown constants for each phase can be obtained by imposing the appropriate constraints, interface and boundary conditions.

4 Formulation for a Three-Dimensional FG Coated Spherical Inhomogeneity

The geometry of a FG coated spherical inhomogeneity system is shown in Fig. 2. The system is subjected to uniform temperature change, ΔT and far-field uniform tension, T in x_3-direction. Following the methodology explained in the previous section, the coating is divided into N thin layers. The displacement field for a typical layer, say the Kth phase may be written as

$$2\bar{\mu}_K u_i^{(K)} = \Phi_{,i}^{(K)} + x_3\Psi_{,i}^{(K)} - \delta_{3i}(3-4\nu)\Psi^{(K)} + 2\bar{\mu}_K\bar{\alpha}_K \Delta T x_m, \text{ no sum on } K,$$
$$i = 1,2,3, \text{ and } K = 1,2,\ldots,N \qquad (17)$$

where all the ingredients have been defined in the previous sections. In lieu of the above mentioned applied loadings and continuity conditions (11) the legitimate displacement potentials are

$$\Phi^{(m)} = a_m\frac{1}{r} + b_m\frac{1}{r^3}p_2(s) + d_m r^2 p_2(s)$$

$$\Psi^{(m)} = c_m\frac{1}{r^2}p_1(s) + e_m r p_1(s)$$

$$\Phi^{(K)} = a_K\frac{1}{r} + b_K\frac{1}{r^3}p_2(s) + d_K r^2 p_2(s) + h_K r^4 p_4(s)$$

$$\Psi^{(K)} = c_K\frac{1}{r^2}p_1(s) + e_K r p_1(s) + f_K r^3 p_3(s) \qquad (18)$$

$$\Phi^{(f)} = d_f r^2 p_2(s) + h_f r^4 p_4(s)$$

$$\Psi^{(f)} = e_f r p_1(s) + f_f r^3 p_3(s)$$

where $s=\cos\theta$ and $p_n(s)$ is the Legendre function. Also $r^2p_2(s)$ and $rp_1(s)$ are the potentials for the undisturbed field. From Eq. (18)

$$\sigma_{rr}^{(m)} = a_m\frac{2}{r^3} + b_m\frac{12}{r^5}P_2(s) + c_m\frac{2}{r^3}\big(s(5-2\nu)P_1(s) - \nu(1-s^2)P_1'(s)\big) + d_m(2P_2(s)) - e_m\big(2(1-\nu)sP_1(s) + 2\nu(1-s^2)P_1'(s)\big)$$

$$\sigma_{r\theta}^{(m)} = \left[b_m\left(\frac{4}{r^5}P_2'(s)\right) + c_m\frac{2}{r^3}\big(s(2-\nu)P_1'(s) - (1-2\nu)P_1(s)\big) - d_m P_2'(s) + e_m(1-2\nu)\big(sP_1'(s) + P_1(s)\big)\right]\sqrt{1-s^2}$$

$$2\mu_m u_r^{(m)} = a_m\frac{-1}{r^2} - b_m\frac{3}{r^4}P_2(s) + c_m\frac{s(4\nu-5)}{r^2}P_1(s) + + 2d_m r P_2(s) - 2e_m rs(1-2\nu)P_1(s) + 2\mu_m\alpha_m r\Delta T$$

$$2\mu_m u_\theta^{(m)} = \left[b_m\frac{-1}{r^4}P_2'(s) - c_m\frac{1}{r^2}\big(sP_1'(s) - (3-4\nu)P_1(s)\big) - d_m r P_2'(s) - e_m r\big(sP_1'(s) - (3-4\nu)P_1(s)\big)\right]\sqrt{1-s^2}$$

$$\sigma_{rr}^{(K)} = a_K\frac{2}{r^3} + b_K\frac{12}{r^5}P_2(s) + c_K\frac{2}{r^3}\big(s(5-2\nu)P_1(s) - \nu(1-s^2)P_1'(s)\big) + 2d_K P_2(s) - 2e_K\big((1-\nu)sP_1(s) + \nu(1-s^2)P_1'(s)\big) + 12f_K r^2 P_4(s) + 2h_K r^2\big(3\nu sP_3(s) - \nu(1-s^2)P_3'(s)\big)$$

$$\sigma_{r\theta}^{(K)} = \left[b_K\left(\frac{4}{r^5}P_2'(s)\right) + c_K\frac{2}{r^3}\big(s(2-\nu)P_1'(s) - (1-2\nu)P_1(s)\big) - d_K P_2'(s) + e_K(1-2\nu)\big(sP_1'(s) + P_1(s)\big) - 3f_K r^2 P_4'(s) - h_K r^2\big((1+2\nu)sP_3'(s) - 3(1-2\nu)P_3(s)\big)\right]\sqrt{1-s^2}$$

$$
\begin{aligned}
2\bar{\mu}_K u_r^{(K)} &= a_K \frac{-1}{r^2} - b_K \frac{3}{r^4} P_2(s) + c_K \frac{s(-5+4\nu)}{r^2} P_1(s) + 2 d_K r P_2(s) \\
&\quad - 2 e_K r s (1-2\nu) P_1(s) + 4 f_K r^3 P_4(s) + 4 h_K s \nu r^3 P_3(s) \\
&\quad + 2\bar{\mu}_K \bar{\alpha}_K r \Delta T \\
2\bar{\mu}_K u_\theta^{(K)} &= \Bigg[b_K \frac{-1}{r^4} P_2'(s) - c_K \frac{1}{r^2}\big(s P_1'(s) - (3-4\nu) P_1(s)\big) \\
&\quad - d_K r P_2'(s) - f_K r^3 P_4'(s) - e_K r \big(s P_1'(s) - (3-4\nu) P_1(s)\big) \\
&\quad - r^3 h_K \big(s P_3'(s) - (3-4\nu) P_3(s)\big) \Bigg] \sqrt{1-s^2}
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{rr}^{(f)} &= 2 d_f P_2(s) - 2 e_f \big((1-\nu) s P_1(s) + \nu (1-s^2) P_1'(s)\big) + 12 f r^2 P_4(s) \\
&\quad + 2 h_f r^2 \big(3 \nu s P_3(s) - \nu (1-s^2) P_3'(s)\big) \\
\sigma_{r\theta}^{(f)} &= \big[- d_f P_2'(s) + e_f (1-2\nu)\big(s P_1'(s) + P_1(s)\big) - 3 f_f r^2 P_4'(s) \\
&\quad - h_f r^2 \big((1+2\nu) s P_3'(s) - 3(1-2\nu) P_3(s)\big)\big] \sqrt{1-s^2} \\
2\mu_f u_r^{(f)} &= 2 d_f r P_2(s) - 2 e_f r s (1-2\nu) P_1(s) + 4 f_f r^3 P_4(s) \\
&\quad + 4 h_f s \nu r^3 P_3(s) + 2 \mu_f \alpha_f r \Delta T \\
2\mu_f u_\theta^{(f)} &= \big[d_f r P_2'(s) - f_f r^3 P_4'(s) - e_f r \big(s P_1'(s) - (3-4\nu) P_1(s)\big) \\
&\quad - r^3 h_f (s P_3'(s) - (3-4\nu) P_3(s)) \big] \sqrt{1-s^2} \qquad (19)
\end{aligned}
$$

The $6(N+2)$ unknown coefficients could be obtained from the imposition of interface and boundary conditions as it is presented in Appendix B.

5 Numerical Results and Discussion

This section presents several studies to assess the effect of thermo-elastic properties of the phases and interface conditions on the stresses associated with the FG coated cylindrical/spherical inhomogeneity system. The Poisson's ratio is assumed to be equal to 0.3 for all phases, $\mu_f = 10$ GPa, $\mu_m = 1$ GPa, $\alpha_f = 2\times 10^{-5}/{}^\circ\mathrm{C}$, and $\alpha_m = 10^{-5}/{}^\circ\mathrm{C}$; the coating properties obey the exponential form (9), unless stated otherwise. The ratio of coating radius to the core radius is $a/b = 1.50$. A uniform far-field tension of 10 MPa in the x_1-direction, and a uniform temperature change $\Delta T = 100\ {}^\circ\mathrm{C}$ is applied to the system. To obtain an accurate numerical solution within a specified tolerance ε, the thickness of the subdivisions within the FG coating is refined until one-half the percent difference between the solutions obtained using n and $n-2$ subdivisions become smaller than ε. In the pertinent examples $\varepsilon = 10^{-6}$ is employed, and so depending on the relative thickness of the transition zone to the radius of the core particle the number of layers has been varied until the desired accuracy has been obtained; for the worst scenarios encountered, 40 to 50 subdivisions suffice for convergence.

5.1 Examples Involving Circular Cylindrical Inhomogeneity with FG Coating. A coated cylindrical inhomogeneity system under thermo-mechanical loading is considered in this section. The effect of interface conditions, SS and PP on the variation of hoop stress in and around a cylindrical inhomogeneity with FG transition zone is displayed in Fig. 3. It is observed that for perfect fiber-coating and coating-matrix interfaces denoted by pp, the variation of the hoop stress $\sigma_{\theta\theta}$ is slow inside the fiber and rather notable within the FG coating. For SS interface condition, on the other hand, the variation of $\sigma_{\theta\theta}$ within both the core and its surrounding coating is very abrupt. The magnitude of the hoop stress $\sigma_{\theta\theta}$ at the field points $r/b = 0, 1^-, 1^+, 1.5^-$ pertinent to the SS conditions is much higher than that for the PP conditions. It should be noted that for the case PP, $\sigma_{\theta\theta}$ is continuous across the interfaces, whereas it has remarkable jumps for SS interface conditions. Figures 4 and 5 are the respective distributions of the tangential and radial stresses for various combinations of interface types. From Fig. 4, it can be seen that the variation of the shear stress $\sigma_{r\theta}$ associated with the cases involving at least one sliding interface, SS, SP, or PS is rather abrupt within the core and the coating, whereas for perfect interfaces, PP the variation of $\sigma_{r\theta}$ is fairly moderate throughout the entire domain. Note that $\sigma_{r\theta}$ takes on the value of zero at the sliding interface. The maximum absolute value of $\sigma_{r\theta}$ corresponding to SS, PS, and SP occurs at the origin, $r/b = 0$ and is highest for SS conditions. Figure 5 examines the effect of interface condition on the variation of σ_{rr} along the x_1-axis. Along this axis, for each of the cases PP, PS, SP, and SS, σ_{rr} is uniform but different inside the core. The value of σ_{rr} inside the core is the smallest for PP and largest for SS interface conditions. For PS and SP cases σ_{rr} is nearly the same within the fiber. Inside the coating the variation of σ_{rr} is more remarkable for PS; this trend changes inside the matrix. The highest variation of σ_{rr} is attributed to SS interface condition within the coating and matrix regions. From Figure 6, it is observed that for such a system, the thermal loading gives rise to negative radial stress, whereas the applied mechanical loading produces positive σ_{rr}. Figure 7 displays the radial stress distribution for various fiber stiffness and interface conditions. It is evident that for a given interface condition SS or PP, the magnitude of the radial stress inside the fiber is

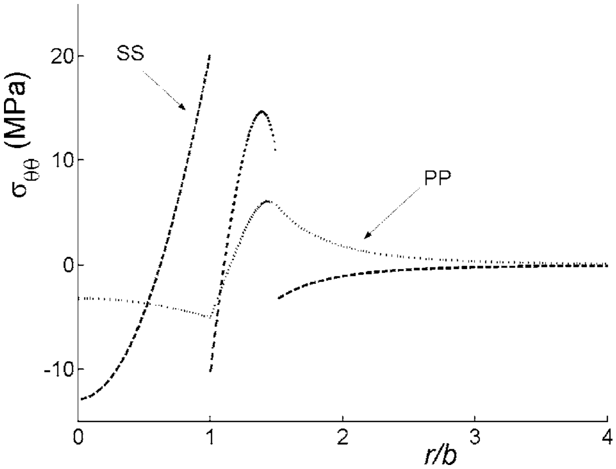


Fig. 3 Variation of the hoop stress with distance from the center of the FG coated cylindrical fiber along the x_1-axis; effect of interface conditions

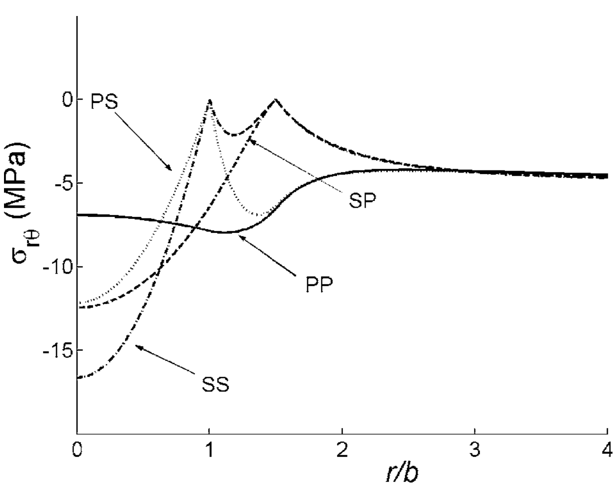


Fig. 4 Variation of the tangential stress with distance from the center of the FG coated cylindrical fiber along the line $\theta = \pi/4$; effect of interface conditions

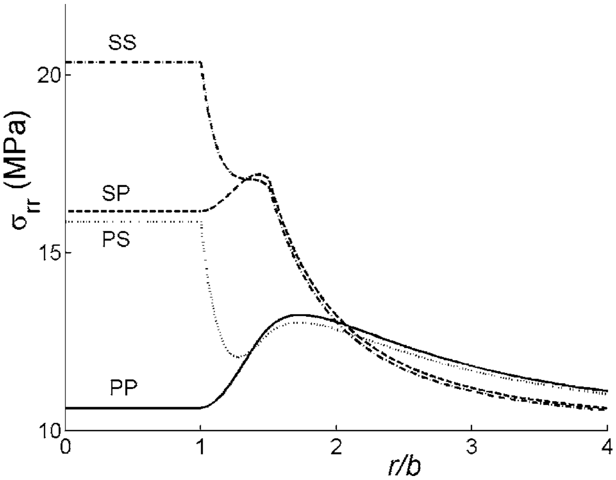


Fig. 5 Variation of the radial stress with distance from the center of the FG coated cylindrical fiber along the x_1-axis; effect of interface conditions

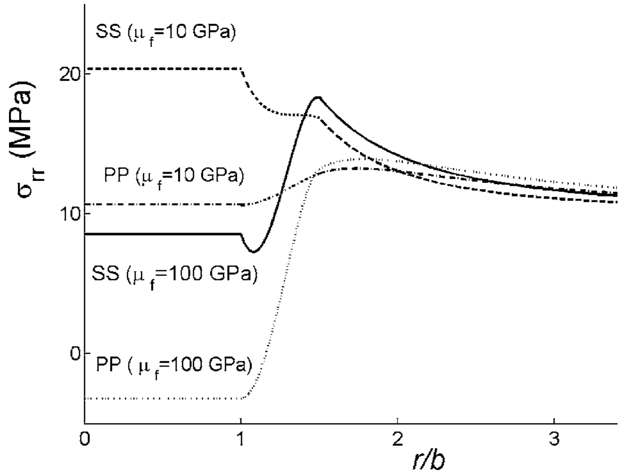


Fig. 7 Variation of the radial stress with distance from the center of the FG coated cylindrical fiber along the x_1-axis; influence of fiber stiffness and interface conditions

lowered with increasing the shear modulus of the fiber. The variation of σ_{rr} within the coating, however, becomes more pronounced as the stiffness of the fiber increases. For a system with SS type interface conditions, the effect of spatial variation of the thermoelastic properties of the transition layer is evident from Fig. 8. In this figure, the power in the power law relation is set equal to 5.

5.2 Examples Involving FG Coated Spherical Inhomogeneity. An interesting field quantity is the amount of slippage along the perimeters of the sliding interfaces. The jumps in the hoop stress across the sliding interfaces of an FG coated spherical inhomogeneity denoted by $[u_\theta]$ are depicted in Fig. 9. It is evident that the jump along the core inhomogeneity—coating interface pertinent to the case SS, $r=b$ is comparable to that of SP, $r=b$. Similar observation is made for the case SS, $r=a$ and the case where the FG coating-matrix is the sliding interface, PS, $r=a$. It is noted that at $\theta=0,90,180°$ the amount of slippage for all the cases shown in Fig. 9 is equal to zero; and the absolute values of the maximum jumps pertinent to the cases SS, $r=a$ and PS, $r=a$ are higher than those corresponding to SS, $r=b$ and SP, $r=b$. The other studies in this section are similar to the studies on FG coated cylindrical fiber discussed in Sec. 5.1, hence the redundant discussions are avoided by noting that Figs. 10–14 of this section are counterparts of Figs. 4–8 of Sec. 5.1, respectively.

5.3 Stress Field of an FG Pressure Vessel. The elastic fields of a cylindrical/spherical functionally graded pressure vessel given by Tuntuncu and Ozturk [22] will be reexamined in this section. For this special case, the boundary conditions (15) and (16) shall be replaced by

$$\sigma_{rr}^{(1)}(a)=0, \quad \sigma_{rr}^{(N)}(b)=-p$$

$$\sigma_{r\theta}^{(1)}(a)=\sigma_{r\theta}^{(N)}(b)=0 \tag{20}$$

where p is the internal pressure. Figures 15 and 16 depict the respective stress fields of FG cylindrical and spherical vessels with $b/a=0.6$, $\nu=0.3$ and elastic modulus of

$$E=E_0 r^{\beta}, \tag{21}$$

in which E_0 is the Young's modulus at the outer surface. The stress field of an isotropic pressure vessel, denoted by superscript H, has been used to normalize the stresses shown in these plots.

Unlike the radial stresses, the hoop stresses obtained by the present method are substantially different from the ones given by Tuntuncu and Ozturk. A close scrutiny of their closed-form hoop

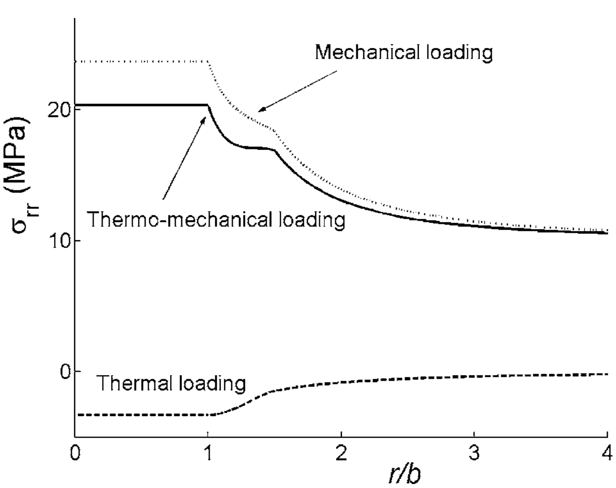


Fig. 6 Variation of the radial stress with distance from the center of the FG coated cylindrical fiber along the x_1-axis; effect of loading conditions

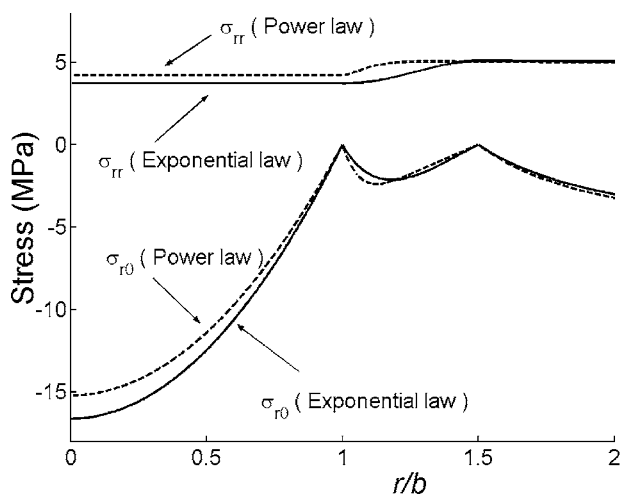


Fig. 8 Variation of σ_{rr} and $\sigma_{r\theta}$ with distance from the center of the FG coated cylindrical fiber along the line $\theta=\pi/4$; influence of spatial distributions of thermal expansion coefficient and shear modulus of FGM coating

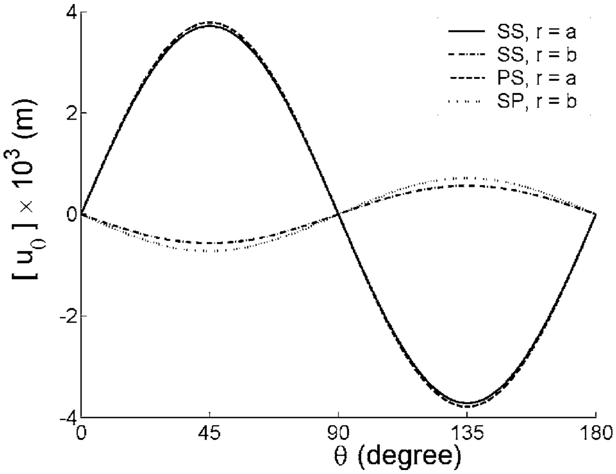


Fig. 9 The jump in hoop displacement along the sliding interfaces of the FG coated spherical particle; effect of interface conditions

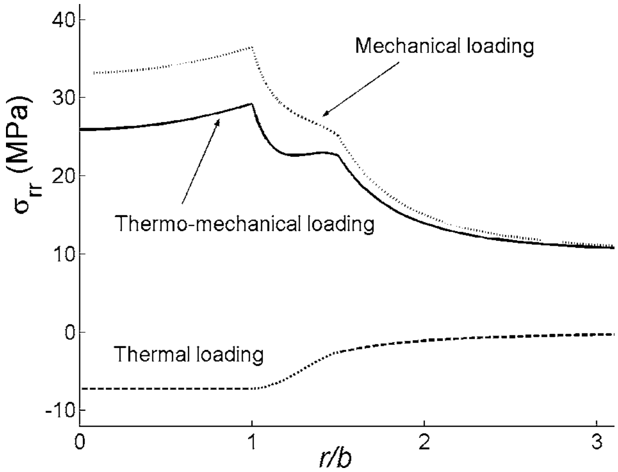


Fig. 12 Variation of the radial stress with distance from the center of the FG coated spherical particle along the x_3-axis; effect of loading conditions

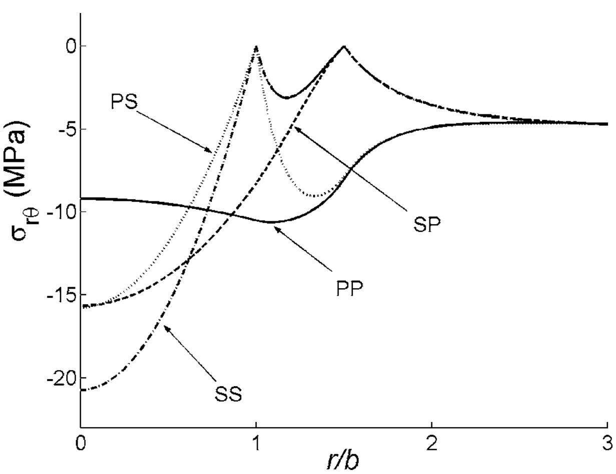


Fig. 10 Variation of $\sigma_{r\theta}$ with distance from the center of the FG coated spherical particle along the line $\theta=\pi/4$; effect of interface conditions

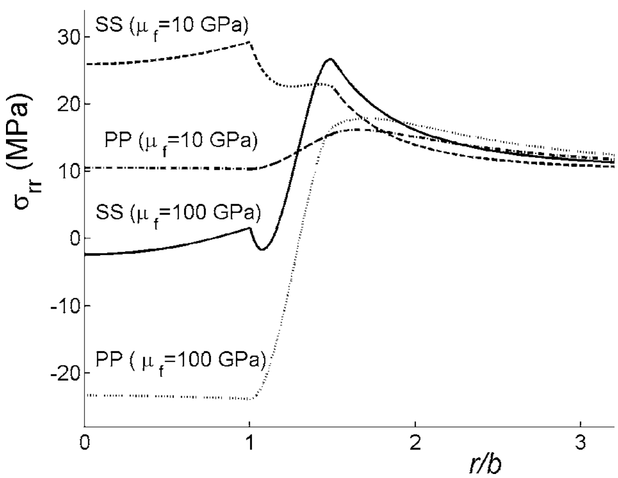


Fig. 13 Variation of the radial stress with distance from the center of the FG coated spherical particle along the x_3-axis; influence of particle stiffness and interface conditions

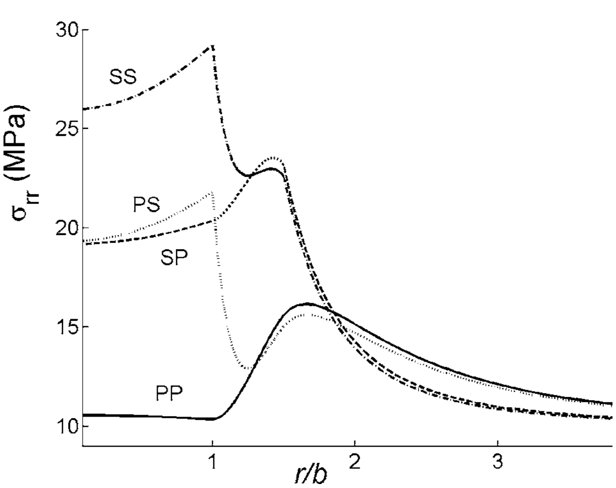


Fig. 11 Variation of the radial stress with distance from the center of the FG coated spherical particle along the x_3-axis; effect of interface conditions

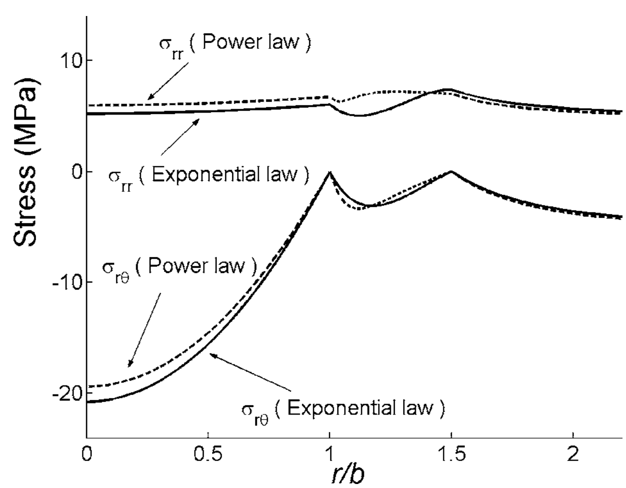


Fig. 14 Variation of σ_{rr} and $\sigma_{r\theta}$ with distance from the center of the FG coated spherical particle along the line $\theta=\pi/4$; influence of spatial distributions of thermal expansion coefficient and shear modulus of FGM coating

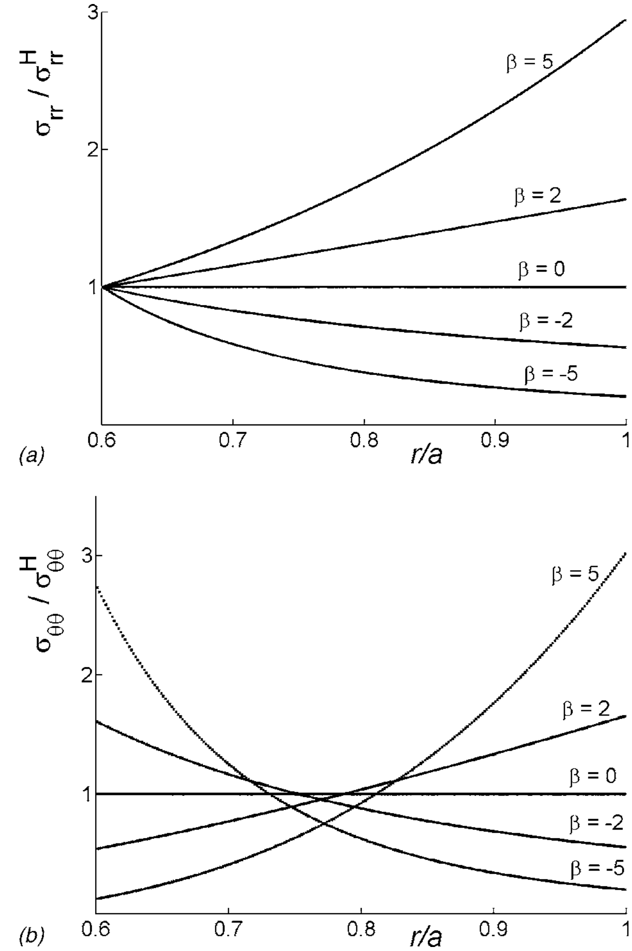


Fig. 15 (*a*) Variation of normalized radial stress for an FG cylindrical pressure vessel. (*b*) Variation of normalized hoop stress for an FG cylindrical pressure vessel.

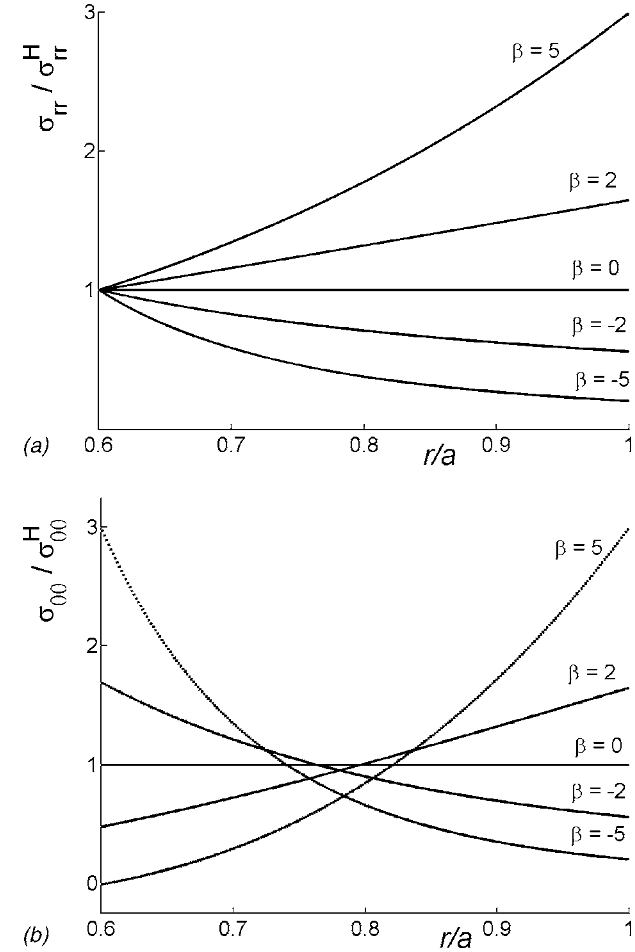


Fig. 16 (*a*) Variation of normalized radial stress for an FG spherical pressure vessel. (*b*) Variation of normalized hoop stress for an FG spherical pressure vessel.

stress distributions in both cylindrical and spherical cases reveals that they are not correct; for example, contrary to their claim, the hoop stress distribution of an isotropic cylindrical vessel cannot be obtained by setting $\beta=0$. The correct closed-form hoop stress expressions have been given in the Appendix C. Finally, it should be noted that, while the formulation of Tuntuncu and Ozturk is valid only for an elastic modulus distribution (21) and $|\beta| \leq 2$, the present method does not pose any restriction on the elastic modulus and/or Poisson's ratio distribution(s) associated with the FG layer.

6 Conclusion

The potential functions associated to coated cylindrical and spherical inhomogeneities under thermo-mechanical loadings are presented. This enables one to obtain the exact solutions pertinent to various combinations of matrix-coating and coating-fiber interface conditions, PP, PS, SP, and SS rigorously. There is no limitation on the number of phases, and the exact solution can be obtained when such conditions are imposed at any internal boundaries, for example at some internal boundaries within the coating, as in the case of multiple coatings. The excellent features and robustness of the potential function approach proved advantageous in arriving at a general accurate analytical solution of an FG coated cylindrical/spherical inhomogeneity with sliding/perfect interfaces. In this, the only condition on the composition of the FG coating is that to be piecewise continuous. Whereas the relevant existing contributions are limited to only perfect interfaces; moreover, either very special distributions for the material properties of the FG layer or very special kinds of loadings needs to be considered in most of the previous studies on the subject; see, for example, [5–12]. It is hoped that the results presented in this manuscript can serve as a benchmark not only for the relevant future work, but also revisiting the existing contributions because they can be treated much more simply and efficiently with the current approach. Since the discontinuities of some components of the elastic field across the FG coatings with sliding interfaces pose a challenge in performing pertinent experimentations and numerical studies, such as finite element, the present work can be particularly beneficial for validation purposes.

Acknowledgment

This research was in part supported by the center of excellence in structures and earthquake engineering at Sharif University of Technology.

Appendix A

In this Appendix, a systematic approach is developed to obtain the $6(N+2)$ unknown constants, where N can be an arbitrarily large number, involved in the analysis of an FG coated cylindrical inhomogeneity system as presented in Sec. 3. The remote boundary condition yields $d_m=T/4$, $e_m=-T/4$, and the conditions (11) are cast into the following form

$$A_K(R_K)X_K + D_K^T(R_K) = A_{K+1}(R_K)X_{K+1} + D_{K+1}^T(R_K) \qquad \text{(A1)}$$

where $D_K^T(r)$ is the thermal displacement field within the Kth layer due to the mismatch between the CTE's of the constituent phases

$$
A_K(r) = \begin{pmatrix}
-\dfrac{1}{2\bar{\mu}_K r} & 0 & \dfrac{\kappa-1}{\bar{\mu}_K} r & 0 & 0 & 0 \\
0 & \dfrac{1}{\bar{\mu}_K r^3} & 0 & -\dfrac{r}{\bar{\mu}_K} & \dfrac{1+\kappa}{2\bar{\mu}_K r} & \dfrac{\kappa-3}{2\bar{\mu}_K} r^3 \\
0 & \dfrac{1}{\bar{\mu}_K r^3} & 0 & \dfrac{r}{\bar{\mu}_K} & \dfrac{1-\kappa}{2\bar{\mu}_K r} & \dfrac{\kappa+3}{2\bar{\mu}_K} r^3 \\
\dfrac{1}{r^2} & 0 & 2 & 0 & 0 & 0 \\
0 & -\dfrac{6}{r^4} & 0 & -2 & -\dfrac{4}{r^2} & 0 \\
0 & -\dfrac{6}{r^4} & 0 & 2 & -\dfrac{2}{r^2} & 6r^2
\end{pmatrix},
$$

$$
X_K = \begin{pmatrix} a_K \\ c_K \\ d_K \\ e_K \\ f_K \\ g_K \end{pmatrix}, \quad D_K^T(r) = \begin{pmatrix} \bar{\alpha}_K r \Delta T \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{A2}
$$

Eq. (A1) can be written as

$$
X_{K+1} = B_K(R_K) X_K + C_K(R_K) \tag{A3}
$$

where

$$
B_k(R_K) = A_{K+1}^{-1}(R_K) A_K(R_K),
$$

$$
C_K(R_K) = A_{K+1}^{-1}(R_K)(D_K^T(R_K) - D_{K+1}^T(R_K)). \tag{A4}
$$

Now from Eq. (A3)

$$
X_{K+1} = M_1^K X_1 + P^K, \tag{A5}
$$

where

$$
M_l^K = \prod_{i=l}^{K} B_i(R_i), \tag{A6}
$$

$$
P^K = M_2^K C_1(R_1) + M_3^K C_2(R_2) + \cdots + M_K^K C_{K-1}(R_{K-1}) + C_K(R_K). \tag{A7}
$$

Letting $K=N-1$ in Eq. (A5),

$$
X_N = M_1^{N-1} X_1 + P^{N-1}. \tag{A8}
$$

Hence, Eq. (A8) along with (14)–(16) can be used to calculate the unknown constants for each phase.

Appendix B

The same approach developed in the previous appendix is followed to determine the unknown constants involved in the formulation of an FG coated spherical inhomogeneity system studied in Sec. 4. The far-filed boundary conditions yield $d_m = -2\nu e_m = T\nu/(1+\nu)$. Also the continuity conditions (11) can be rewritten in the form given by Eq. (A1) with $A_K(r)$ and X_K redefined as

$$
A_K(r) = \begin{pmatrix}
-\dfrac{1}{2\bar{\mu}_K r^2} & \dfrac{3}{4\bar{\mu}_K r^4} & 0 & -\dfrac{r}{2\bar{\mu}_K} & 0 & -\dfrac{\nu r^3}{\bar{\mu}_K} \\
0 & \dfrac{-9}{4\bar{\mu}_K r^4} & \dfrac{4\nu-5}{2\bar{\mu}_K r^2} & \dfrac{3r}{2\bar{\mu}_K} & \dfrac{(2\nu-1)r}{\bar{\mu}_K} & \dfrac{3\nu r^3}{\bar{\mu}_K} \\
0 & \dfrac{-3}{r^4} & \dfrac{(1-2\nu)}{\bar{\mu}_K r^2} & \dfrac{-3r}{2\bar{\mu}_K} & \dfrac{(1-2\nu)r}{\bar{\mu}_K} & \dfrac{(4\nu-7)r^3}{2\bar{\mu}_K} \\
\dfrac{2}{r^3} & \dfrac{-6}{r^5} & \dfrac{-2\nu}{r^3} & -1 & -2\nu & \nu r^2 \\
0 & \dfrac{18}{r^5} & \dfrac{2(5-\nu)}{r^3} & 3 & -2(1-2\nu) & -3\nu r^2 \\
0 & \dfrac{12}{r^5} & \dfrac{2(1+\nu)}{r^3} & -3 & 2(1-2\nu) & -(7+2\nu)r^2
\end{pmatrix}
$$

and

$$X_K = \begin{pmatrix} a_K \\ b_K \\ c_K \\ d_K \\ e_K \\ f_K \end{pmatrix}, \quad h_K = -\frac{4}{7}\nu f_K \tag{B1}$$

and D_k remains unchanged. A similar procedure described in Appendix A along with stress and displacement fields (19) can be used to compute all the unknown constants.

Appendix C

As mentioned in Sec. 5.3, the expressions derived by Tuntuncu and Ozturk [22] for the hoop stress distributions in FG cylindrical and spherical pressure vessels are incorrect. Following their notations, the correct closed-form expressions must read, respectively, as

$$\sigma_{\theta\theta} = \frac{P\left(\frac{b}{a}\right)^{1-\beta}(\chi_1 - \chi_2)}{\left(\left(\frac{b}{a}\right)^{m_1} - \left(\frac{b}{a}\right)^{m_2}\right)(m_1(\nu-1)-\nu)(m_2(\nu-1)-\nu)} r^{\beta-1} \tag{C1}$$

$$\chi_1 = r^{m_1}(m_2(\nu-1)-\nu)(1+(m_1-1)\nu)$$

$$\chi_2 = r^{m_2}(m_1(\nu-1)-\nu)(1+(m_2-1)\nu) \tag{C2}$$

and

$$\sigma_{\theta\theta} = \frac{P\left(\frac{b}{a}\right)^{1-\beta}(\psi_1 - \psi_2)}{\left(\left(\frac{b}{a}\right)^{s_1} - \left(\frac{b}{a}\right)^{s_2}\right)(s_1(\nu-1)-2\nu)(s_2(\nu-1)-2\nu)} \tag{C3}$$

$$\psi_1 = r^{s_1-1}(s_2(v-1)-2v)(1+s_1 v)r^{\beta}$$

$$\psi_2 = r^{s_2-1}(s_1(v-1)-2v)(1+s_2 v)r^{\beta} \tag{C4}$$

Now, in lieu of the above corrections, by setting $\beta=0$ the expressions for the hoop stress corresponding to homogeneous cylindrical and spherical pressure vessels (Eqs. (20) and (22) given in the paper of Tuntuncu and Ozturk [22]) are readily available.

References

[1] Mura, T., 1988, "Inclusion Problems," Appl. Mech. Rev., **41**, pp. 15–20.
[2] Mura, T., Shodja, H. M., and Hirose, Y., 1996, "Inclusion Problems," Appl. Mech. Rev., **49**, pp. S118–S127.
[3] Christensen, R. M., and Lo, K. H., 1979, "Solution for Effective Shear Properties in Three Phase Sphere and Cylinder Models," J. Mech. Phys. Solids, **27**, pp. 315–330.
[4] Love, A. E. H., 1927, *A Treatise on the Mathematical Theory of Elasticity*, Cambridge University Press, Cambridge.
[5] Herve, E., and Zaoui, A., 1993, "*n*-Layered Inclusion-Based Micromechanical Modeling," Int. J. Eng. Sci., **31**, pp. 1–10.
[6] Benveniste, Y., Dvorak, G. J., and Chen, T., 1989, "Stress Fields in Composites with Coated Inclusions," Mech. Mater., **7**, pp. 305–317.
[7] Lutz, M. P., and Ferrari, M., 1993, "Compression of a Sphere with Radially Varying Elastic Moduli," Composites Eng., **3**, pp. 873–884.
[8] Lutz, M. P., and Zimmerman, R. W., 1996, "Effect of the Interphase Zone on the Bulk Modulus of a Particulate Composite," J. Appl. Mech., **63**, pp. 855–861.
[9] Zimmerman, R. W., and Lutz, M. P., 1999, "Thermal Stresses and Thermal Expansion in a Uniformly Heated Functionally Graded Cylinder," J. Therm. Stresses, **22**, pp. 177–188.
[10] Mikata, Y., 1994, "Stress Fields in a Continuous Fiber Composite with a Variable Interphase under Thermo-Mechanical Loadings," J. Eng. Mater. Technol., **116**, pp. 367–377.
[11] Jayaraman, K., and Reifsnider, K. L., 1992, "Residual Stresses in a Composite with Continuously Varying Young's Modulus in the Matrix/Matrix Interphase," J. Compos. Mater., **26**, pp. 770–791.
[12] Jasiuk, I., and Kouider, M. W., 1993, "The Effect of an Inhomogeneous Interphase on the Elastic Constants of Transversely Isotropic Composites," Mech. Mater., **15**, pp. 53–63.
[13] Sadowsky, M. A., and Sternberg, E., 1949, "Stress Concentration Around a Triaxial Ellipsoidal Cavity," J. Appl. Mech., **16**, pp. 149–157.
[14] Mikata, Y., and Taya, M., 1985, "Stress Field in and Around a Coated Short Fiber in an Infinite Matrix Subjected to Uniaxial and Biaxial Loadings," J. Appl. Mech., **52**, pp. 19–25.
[15] Mikata, Y., and Taya, M., 1986, "Thermal Stress in a Coated Short Fibre Composite," J. Appl. Mech., **53**, pp. 681–689.
[16] Mura, T., Jasiuk, I., and Tuschida, B., 1985, "The Stress Field of a Sliding Inclusion," Int. J. Solids Struct., **21**(12), pp. 1165–1179.
[17] Ghahremani, F., 1980, "Effect of Grain Boundary Sliding on Anelasticity of Polycrystals," Int. J. Solids Struct., **16**(9), pp. 825–845.
[18] Lur'e, A. I., 1964, *Three-Dimensional Problems of the Theory of Elasticity*, Interscience Publishers, New York.
[19] Hashin, Z., 1991, "Thermoelastic Properties of Particulate Composites with Imperfect Interface," J. Mech. Phys. Solids, **39**, pp. 745–762.
[20] Hashin, Z., 2002, "Thin Interphase/Imperfect Interface in Elasticity with Application to Coated Fiber Composites," J. Mech. Phys. Solids, **50**, pp. 2509–2537.
[21] Gao, J., 1995, "A Circular Inclusion with Imperfect Interface: Eshelby's Tensor and Related Problems," J. Appl. Mech., **62**, pp. 860–866.
[22] Tuntuncu, N., and Ozturk, M., 2001, "Exact Solution for Stresses in Functionally Graded Pressure Vessels," Composites, Part B, **32**, pp. 683–686.

Thermal Buckling of Multi-Walled Carbon Nanotubes by Nonlocal Elasticity

Renfu Li
Post-Doctoral Fellow

George A. Kardomateas
Professor of Aerospace Engineering
Fellow ASME

Georgia Institute of Technology,
Atlanta, Georgia 30332-0150

The small internal length scales of nanomaterials/nano-devices may call the direct application of classical continuum models into question. In this research, a nonlocal elastic shell model, which takes the small scale effects into account, is developed to study the thermal buckling behavior of multi-walled carbon nanotubes. The multi-walled carbon nanotubes are considered as concentric thin shells coupled with the van der Waals forces between adjacent nanotubes. Closed form solutions are formulated for two types of thermal buckling of a double-walled carbon nanotube: Radial thermal buckling (as in a shell under external pressure) and axial thermal buckling. The effects of small scale effects are demonstrated, and a significant influence of internal characteristic parameters such as the length of the C-C *bond has been found on the thermal buckling critical temperature. The study interestingly shows that the axial buckling is not likely to happen, while the "radial" buckling may often take place when the carbon nano-tubes are subjected to thermal loading. Furthermore, a convenient method to determine the material constant, "e_0" and the internal characteristic parameter, "a," is suggested.*
[DOI: 10.1115/1.2200656]

1 Introduction

The discovery of carbon nanotubes by Ijima [1] has inspired the promise of a new generation of novel composites and devices such as nano-electronics, nano-devices, and nano-composites [2,3]. Attention has been attracted to the subject of understanding the material properties of carbon nanotubes and their applications [4–7]. In most of the studies, experimental and molecular-dynamics simulation approaches are often employed and many relevant results have been obtained. However, the experiments at the nano-scale are often hard to control, and the simulations by molecular-dynamics are difficult to accurately formulate and extremely expensive for large-scale atomic systems [8]. Therefore, many researchers have attempted to expand the classical continuum mechanics approach to the atomic or molecular-based discrete systems [8–11]. For example, adding the van der Waal's forces to the classical shell models in order to simulate the carbon nanotubes, etc. The classical continuum models are efficient and accurate in computations for a material system in large length scales. But the length scales at nano-meters such as in nano-materials or nano-devices may not be sufficiently big enough to homogenize the discrete structure into a continuum [12]. The assumption upon which the classical continnum mechanics is built may no longer be fully satisfied. Therefore, modifications which would take the small scale effect into account are needed in order to make use of the virtues of the well-developed classical continuum mechanics.

In 1972, Eringen proposed a theory, called nonlocal continuum mechanics [13], in an effort to deal with the small-scale structure problems. In the classical (local) continuum elasticity, the material particles are assumed to be continuously distributed and the stress tensor at a reference point is uniquely determined by the strain tensor at the same point. On the contrary, the nonlocal continuum mechanics is based on the constitutive functionals being functionals of the past deformation histories of all material points of the body concerned. The small length scale effects are counted by incorporating the internal characteristic length such as the length of the C-C bond into the constitutive relationship. Solutions from various problems support this theory [13–15]. For examples, the dispersion curves by the nonlocal model are in excellent agreement with those by the Born-Karman theory of lattice dynamics; the dislocation core and cohesive stress predicted by the nonlocal theory are close to those known in the physics of solids [14,15]. Recently, some researchers such as Peddieson et al. [16] and Sudak [17] applied Eringen's theory to the nano-scales structures. In particular, Peddieson et al. [16] illustrated by the Benoulli/Euler beam model that the small scale effects manifest themselves in the range of nano-meters. Sudak in [17] applied the nonlocal elasticity to the column buckling study of the axial buckling of carbon nano-tubes. The results in these papers indicate that the small length scales would have significant influences and the nonlocal continuum model can effectively capture these influences in the study of nano-structures.

The importance and significant results of thermal buckling for conventional structures such as plates and shells have been reported by many researchers [18]. But in our literature search, the study of thermal buckling of nano-tubes has received relatively little attention. In the current paper, a nonlocal multiple shell model is developed to investigate the thermal buckling problems of multi-walled carbon nanotubes. In this model, not only the terms concerning the van der Waals forces between adjacent nanotubes are incorporated into the Donnell shell model [19], but the full nonlocal constitutive relationship is also adopted in the derivation of the formulas. Therefore, this model includes both the interactions from the van der Waals forces and the effects from the internal small scales of the nano-devices. Compared with some nonlocal models in the literature for nano-tubes subjected to mechanical loading, our model is a comprehensive nonlocal elastic model in the sense that no approximation has been made in the use of the nonlocal elastic constitutive equations and each tube is treated as a shell, not a one-dimensional column. Since the thermal expansion coefficients of the carbon nano-tube are negative [7], the tube usually contracts when its temperature is elevated. Therefore, its thermal buckling modes are different from those of

Contributed by the Applied Mechanics Division of ASME for publication in the ASME Journal of Applied Mechanics. Manuscript received August 31, 2005; final manuscript received March 15, 2006. Review conducted by S. Mukherjee. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California – Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME Journal of Applied Mechanics.

conventional shell structures. For example, a simply supported conventional shell structure often shows axial buckling modes under increased temperature without radial constraints [18]. However, a fixed simply supported nano-tube is in tension in the axial direction, but may fall into a "radial" buckling mode because of the contraction, like a traditional shell under external radial compressive pressure. In this study two cases of temperature variations are investigated under fixed simply edge-supported boundary conditions: Radially elevated temperature and uniformly reduced temperature. Closed form expressions for the critical temperatures are obtained and numerical results for double-walled nanotubes are presented. The results show that the internal small scales have significant effects on the critical buckling temperature. A method to determine the internal characteristic parameter, "a" and material constant, "e_0" (defined in see Sec. 2) is also proposed.

This paper is organized as follows: Section 2 is a brief summary of the nonlocal continuum mechanics; Sec. 3 is the derivation of a nonlocal continuum elastic shell model for the multi-walled nanotubes; in Sec. 4, the thermal buckling for two types of temperature distributions are analyzed under simply end-supported boundary conditions, and closed form solutions are formulated for the double-walled nano-tubes; numerical results for double-walled nano-tubes and discussions are presented in Sec. 5; and some conclusions are suggested in Sec. 6.

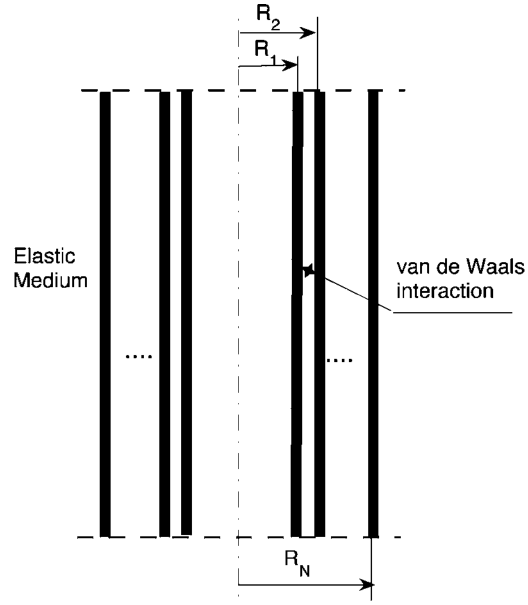


Fig. 1 A shell model of multi-walled nanotubes in an elastic medium

2 Summary of the Nonlocal Elasticity

In the classical (local) theory of elasticity, the stress tensor at a reference point x of a body can be determined by the strain tensor at that point. However, Eringen [13,14] pointed out that when we deal with a small scale structure, the media may no longer be considered continuously distributed and the internal characteristic length such as the length of the C-C bond in carbon nano-tubes should be considered. Based on this motivation, he developed a theory of nonlocal elasticity. The nonlocal mechanics says that the stress tensor at a reference point x in a body depends not only on the strain tensor at that point x, but also on the strain tensor at all other points, x' of this body. The constitutive equations of nonlocal elasticity read as

$$\sigma_{ij}(x) = \int_v \alpha(|x' - x|, \tau) c_{ijkl} \epsilon_{kl}(x') dv(x') \tag{1}$$

where, $\alpha(|x'-x|,\tau)$ is the nonlocal moduli; $\tau = e_0 a/l$ with a an internal characteristic length (e.g., lattice parameter, granular distance, length of C-C bonds), l an external characteristic length (e.g., crack length, wave-length), and e_0 a constant appropriate to each material and being determined by experiment or by matching dispersion curves of plane waves with those of atomic lattice dynamics; c_{ijkl} the elastic moduli tensor; $\epsilon_{kl}(x')$ the strain tensor.

Sometimes, it is hard to find an explicit expression for the kernel, $\alpha(|x'-x|,\tau)$ and do the integration of (1). In 1983, Eringen [15] developed some differential forms for practical use. For the two-dimensional nonlocal elasticity, a differential form corresponding to Eq. (1) is expressed as:

$$(1 - \tau^2 l^2 \nabla^2)\sigma_{ij} = c_{ijkl}\epsilon_{kl} \tag{2}$$

One may see that when the internal characteristic length a is neglected, i.e., the particles of a medium are considered to be continuously distributed, then τ, which is defined as ae_0/l is zero, and Eq. (2) reduces to the constitutive equation of classical elasticity. Also, it should be noted that, through Eq. (2), the ℓ is cancelled from the rest of the analysis, leaving a and e_0 as the internal characteristic constants.

3 A Nonlocal Continuum Model for Multi-Walled Carbon Nanotubes

Let x, s, z be the axial, circumferential and radial coordinates of the nanotube (Fig. 1), respectively. If the Donnell's assumptions [19] are used, then the strains and displacements of a nanotube have the following relations:

$$\epsilon_x = \epsilon_{0x} - zw_{,xx}, \quad \epsilon_s = \epsilon_{0s} - zw_{,ss}, \quad \gamma_{xs} = \gamma_{0xs} - zw_{,xs} \tag{3}$$

where $\epsilon_{0x}, \epsilon_{0s}$, and γ_{0xs} are the mid-surface strains. These are:

$$\begin{aligned} \epsilon_{0x} &= u_{,x} + \tfrac{1}{2}w_{,x}^2 \\ \epsilon_{0s} &= v_{,s} + \frac{1}{2}w_{,s}^2 + \frac{w}{R} \\ \gamma_{0xs} &= u_{,s} + v_{,x} + w_{,x}w_{,s} \end{aligned} \tag{4}$$

in which, u and v are, respectively, the axial and circumferential displacements of mid-surface, and w is the radial displacement; R is the radius of the mid-surface; the comma denotes differentiation with the corresponding coordinates. From Eq. (2), the nonlocal constitutive equations can be written as:

$$\begin{aligned} (1 - \tau^2 l^2 \nabla^2)\sigma_x &= \frac{E}{1-\nu^2}(\epsilon_x + \nu\epsilon_s) - \frac{E\alpha\Delta T}{1-\nu} \\ (1 - \tau^2 l^2 \nabla^2)\sigma_s &= \frac{E}{1-\nu^2}(\nu\epsilon_x + \epsilon_s) - \frac{E\alpha\Delta T}{1-\nu} \\ (1 - \tau^2 l^2 \nabla^2)\sigma_{xs} &= \frac{E}{2(1+\nu)}\gamma_{xs} \end{aligned} \tag{5}$$

in which $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial s^2$; $E, \nu, \alpha, \Delta T$ are the elastic modulus, Poisson's ratio, thermal expansion coefficient, and increase in temperature, respectively.

From Eqs. (3) and (5), one can have the following resultant forces and moments:

$$(1 - \tau^2 l^2 \nabla^2)N_x = K(\epsilon_{0x} + \nu\epsilon_{0s}) - \frac{N^T}{1-\nu}$$

$$(1-\tau^2 l^2\nabla^2)N_s = K(\nu\epsilon_{0x}+\epsilon_{0s}) - \frac{N^T}{1-\nu} \tag{6}$$

$$(1-\tau^2 l^2\nabla^2)N_{xs} = \frac{1-\nu}{2}K\gamma_{0xs}$$

and

$$(1-\tau^2 l^2\nabla^2)M_x = -D(w_{,xx}+\nu w_{,ss}) - \frac{M^T}{1-\nu}$$

$$(1-\tau^2 l^2\nabla^2)M_s = -D(w_{,ss}+\nu w_{,xx}) - \frac{M^T}{1-\nu} \tag{7}$$

$$(1-\tau^2 l^2\nabla^2)M_{xs} = -(1-\nu)Dw_{,xs}$$

where, $K=Eh/(1-\nu^2)$ and $D=Eh^3/12(1-\nu^2)$, with h being the thickness of the nano-tube; $N^T=\int E\alpha\Delta T dz$, and $M^T=\int E\alpha\Delta T z dz$ are the thermal resultant force and moment, respectively.

The nonlinear equilibrium equations, $\sigma_{ij,j}+f_i=0$, with $f_1=f_2=0$ and $f_3=p(x,s)$, can be expressed in terms of resultant forces and moments as:

$$(1-\tau^2 l^2\nabla^2)N_{x,x} + (1-\tau^2 l^2\nabla^2)N_{sx,s} = 0$$

$$(1-\tau^2 l^2\nabla^2)N_{sx,x} + (1-\tau^2 l^2\nabla^2)N_{s,s} = 0 \tag{8}$$

$$(1-\tau^2 l^2\nabla^2)M_{x,xx} + 2(1-\tau^2 l^2\nabla^2)M_{xs,xs} + (1-\tau^2 l^2\nabla^2)M_{s,ss}$$

$$-(1-\tau^2 l^2\nabla^2)\mathrm{N_L} - (1-\tau^2 l^2\nabla^2)\frac{1}{R}N_s + (1-\tau^2 l^2\nabla^2)p(x,s) = 0$$

where the N_{L} is a nonlinear operator defined as

$$\mathrm{N_L} = N_x w_{,xx} + 2N_{xs}w_{,xs} + N_s w_{,ss} \tag{9}$$

and $p(x,s)$ is the force acting in the radial direction. For nanotubes, it is the van der Waals interaction from an adjacent tube. This interaction is considered to be proportional to the radial displacement difference between two neighboring nano-tubes [8,11,17].

To study the buckling behavior of the nanotubes, the perturbation technique [20,21] is employed here. Let u_0, v_0, w_0 be the pre-buckling state of displacements, which satisfy equilibrium equations, u^1, v^1, w^1 a neighboring state, and the ξ is very small constant, then one has:

$$u = u_0 + \xi u^1, \quad v = v_0 + \xi v^1, \quad w = w_0 + \xi w^1 \tag{10}$$

The corresponding resultant forces and moments can be written as:

$$N_x = N_{0x} + \xi N_{1x}, \quad N_s = N_{0s} + \xi N_{1s}, \quad N_{xs} = N_{0xs} + \xi N_{1xs}$$

$$M_x = M_{0x} + \xi M_{1x}, \quad M_s = M_{0s} + \xi M_{1s}, \quad M_{xs} = M_{0xs} + \xi M_{1xs} \tag{11}$$

Considering the w_0 to be zero in the pre-buckling state [22,23] and substituting Eqs. (10) into Eq. (11), then into Eq. (8), leads to the following stability equation for nonlocal continuum elasticity of a nano-tube:

$$\nabla^8 w + 4k^4 w_{,xxxx} + \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4 \mathrm{N_L^0} = \frac{1}{D}[(1-\tau^2 l^2\nabla^2)\nabla^4 p(x,s)] \tag{12}$$

where $k^4=3(1-\nu^2)/(R^2h^2)$, and the N_{L}^0 operator reads as:

$$N_{\mathrm{L}}^0 = N_{0x}w_{,xx} + 2N_{0xs}w_{,xs} + N_{0s}w_{,ss} \tag{13}$$

N_{0x}, N_{0xs}, N_{0s} are the resultant forces at the pre-buckling state that satisfy the equilibrium equation (8). For convenience of writing, the superscript "1" is omitted in Eqs. (12) and (13).

Our structure is a multi-walled nano-tube and we shall denote by w_j, $j=1,2,\ldots,N$ the transverse displacements of each of the N walls. Also, we denote by $p_{j,(j+1)}$ the interaction pressure exerted on the tube j from the tube $j+1$. Applying Eq. (12) to each of the nano-tubes of the multi-walled carbon nanotube structure, one can obtain a set of nonlocal equations for the multi-walled carbon nanotube (i.e., w_1 is the displacement of the first nano-tube; w_2 is the displacement of the second nano-tube, etc.):

$$\begin{aligned}
&\nabla^8 w_1 + 4k_1^4 w_{1,xxxx} + \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4 \mathrm{N}_{L1}^0 \\
&= \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4 p_{12}(x,s) \\
&\nabla^8 w_2 + 4k_2^4 w_{2,xxxx} + \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4 \mathrm{N}_{L2}^0 \\
&= \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4\left[p_{23}(x,s) - \frac{R_1}{R_2}p_{12}(x,s)\right] \\
&\nabla^8 w_N + 4k_N^4 w_{N,xxxx} + \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4 \mathrm{N}_{LN}^0 \\
&= \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4\left[p_N(x,s) - \frac{R_{N-1}}{R_N}p_{(N-1)N}(x,s)\right]
\end{aligned} \tag{14}$$

where,

$$\begin{aligned}
&p_{12}(x,s) = c[w_2(x,s) - w_1(x,s)] \\
&p_{N(N-1)}(x,s) = c[w_N(x,s) - w_{N-1}(x,s)] \\
&p_N(x,s) = -k_0[w_N(x,s) - w_{N-1}(x,s)]
\end{aligned} \tag{15}$$

in which, $p_{j(j+1)}(x,s)$ is the interaction pressure exerted on the tube j from the tube $j+1$, while $p_{(j+1)j}(x,s)$ is the interaction pressure exerted on the tube $j+1$ from the tube j; these have the following relationship:

$$R_j p_{j(j+1)}(x,s) = -R_{j+1}p_{(j+1)j}(x,s), \quad j=1,2,\ldots,N-1 \tag{16}$$

p_N is the interaction pressure between the outmost tube and the surrounding elastic medium; k_0 is the spring constant of the surrounding elastic medium; c is the van der Waals interaction coefficient and can be estimated from the data given in [12,24], and it reads as:

$$c = \frac{200}{0.16\pi d^2}\ \mathrm{erg/cm^2}, \quad d = 0.142\ \mathrm{nm} \tag{17}$$

where d is the bond distance between carbon atoms in a graphite sheet. One may realize that the buckling phenomenon can be described as an infinitesimal deflection perturbation, so the van der Waals interaction and interaction between the outer tube and the elastic surrounding media can be estimated from a linear function of the deflection jump at two points, and the interactions in the tangential direction can be neglected, as discussed in Ref. [17]. But for post-buckling behavior, the nonlinear higher order terms and effects from the tangential force should be included in these interaction expressions.

Substitution of (15) into Eq. (14) yields the nonlocal model for the multi-walled nano-tubes:

$$\nabla^8 w_1 + 4k_1^4 w_{1,xxxx} + \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4 \mathrm{N}_L^0$$

$$= \frac{c}{D}(1-\tau^2 l^2\nabla^2)\nabla^4[w_2 - w_1]$$

$$\nabla^8 w_2 + 4k_2^4 w_{2,xxxx} + \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4 \mathrm{N}_L^0$$

$$= \frac{c}{D}(1-\tau^2 l^2\nabla^2)\nabla^4\left[w_3 - w_2 - \frac{R_1}{R_2}(w_2 - w_1)\right]$$

$$\nabla^8 w_{N-1} + 4k_{N-1}^4 w_{N-1,xxxx} + \frac{1}{D}(1-\tau^2 l^2 \nabla^2)\nabla^4 \mathrm{N}_L^0$$
$$= \frac{c}{D}(1-\tau^2 l^2\nabla^2)\nabla^4\left[w_N - w_{N-1} - \frac{R_{N-2}}{R_{N-1}}(w_{N-1}-w_{N-2})\right] \tag{18}$$

$$\nabla^8 w_N + 4k_N^4 w_{N,xxxx} + \frac{1}{D}(1-\tau^2 l^2\nabla^2)\nabla^4 \mathrm{N}_L^0$$
$$= \frac{c}{D}(1-\tau^2 l^2\nabla^2)\nabla^4\left[-\frac{k_0}{c}w_N - \frac{R_{N-1}}{R_N}(w_N - w_{N-1})\right]$$

Here in the set of Eqs. (18), one can clearly see that if the internal characteristic parameter, a is negligible, then $1-\tau^2 l^2\nabla^2 = 1$, and this model returns to the classical continuum elastic shell model including the van de Waals interaction via the coefficient c, and the interaction from the surrounding elastic medium via the coefficient k_0.

4 Thermal Buckling of the Multi-Walled Carbon Nanotubes

In the set of Eqs. (18), the thermal stress is related through the pre-buckling terms such as N_{0x} and N_{0s} in N_L^0. The edges of the nano-tubes are assumed to be simply supported and the buckling of the carbon nanotubes under two types of temperature distributions will be analyzed in the following study: (a) Radially elevated temperature and (b) uniformly reduced temperature. Without loss of generality, let us first consider the double-walled carbon nanotubes (DWCNT), i.e., $N=2$.

4.1 Radially Elevated Temperature. The thermal expansion coefficient of carbon nanotubes depends on the temperature range, radius, the number of layers and helix [6]. But in most cases, they are likely negative [7]. Hence, the nano-tubes contract when the temperature rises, while the surrounding media expands. Therefore, the tube can be viewed as a shell under external compressive pressure.

If the temperature rise varies linearly across the nano-tube thickness as

$$\Delta T(z) = \Delta T_0 \frac{z+h/2}{h}, \quad -\frac{h}{2} < z < \frac{h}{2} \tag{19}$$

then from pre-buckling equilibrium, one can have $N_{0x}=N_{0xs}=0$ in the operator N_L^0, and

$$N_{0s} = -\frac{E\alpha h}{1-\nu}\int_{-h/2}^{h/2}\Delta T(z)dz = -\frac{E\alpha h}{2(1-\nu)}\Delta T_0 \tag{20}$$

The stability equations from (18) can then be written as:

$$\nabla_{R1}^8 w_1 + 4k_1^4 w_{1,xxxx} = -\frac{N_{0s}}{D}(1-\tau^2 l^2\nabla_{R1}^2)\nabla_{R1}^4 w_{1,ss} + \frac{1}{D}(1-\tau^2 l^2\nabla_{R1}^2)\nabla_{R1}^4[c(w_2-w_1)] \tag{21}$$
$$\nabla_{R2}^8 w_2 + 4k_2^4 w_{2,xxxx} = -\frac{N_{0s}}{D}(1-\tau^2 l^2\nabla_{R2}^2)\nabla_{R2}^4 w_{2,ss} - \frac{1}{D}(1-\tau^2 l^2\nabla_{R2}^2)\nabla_{R2}^4\left[k_0 w_2 + c\frac{R_1}{R_2}(w_2-w_1)\right]$$

where,

$$\nabla_{R_j}^2 = \frac{\partial^2}{\partial x^2} + \frac{1}{R_j^2}\frac{\partial^2}{\partial\theta^2}, \quad j=1,2 \tag{22}$$

One can see that a solution for each of the concentric nanotubes can take the form:

$$w_1 = A_1\sin\left(\frac{m\pi}{L}x\right)\sin(n\pi\theta), \quad w_2 = A_2\sin\left(\frac{m\pi}{L}x\right)\sin(n\pi\theta) \tag{23}$$

which satisfies the edge conditions

$$w_1 = w_{1,xx} = 0, \quad w_2 = w_{2,xx} = 0 \tag{24}$$

where m is the axial half wave-number, and n the circumferential half wave-number.

If one defines:

$$\Lambda_{R1} = \frac{m^2\pi^2}{L^2} + \frac{n^2\pi^2}{R_1^2}, \quad \Lambda_{R2} = \frac{m^2\pi^2}{L^2} + \frac{n^2\pi^2}{R_2^2}$$
$$\beta_1 = \frac{D}{(1+\tau^2 l^2\Lambda_{R1})\Lambda_{R1}^2}\left(\Lambda_{R1}^4 + \frac{m^4\pi^4}{L^4}4k_1^4\right) \tag{25}$$
$$\beta_2 = \frac{D}{(1+\tau^2 l^2\Lambda_{R2})\Lambda_{R2}^2}\left(\Lambda_{R2}^4 + \frac{m^4\pi^4}{L^4}4k_2^4\right)$$

Then, substitution of (23) into Eq. (21) yields:

$$\left[\beta_1 + c - \left(\frac{n\pi}{R_1}\right)^2 N_{0s}\right]A_1 - cA_2 = 0 \tag{26}$$
$$-c\frac{R_1}{R_2}A_1 + \left[\beta_2 + c\frac{R_1}{R_2} + k_0 - \left(\frac{n\pi}{R_2}\right)^2 N_{0s}\right]A_2 = 0$$

For non-trivial solutions of A_1 and A_2, the determinant of Eq. (26) must vanish, i.e.,

$$\left[\beta_1 + c - \left(\frac{n\pi}{R_1}\right)^2 N_{0s}\right]\left[\beta_2 + c\frac{R_1}{R_2} + k_0 - \left(\frac{n\pi}{R_2}\right)^2 N_{0s}\right] - \frac{R_1}{R_2}c^2 = 0 \tag{27}$$

which leads to:

$$\Delta T_0^{NL} = \frac{1-\nu}{E\alpha h}\left(\frac{1}{n\pi}\right)^2\left[\left(\beta_1 R_1^2 + \beta_2 R_2^2 + k_0 R_2^2 + cR_1^2 + \frac{R_1}{R_2}cR_2^2\right) - \sqrt{\left(\beta_2 R_2^2 - \beta_1 R_1^2 + k_0 R_2^2 - cR_1^2 + \frac{R_1}{R_2}cR_2^2\right)^2 + 4cR_1^2\frac{R_1}{R_2}cR_2^2}\right] \tag{28}$$

If the effects of small scale are neglected, the constants β_1 and β_2 in (25) are redefined as:

$$\beta_1^o = D\left(\Lambda_{R1}^2 + \frac{m^4\pi^4}{L^4\Lambda_{R1}^2}4k_1^4\right)$$
$$\beta_2^o = D\left(\Lambda_{R2}^2 + \frac{m^4\pi^4}{L^4\Lambda_{R2}^2}4k_2^4\right) \tag{29}$$

and the buckling temperature variation reads:

$$\Delta T_0^L = \frac{1-\nu}{E\alpha h}\left(\frac{1}{n\pi}\right)^2\left[\left(\beta_1^o R_1^2 + \beta_2^o R_2^2 + k_0 R_2^2 + cR_1^2 + \frac{R_1}{R_2}cR_2^2\right) - \sqrt{\left(\beta_2^o R_2^2 - \beta_1^o R_1^2 + k_0 R_2^2 - cR_1^2 + \frac{R_1}{R_2}cR_2^2\right)^2 + 4cR_1^2\frac{R_1}{R_2}cR_2^2}\right] \tag{30}$$

To investigate the effects of small length scale on the thermal buckling temperature variation, the ratio $\phi(a,e_0,c,k_0,R_1,R_2,L,E,\alpha)$ can be defined as:

$$\phi(a,e_0,c,k_0,R_1,R_2,L,E,\alpha) = \frac{\Delta T_0^{NL}}{\Delta T_0^L} \tag{31}$$

The superscripts NL and L denote the results obtained by the nonlocal elastic model and the classical model, respectively.

4.2 Uniformly Reduced Temperature. Now let us consider the case that the temperature of the nano-tubes uniformly decreases by:

$$T(x) = \Delta T_0, \quad 0 < x < L \tag{32}$$

Due to this temperature change, a compressive axial force is developed at the fixed ends of the nano-tubes. In the radial direction, the surrounding elastic medium contracts away from the nano-tubes while the tubes expand. Therefore, the pre-buckling forces are: $N_{0s}=N_{0xs}=0$ in the operator N_L^0, and

$$N_{0x} = \frac{E\alpha h}{1-\nu}\Delta T_0 \tag{33}$$

Now, the stability equations become:

$$\nabla_{R1}^8 w_1 + 4k_1^4 w_{1,xxxx} = -\frac{N_{0x}}{D}(1-\tau^2 l^2 \nabla_{R1}^2)\nabla_{R1}^4 w_{1,xx} + \frac{1}{D}(1-\tau^2 l^2 \nabla_{R1}^2)\nabla_{R1}^4[c(w_2-w_1)]$$

$$\nabla_{R2}^8 w_2 + 4k_2^4 w_{2,xxxx} = -\frac{N_{0x}}{D}(1-\tau^2 l^2 \nabla_{R2}^2)\nabla_{R2}^4 w_{2,xx} - \frac{1}{D}(1-\tau^2 l^2 \nabla_{R2}^2)\nabla_{R2}^4\left[k_0 w_2 + c\frac{R_1}{R_2}(w_2-w_1)\right] \tag{34}$$

The expression (23) can also be a solution to this case. The argument similar to the one used in obtaining (26) yields:

$$\left(\beta_1 + c - \frac{m^2\pi^2}{L^2}N_{0x}\right)A_1 - cA_2 = 0$$
$$-c\frac{R_1}{R_2}A_1 + \left(\beta_2 + c\frac{R_1}{R_2} + k_0 - \frac{m^2\pi^2}{L^2}N_{0x}\right)A_2 = 0 \tag{35}$$

which leads to:

$$\Delta T_0 = \frac{1-\nu}{2E\alpha h}\left(\frac{L}{m\pi}\right)^2\left\{\beta_1+\beta_2+k_0+\left(1+\frac{R_1}{R_2}\right)c - \sqrt{\left[\beta_2-\beta_1+k_0-\left(1-\frac{R_1}{R_2}\right)c\right]^2 + 4\frac{R_1}{R_2}c^2}\right\} \tag{36}$$

A ratio similar to (31) can also be defined for this case, in order to compare with the classical (as opposed to the nonlocal elastic) model.

4.3 Number of the Carbon Nanotubes, $N>2$. Though the above study is for $N=2$, the procedure can be extended to multi-walled carbon nanotubes ($N>2$). For example, if the multi-walled nano-tube is subjected to the thermal loading defined in Eq. (19), the critical temperature variation ΔT_0, through Eq. (20), is found by setting equal to zero the determinant of the following matrix:

$$\begin{bmatrix} a_{1,1}+\left(\frac{n\pi}{R_1}\right)^2 N_{os} & -c & 0 & \ldots & 0 \\ -c\frac{R_1}{R_2} & a_{2,2}-\left(\frac{n\pi}{R_2}\right)^2 N_{os} & -c & 0 & \ldots \\ 0 & \ldots\ldots\ldots & & & \\ \ldots & 0 & -\frac{R_{N-2}}{R_{N-1}}c & a_{N-1,N-1}-\left(\frac{n\pi}{R_{N-1}}\right)^2 N_{os} & -c \\ 0 & \ldots & 0 & -\frac{R_{N-1}}{R_N}c & a_{N,N}-\left(\frac{n\pi}{R_N}\right)^2 N_{os} \end{bmatrix} \tag{37}$$

where,

$$a_{1,1} = \beta_1 + c; \quad a_{2,2} = \beta_2 + \left(1+\frac{R_1}{R_2}\right)c$$
$$a_{N-1,N-1} = \beta_{N-1} + \left(1+\frac{R_{N-2}}{R_{N-1}}\right)c; \quad a_{N,N} = \beta_N + \frac{R_{N-1}}{R_N}c + k_o \tag{38}$$

Setting the determinant of (37) to zero results in an Nth polynomial equation in terms of ΔT_0, which can be solved numerically for $N>2$. The minimum of the solutions for ΔT_0 is the desired critical solution.

5 Results and Discussion

Numerical results are presented here for double-walled carbon nano-tubes in an elastic medium. The original data are chosen as: The length of a C-C bond is $a=1.42$ nm [13]; the material thermal expansion coefficient $\alpha=-1.60\times10^{-6}/{}^\circ$K, $E=742$ GPa, $\nu=0.17$ [7]; the radius of the inner carbon tube $R_1=R_1^o=0.35$ nm and the radius of the outer carbon tube $R_2=R_2^o=0.79$ nm (Fig. 1), the thickness of each nano-tube $h=(R_2-R_1)/4$, the length of the nano-tubes $L=10\times R_2$; the van de Waals interaction coefficient $c=0.0694$ TPa, the spring constant from the surrounding elastic medium $k_0=0.001\times c$.

Presented in Fig. 2 is the influence of the internal characteristic parameter, $a:\phi(a,e_0,\ldots,\alpha)$, defined in Eq. (31), versus a for different sizes of the nano-tubes. Both the axial half wave-number, m and circumferential half wave-number, n are 1, which corresponds to the critical value of ΔT_0 (see Fig. 5 for details). Here, the material constant, e_0 is assumed as 0.39 [15].

Three observations can be made here: First, for each of the tubes (except the one with $R_1=300R_1^o, R_2=300R_2^o$), the $\phi(a,e_0,\ldots,\alpha)$ decreases notably as the internal characteristic parameter, a increases. This means that the classical shell model may give over-estimated predictions for the thermal buckling temperature when a structure is in the scope of nanometers. Second, when the inner tube diameter of the multi-walled carbon tube is larger than 1.05×10^{-8} m (out of the nanometer range), the length of the C-C bond (an internal characteristic parameter) does not have an influence on the thermal buckling behavior (see the dotted curve of Fig. 2). In other words, if the structure considered falls above the nanometer range, the nonlocal model and classical shell model will give very close predictions. Third, for carbon nano-tubes, the internal parameter a (=1.42 nm) has a significant influ-

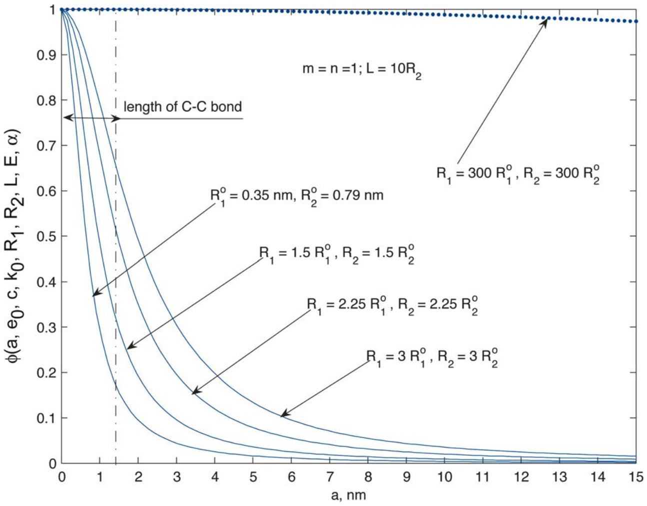


Fig. 2 The influence of the internal characteristic parameter, "*a*"

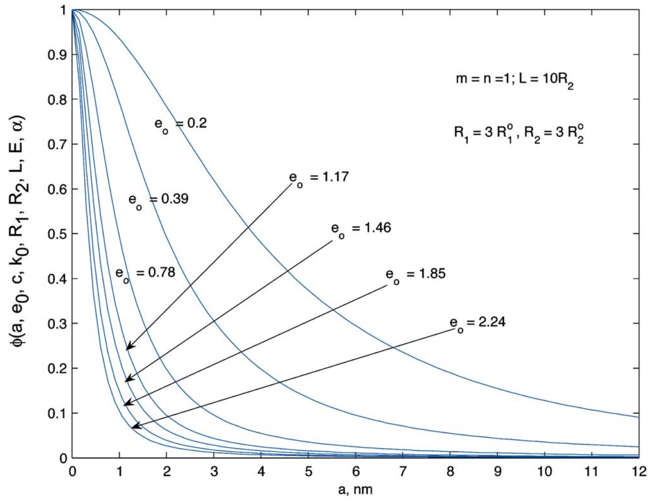


Fig. 4 The influence of the internal parameter "*a*" or various values of "e_0"

ence on the thermal buckling point. For example, if the inner radius of the double-walled carbon nanotubes is 0.35 nm and the outer radius is 0.79 nm, the ratio function $\phi(a, e_0, \ldots, \alpha) = 0.18$.

Plotted in Fig. 3 are the influences from the combination of the internal characteristic parameter and the material constant, ae_0, while the curves depicted in Fig. 4 are the influences of the internal parameter a under different material constants, e_0. The decreasing tendency similar to the one in Fig. 2 is also shown in Figs. 3 and 4, i.e., the value of $\phi(a, e_0, \ldots, \alpha)$ decreases as the value of ae_0 increases (Fig. 3), or as the internal length parameter, a increases for each given e_0 (Fig. 4). One may further observe that the curves in Figs. 3 and 4 together can provide a convenient way to determine the values of the material constant, e_0 and the internal characteristic parameter, a for each material. Here is the procedure: (1) Experimentally find the critical buckling temperature variation ΔT_0^{NL}, then calculate the value of $\phi_{ae_0} = \phi(a, e_0, c, k_0, R_1, R_2, L, E, \alpha)$ via dividing ΔT_0^{NL} by ΔT_0^{L}; (2) Using the ϕ_{ae_0} to locate the value of ae_0 from the curves in Fig. 3; (3) Using the same ϕ_{ae_0} to locate the values of the pair (a, e_0) from the curves in Fig. 4; (4) the values of the pair (a, e_0), whose product equals the value of ae_0 in step two are the values for a and e_0 of the material, respectively. Sudak [17] discussed the determination of the material constant e_0 for the known a. But the suggested procedure solves for the e_0 and a simultaneously.

Results in Figs. 5 and 6 are the critical ΔT_0s of two types of thermal buckling modes (radial buckling due to elevated radial temperature and axial buckling due to reduced temperature [18]) versus the axial half wave-number, m and the circumferential half wave-number, n. The material constant e_0 used in Figs. 5 and 6 is assumed to be 0.78, which is very close to the estimated value of 0.775 from the data given in [25]. The other data are the original ones described in the beginning of the section. Figure 5 shows that thermal buckling happens when both the axial half wave-number, m and circumferential half wave-number are 1, and the ΔT_0 is increased by 840.25°C for this double-walled nano-tubes. However, from the result in Fig. 6, one can see that the absolute value of ΔT_0 ("–" means temperature decreased) is very big. Physically speaking, this type of buckling mode will never happen for the carbon nano-tubes, which by itself is an interesting observation.

Finally, it should be mentioned that in this paper we study buckling and therefore, as is well known in bifurcation theory, we do not need to consider the geometric nonlinearities. Of course, if post-buckling is studied, nonlinearities need to be included, as in [26,27].

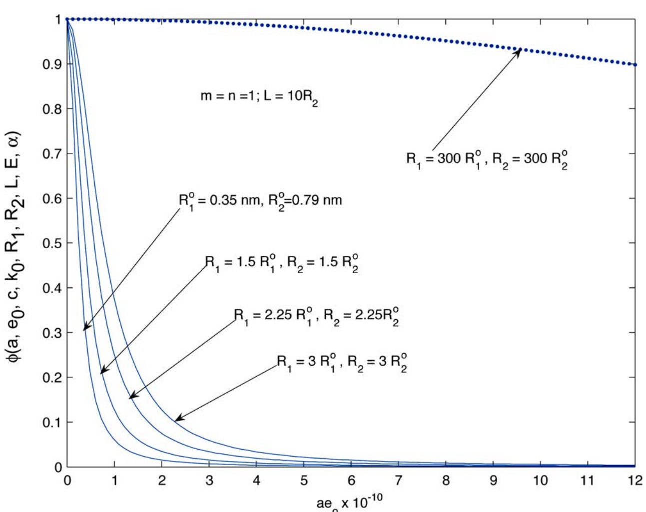


Fig. 3 The influence of the combined parameter "ae_0"

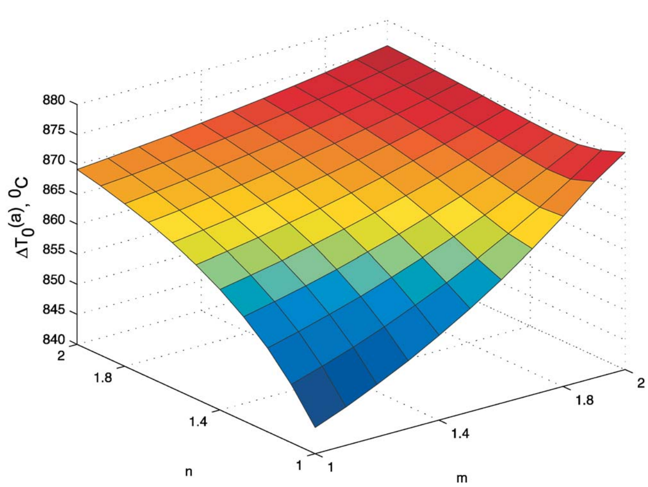


Fig. 5 Critical temperature $\Delta T_0^{(a)}$ versus (m, n)

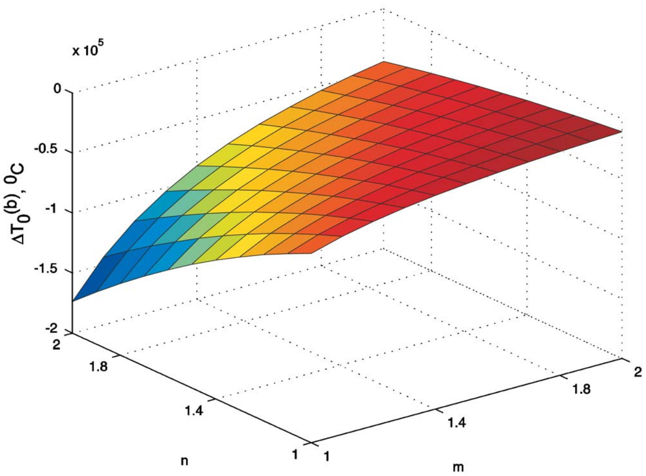


Fig. 6 Critical temperature $\Delta T_0^{(b)}$ versus (m,n)

6 Conclusions

In this paper, a nonlocal multi-walled shell model is developed to investigate the thermal buckling phenomenon of multi-walled carbon nano-tubes in an elastic medium. The tubes are considered concentric shells coupled by the van der Waals' interaction between adjacent tubes. Closed form expressions for the critical temperature variation are formulated for double-walled carbon nano-tubes, and numerical results are presented to demonstrate the influences of the small length scales in the study of nano-devices. Observations in this research suggest the following conclusions: (1) When the structure considered falls into the nano-meter range, the material internal characteristic parameter has a significant influence on the outcomes of this study. The nonlocal mechanics model is an appropriate model. But when the geometrical sizes of the structure are bigger than nano-meter scales, the classical elastic model is valid; (2) the curves of $\phi(a,e_0,\ldots,\alpha)$ versus ae_0 and $\phi(a,e_0,\ldots,\alpha)$ versus a under different e_0 can be used to determine the internal parameter a, and the material constant, e_0; (3) for carbon nano-tubes, the possible thermal buckling mode is the "radial" buckling mode (as in a shell under external pressure).

Acknowledgments

The financial support of the Office of Naval Research, Ship Structures and Systems, S and T Division, Grants Nos. N00014-90-J-1995 and N00014-0010323, and the interest and encouragement of the Grant Monitor, Dr. Y.D.S. Rajapakse, are both gratefully acknowledged.

References

[1] Ijima, S., 1991, "Helical micro-tubes of graphite carbon," Nature (London), **354**, pp. 56–58.
[2] Frankland, S. J. V., Harik, V. M., Odegard, G. M., Brenner, D. W., and Gates, T. S., 2003, "The Stress-Strain Behavior of Polymer-Nanotube Composites from Molecular Dynamics Simulation," Compos. Sci. Technol., **63**, pp. 1655–1661.
[3] Ma, H., Zeng, J., Lynn, M., Kumar, S., and Schiraldi, D., 2003, "Processing, Structure, and Properties of Fiber from Polyester/Carbon Nanofiber Composites," Compos. Sci. Technol., **63**, pp. 1617–1628.
[4] Krishnam, A., Dujardin, E., Ebbesen, T., Yianilos, P. N., and Treacy, M. M. J., 1998, "Young's Modulus of Singled-Walled Nanotubes," Phys. Rev. B, **58**, pp. 14013.
[5] Luo, J., and Daniel, I. M., 2003, "Characterization and Modeling of Mechanical Behavior of Polymer/Clay Nano-Composites," Compos. Sci. Technol., **63**, pp. 1607–1616.
[6] Schelling, P. K., and Keblinski, P., 2003, "Thermal Expansion of Carbon Structures," Phys. Rev. B, **68**, pp. 035425–035432.
[7] Pipes, R. B., and Hubert, P. 2003, "Helical Carbon Nanotubes Arrays: Thermal Expansion," Compos. Sci. Technol., **63**, pp. 1571–1579.
[8] Ru, C. Q., 2000, "Degraded Axial Buckling Strain of Multiwalled Carbon Nanotubes Due to Interlayer Slips," J. Appl. Phys., **89**, pp. 3426–3433.
[9] Harik, V. M., 2001, "Ranges of Applicability for Continuum Beam Model in the Mechanics of Carbon Nanotubes and Nano-rod," Solid State Commun., **120**, pp. 331–335.
[10] Govindjee, S., and Sackman, J. L., 1999, "On the Use of Continuum Mechanics to Estimate the Properties of Nanotubes," Solid State Commun., **110**, pp. 227–230.
[11] Shen, H-S, 2004, "Postbuckling Prediction of Double-Walled Carbon Nanotubes Under Hydrostatic Pressure," Int. J. Solids Struct., **41**, pp. 2643–2657.
[12] Girifalco, L. A., and Lad, R. A., 1955, "Energy of Cohesion, Compressibility, and the Potential Energy Functions of the Graphite System," J. Chem. Phys., **25**, pp. 693–697.
[13] Eringen, A. C., 1972, "Nonlocal Polar Elastic Continua," Int. J. Eng. Sci., **10**, pp. 1–16.
[14] Eringen, A. C., and Edelen, D. G. B., 1972, "On Nonlocal Elasticity," Int. J. Eng. Sci., **10**, pp. 233–248.
[15] Eringen, A. C., 1983, "On Differential Equations of Nonlocal Elasticity and Solutions of Screw Dislocation and Surface Waves," J. Appl. Phys., **54**, pp. 4703–4710.
[16] Peddieson, J., Buchanan, R., and McNitt, R. P., 2003, "Application of Nonlocal Continuum Models to Nanotechnology," Int. J. Eng. Sci., **41**, pp. 305–312.
[17] Sudak, L. J., 2003, "Column Buckling of Multi-Walled Carbon Nanotubes Using Nonlocal Continuum Mechanics," J. Appl. Phys., **94**, pp. 7281–7287.
[18] Thornton, E. A., 1993, "Thermal Buckling of Plates and Shells," Appl. Mech. Rev., **46**(10), pp. 485–506.
[19] Brush, D. O., and Almroth, B. O., 1975, "*Buckling of Bars, Plates and Shells*, McGraw-Hill, New York.
[20] Li, R., Frostig, Y., and Kardomateas, G. A., 2001, "Nonlinear High-Order Response of Imperfect Sandwich Beams with Delaminated Faces," AIAA J., **39**(9), pp. 1782–1787.
[21] Kardomateas, G. A., 1993, "The Initial Post-Buckling and Growth Behavior of Internal Delaminations in Composite Plates," J. Appl. Mech., **60**, pp. 903–910.
[22] Meyers, C. A., and Hyer, M. W., 1991, "Thermal Buckling and Postbuckling of Symmetrically Laminated Composites Plates," J. Therm. Stresses, **14**, pp. 519–540.
[23] Eslami, M. R., Ziaili, A. R., and Ghorbanpour, A., 1996, "Thermal Buckling of Thin Cylindrical Shells on Improved Stability Equations," J. Therm. Stresses, **19**, pp. 299–315.
[24] Saito, R., Matsuo, R., Kimura, T., Dresselhaus, G., and Dresselhaus, M. S., 2001, "Anomalous Potential Barrier of Double-Wall Carbon Nanotubes," Chem. Phys. Lett., **348**(9), pp. 187–193.
[25] Wang, L., and Hu, H., 2005, "Flexural Wave Propagation in Single-Walled Carbon Nanotubes," Phys. Rev. B, **71**, pp. 195412–195419.
[26] Yakobson, B. I., Brabec, C. J., and Bernholc, J., 1996, "Nanomechanics of Carbon Tubes: Instabilities Beyond Linear Response," Phys. Rev. Lett., **76**, pp. 2511–2514.
[27] 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**, pp. 341–348.

Sen Yung Lee[1]
Professor of Mechanical Engineering Department,
National Cheng Kung University,
Department of Mechanical Engineering,
Tainan, 701 Taiwan
e-mail: sylee@mail.ncku.edu.tw

Jer Jia Sheu
National Cheng Kung University,
Tainan, Taiwan,
Republic of China

Free Vibrations of a Rotating Inclined Beam

By utilizing the Hamilton principle and the consistent linearization of the fully nonlinear beam theory, two coupled governing differential equations for a rotating inclined beam are derived. Both the extensional deformation and the Coriolis force effect are considered. It is shown that the vibration system can be considered as the superposition of a static subsystem and a dynamic subsystem. The method of Frobenius is used to establish the exact series solutions of the system. Several frequency relations that provide general qualitative relations between the natural frequencies and the physical parameters are revealed without numerical analysis. Finally, numerical results are given to illustrate the general qualitative relations and the influence of the physical parameters on the natural frequencies of the dynamic system. [DOI: 10.1115/1.2200657]

1 Introduction

Rotating beams, which have importance in many engineering applications such as turbine blades, helicopter rotor blades, airplane propellers, robot manipulators, and various cooling fans, have been studied for a long time. An interesting review can be found in the related papers by Leissa [1], Ramamurti et al. [2] and Rao [3].

In the vibrational analysis, the structures were often modeled as beams vibrating in flexural motion. The influence of tip mass, rotating speed, shear deformation and rotatory inertia, hub radius, setting angle, taper ratio, pretwisted angle, and elastic root restraints on the natural frequencies of flexural vibration of a rotating beam were investigated by many investigators [4–17]. However, most of the qualitative conclusions are not general conclusions and are based on tremendous numerical analysis, except the studies on the upper and lower bounds of the fundamental frequency of a rotating beam [15–17]. The upper and lower bounds of the fundamental frequencies of a rotating uniform beam with a clamped root were obtained analytically by Schihansl [15] and Pnueli [16]. They concluded that the bending frequency of a rotating beam with clamped root, because of centrifugal stiffening, is higher than that of a nonrotating beam. Lee and Kuo [17] showed that if the setting angle is not equal to zero, the fundamental frequency of a rotating uniform beam with an elastically restrained root can be less than that of a nonrotating beam, and the phenomenon of rotating instability may occur.

Even though the vibration of a rotating beam has been extensively studied, to the authors' knowledge, a review on the literature reveals that the effect of inclination angle, which is considered in the recent computer cooling fan design on the natural frequencies of rotating beams, has never been studied before.

In this paper, based on the Bernoulli-Euler beam theory, one considers the free vibration of a rotating uniform beam with an inclination angle. The extensional deformation and the Coriolis force effect are considered. The beam considered is doubly symmetric, such that the centroidal axis and the neutral axis are coincident. By utilizing the Hamilton principle and the consistent linearization of the fully nonlinear beam theory [18,19], two coupled nonhomogeneous governing differential equations for the rotating inclined beam are derived. It is shown that the analysis of the problem can be simplified by considering the system as the superposition of two subsystems. One of them is a static system and the other is a dynamic system. The natural frequencies of the inclined beam will be determined from the dynamic subsystem governed by two coupled homogeneous differential equations. One of the coefficients of the differential equations is variable. In general, the exact solutions of the system are not available. The kind of problem was mainly solved by approximated methods such as the Galerkin method [4], the Rayleigh–Ritz method [5], the finite difference method [6], the dynamic stiffness method [11], and the finite element method [9,12–14]. In the present study, it is shown that the coupled characteristic differential equations can be decoupled and reduced to a sixth order ordinary differential equation with variable coefficients. By using the method of Frobenius and extending the works of Stafford and Giurgiutiu [20,21] and Lee and Kuo [22], the exact series solution of the system is developed. Several frequency relations that provide general qualitative relations between the natural frequencies and the physical parameters are revealed without a numerical analysis. Finally, numerical results are given to illustrate the general qualitative relations and the influence of the physical parameters on the natural frequencies of the dynamic system.

2 Governing Differential Equations

Consider the free vibration of a rotating inclined Bernoulli-Euler beam, as shown in Fig. 1. The beam is mounted with a setting angle ψ and an inclination angle θ on a hub with radius r_h. It rotates with a constant angular velocity Ω. Three coordinate systems, $O-X_0Y_0Z_0$, $A-X_1Y_1Z_1$, and $A-X_2Y_2Z_2$, are used in the expression of the configuration. The centroidal axis of the beam is coincident with the X_2 axis. It is assumed that the thickness of the beam is relatively small to the width of the beam. The deflection of the beam is in the X_2-Y_2 plane. The position vector of point P at an arbitrary cross section in the rotating beam, after deformation, can be expressed as

$$\overrightarrow{OP} = [r_h + (x+u)\cos\theta - v\sin\theta\sin\psi]\vec{i} + [(x+u)\sin\theta + v\cos\theta\sin\psi]\vec{j} + v\cos\psi\vec{k} \quad (1)$$

where u and v are the centroidal axial and transverse displacements, respectively, x is the distance from the origin point A to the position of point P, and $\vec{i}$, $\vec{j}$, $\vec{k}$ are unit vectors in the $O-X_0Y_0Z_0$ coordinate system. The velocity of point P is

[1]Author to whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received February 25, 2005; final manuscript received March 24, 2006. Review conducted by M. P. Mignolet. 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.

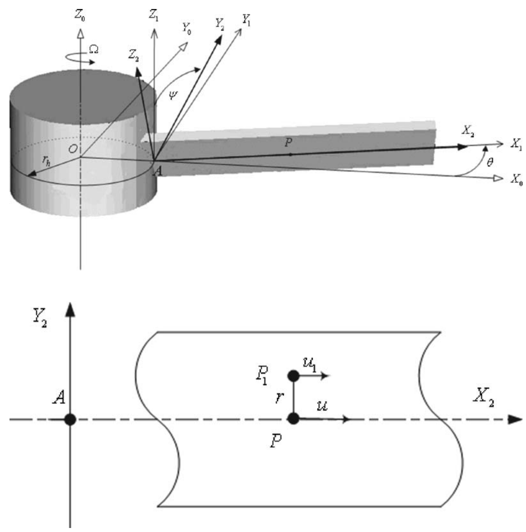


Fig. 1 Geometry and coordinate system of a rotating inclined beam

$$\vec{v_p} = \left\{ \frac{du}{dt}\cos\theta - \frac{dv}{dt}\sin\theta\sin\psi - \Omega[(x+u)\sin\theta + v\cos\theta\sin\psi] \right\}\vec{i} + \left\{ \frac{du}{dt}\sin\theta + \frac{dv}{dt}\cos\theta\sin\psi + \Omega[r_h + (x+u)\cos\theta - v\sin\theta\sin\psi] \right\}\vec{j} + \frac{dv}{dt}\cos\psi\vec{k} \quad (2)$$

and the kinetic energy T of the rotating beam can be expressed as

$$T = \frac{1}{2}\int_0^L \rho A(\vec{v_p}\cdot\vec{v_p})dx \quad (3)$$

where ρ, A, and L are the mass per unit volume, the cross sectional area, and the length of the beam, respectively.

The displacement for arbitrary point P_1 in the cross section of the rotating beam is

$$u_1 = u - r\frac{\partial v}{\partial x} \quad (4)$$

where r is the lateral distance of point P_1 to the centroidal axis. Based on the Bernoulli-Euler beam theory, only the normal strain ε_x is considered. The nonlinear strain-displacement relation yields

$$\varepsilon_x = \frac{\partial u}{\partial x} - r\frac{\partial^2 v}{\partial x^2} + \frac{1}{2}\left(\frac{\partial v}{\partial x}\right)^2 \quad (5)$$

The potential energy U of the rotating beam is

$$U = \frac{1}{2}\int\int E\varepsilon_x{}^2\, dA\, dx, \quad (6)$$

where E is the Young's modulus of the beam.

By utilizing the Hamilton principle and the consistent linearization of the fully nonlinear beam theory, two coupled governing differential equations for the rotating inclined beam are derived as

$$EA\frac{\partial^2 u}{\partial x^2} + \rho A\Omega^2 u - \rho A\frac{\partial^2 u}{\partial t^2} + 2\rho A\Omega\sin\psi\frac{\partial v}{\partial t} = -\rho A\Omega^2(x + r_h\cos\theta) \quad (7a)$$

$$EI\frac{\partial^4 v}{\partial x^4} - \frac{\partial}{\partial x}\left(N_p\frac{\partial v}{\partial x}\right) - \rho A\Omega^2 v\sin^2\psi + \rho A\frac{\partial^2 v}{\partial t^2} + 2\rho A\Omega\sin\psi\frac{\partial u}{\partial t} = -\rho A\Omega^2 r_h\sin\theta\sin\psi \quad (7b)$$

where I is the moment of inertia of the beam. The last terms on the left hand side of Eq. (7a) and (7b) are the Coriolis forces in the axial and the transverse directions, respectively. The associated boundary conditions are as follows:

$$\text{at } x = 0: \quad v = 0 \quad (8a)$$

$$\frac{\partial v}{\partial x} = \tan\theta \quad (8b)$$

$$u = 0 \quad (8c)$$

and

$$\text{at } x = L: \quad EI\frac{\partial^3 v}{\partial x^3} = 0 \quad (9a)$$

$$EI\frac{\partial^2 v}{\partial x^2} = 0 \quad (9b)$$

$$EA\frac{\partial u}{\partial x} = 0 \quad (9c)$$

where N_p is a centrifugal stiffened force and used to be considered as the steady state normal force $N_p = EA(du/dx)$ [14,18]. The second term in Eq. (7b) is a nonlinear term induced from the nonlinear strain displacement relation (5). The steady state normal force N_p is determined in the following.

For steady state deformations, the differential equation (7a) is reduced to

$$EA\frac{d^2 u}{dx^2} + \rho A\Omega^2 u = -\rho A\Omega^2(x + r_h\cos\theta) \quad (10)$$

The solution that satisfies the associated boundary conditions (8c) and (9c) is

$$u(x) = r_h\cos\theta\cdot\cos\frac{\lambda}{L}x + \left(\frac{L}{\lambda\cos\lambda} + r_h\cos\theta\cdot\tan\lambda\right)\sin\frac{\lambda}{L}x - (x + r_h\cos\theta) \quad (11)$$

where $\lambda = \sqrt{\rho/E}\Omega L$ is a dimensionless quantity. As a result,

$$N_p = EA\frac{du}{dx} = EA\left[-\frac{\lambda}{L}r_h\cos\theta\cdot\sin\frac{\lambda}{L}x + \left(\frac{1}{\cos\lambda} + \frac{\lambda r_h\cos\theta\cdot\tan\lambda}{L}\right)\cos\frac{\lambda}{L}x - 1\right] \quad (12)$$

If λ is relatively small and one can retain two terms in the power series approximation of the above trigonometric functions, it is reduced to

$$N_p = \rho A\Omega^2\left[r_h(L-x)\cos\theta + \frac{1}{2}(L^2 - x^2)\right] \quad (13)$$

If the inclination angle θ is zero, the governing differential equations (7a) and (7b) and N_p will be the same as those given by Lin [19], after ignoring the shear deformation and the rotary inertia effects.

3 Two Subsystems

The vibration system, Eqs. (7)–(9), can be considered as the superposition of a static subsystem and a dynamic subsystem by assuming the displacements as

$$u = u_s(s) + u_d(s,t)$$

$$v = v_s(s) + v_d(s,t) \quad (14)$$

where u_s and v_s are the centroidal static displacements in the axial and the transverse directions, respectively. They are time independent variables. u_d and v_d are the centroidal dynamic displacements. in the axial and the transverse directions, respectively.

3.1 Static Subsystem. The static subsystem contains two uncoupled subsystems in terms of the transverse and the axial displacements, respectively. The governing equation for the transverse static displacement is

$$EI\frac{d^4v_s}{dx^4} - \frac{d}{dx}\left(N_p\frac{dv_s}{dx}\right) - \rho A\Omega^2(\sin^2\psi)v_s = -\rho A\Omega^2 r_h \sin\theta\cdot\sin\psi \quad (15)$$

and the associated boundary equations are

$$\text{at } x=0\text{: } v_s = 0 \quad (16a)$$

$$\frac{dv_s}{dx} = \tan\theta \quad (16b)$$

and

$$\text{at } x=L\text{:}\quad EI\frac{d^3v_s}{dx^3} = 0 \quad (17a)$$

$$EI\frac{d^2v_s}{dx^2} = 0 \quad (17b)$$

The governing equation for the axial static displacement is

$$EA\frac{d^2u_s}{dx^2} + \rho A\Omega^2 u_s = -\rho A\Omega^2(x + r_h\cos\theta) \quad (18)$$

and the associated boundary equations are:

$$\text{at } x=0\text{:}\quad u_s = 0 \quad (19)$$

and

$$\text{at } x=L\text{:}\quad EA\frac{du_s}{dx} = 0 \quad (20)$$

This subsystem will have no influence on the free vibration of the beam system.

Dynamic Subsystem. The free vibration of the beam system is governed by dynamic subsystem. The governing differential equations for the sub-system are two coupled differential equations

$$EA\frac{\partial^2u_d}{\partial x^2} + \rho A\Omega^2 u_d - \rho A\frac{\partial^2u_d}{\partial t^2} + 2\rho A\Omega\sin\psi\frac{\partial v_d}{\partial t} = 0 \quad (21a)$$

$$EI\frac{\partial^4v_d}{\partial x^4} - \frac{\partial}{\partial x}\left(N_p\frac{\partial v_d}{\partial x}\right) - \rho A\Omega^2 v_d\sin^2\psi + \rho A\frac{\partial^2v_d}{\partial t^2}$$
$$+ 2\rho A\Omega\sin\psi\frac{\partial u_d}{\partial t} = 0 \quad (21b)$$

The associated dynamic boundary conditions are

$$\text{at } x=0\text{:}\quad v_d = 0 \quad (22a)$$

$$\frac{\partial v_d}{\partial x} = 0 \quad (22b)$$

$$u_d = 0 \quad (22c)$$

and

$$\text{at } x=L\text{: } EI\frac{\partial^3v_d}{\partial x^3} = 0 \quad (23a)$$

$$EI\frac{\partial^2v_d}{\partial x^2} = 0 \quad (23b)$$

$$EA\frac{\partial u_d}{\partial x} = 0 \quad (23c)$$

It is easy to see that when the Coriolis force is not considered, the two governing differential equations of the system are independent.

4 Uncoupled Governing Characteristic Differential Equations

For time-harmonic vibration of a rotating inclined beam with angular frequency ω, one assumes

$$v_d(x,t) = \tilde{V}(x)e^{i\omega t} \quad\text{and}\quad u_d(x,t) = \tilde{U}(x)e^{i\omega t} \quad (24)$$

In terms of the following dimensionless parameters

$$\bar{V} = \frac{\tilde{V}}{L},\quad \bar{U} = \frac{\tilde{U}}{L},\quad \Lambda = \sqrt{\frac{\rho A}{EI}}\omega L^2$$

$$L_z = L\sqrt{\frac{A}{I}},\quad \mu = \frac{r_h}{L},\quad \xi = \frac{x}{L} \quad (25)$$

$$\alpha = L_z\lambda = \sqrt{\frac{\rho A}{EI}}\Omega L^2,\quad \overline{N_p} = \frac{N_p}{\rho A\Omega^2 L^2}$$

the governing characteristic differential equations Eq. (21a) and (21b) can be expressed as

$$\bar{V}^{(4)} - \alpha^2\left(\overline{N_p}\bar{V}'\right)' - (\Lambda^2 + \alpha^2\sin^2\psi)\bar{V} + 2i\Lambda\alpha(\sin\psi)\bar{U} = 0 \quad (26a)$$

$$\bar{U}'' + \frac{(\Lambda^2 + \alpha^2)}{L_z^2}\bar{U} + \frac{2i\Lambda\alpha\sin\psi}{L_z^2}\bar{V} = 0 \quad (26b)$$

where the prime is the derivative with respect to the dimensionless variable ξ and the dimensionless axial centrifugal force is

$$\overline{N_p} = C_a\sin\frac{\alpha}{L_z}\xi + C_b\cos\frac{\alpha}{L_z}\xi - \frac{L_z^2}{\alpha^2} \quad (27)$$

In the above expression, C_a and C_b are, respectively,

$$C_a = -\frac{L_z}{\alpha}\mu\cos\theta$$

$$C_b = \frac{L_z^2}{\alpha^2\cos\dfrac{\alpha}{L_z}} + \frac{L_z}{\alpha}\left(\tan\frac{\alpha}{L_z}\right)\mu\cos\theta \quad (28)$$

When $\lambda = \alpha/L_z$ is relatively small, $\overline{N_p}$ can be reduced to

$$\overline{N_p} = \mu(1-\xi)\cos\theta + \frac{1}{2}(1-\xi^2) \quad (29)$$

The associated dimensionless boundary conditions are

$$\text{at } \xi=0\text{: } \bar{V} = 0 \quad (30a)$$

$$\bar{V}' = 0 \quad (30b)$$

$$\bar{U} = 0 \quad (30c)$$

and

$$\text{at } \xi=1\text{:}\quad \bar{V}''' = 0 \quad (31a)$$

$$\bar{V}'' = 0 \quad (31b)$$

$$\bar{U}' = 0 \tag{31c}$$

By simple arithmetic operation, the two coupled governing equations (26*a*) and (26*b*) can be decoupled and expressed as a six-order ordinary differential equation in terms of $\bar{V}$

$$\bar{V}^{(6)} + a_4\bar{V}^{(4)} + a_3\bar{V}''' + a_2\bar{V}'' + a_1\bar{V}' + a_0\bar{V} = 0 \tag{32}$$

where

$$a_0 = \frac{(\Lambda^2+\alpha^2)(\Lambda^2+\alpha^2\sin^2\psi) + 4\Lambda^2\alpha^2\sin^2\psi}{L_z^2}$$

$$a_1 = \frac{(\Lambda^2+\alpha^2)\alpha^2}{L_z^2}\overline{N_p}' - \alpha^2\overline{N_p}'''$$

$$a_2 = -(\Lambda^2+\alpha^2\sin^2\psi) - 3\alpha^2\overline{N_p}'' + \frac{(\Lambda^2+\alpha^2)\alpha^2}{L_z^2}\overline{N_p} \tag{33}$$

$$a_3 = -3\alpha^2\overline{N_p}'$$

$$a_4 = -\alpha^2\overline{N_p} - \frac{(\Lambda^2+\alpha^2)}{L_z^2}$$

The associated boundary conditions are

$$\text{at } \xi=0: \quad \bar{V} = 0 \tag{34a}$$

$$\bar{V}' = 0 \tag{34b}$$

$$\bar{V}^{(4)} - \alpha^2\overline{N_p}\bar{V}'' - \alpha^2\overline{N_p}'\bar{V}' - (\Lambda^2+\alpha^2\sin^2\psi)\bar{V} = 0 \tag{34c}$$

and

$$\text{at } \xi=1: \ \bar{V}'' = 0 \tag{35a}$$

$$\bar{V}''' = 0 \tag{35b}$$

$$\bar{V}^{(5)} - \alpha^2\overline{N_p}\bar{V}''' - 2\alpha^2\overline{N_p}'\bar{V}'' - (\Lambda^2+\alpha^2\sin^2\psi + \alpha^2\overline{N_p}'')\bar{V}' = 0 \tag{35c}$$

5 Fundamental Solutions and Frequency Equation

The decoupled governing differential equation (32) is a six-order ordinary differential equation with variable coefficients. In general, the exact fundamental solutions are not available. However, if the coefficients of the differential equation can be expressed in the following polynomial form

$$a_0 = \sum_{i=0}^{n0} d_i\xi^i, \quad a_1 = \sum_{i=0}^{n1} e_i\xi^i, \quad a_2 = \sum_{i=0}^{n2} f_i\xi^i, \quad a_3 = \sum_{i=0}^{n3} g_i\xi^i,$$

$$a_4 = \sum_{i=0}^{n4} h_i\xi^i \tag{36}$$

where upper terms in the summations, $n0$, $n1$, $n2$, $n3$, and $n4$, denote the order of the polynomials $a_i(\xi), i=0-4$. Then the six linearly independent fundamental solutions, $w_i(\xi), i=1-6$, of the differential equation (32), which satisfy the following normalization condition at the origin of the coordinate system

$$\begin{bmatrix} w_1(0) & w_2(0) & w_3(0) & w_4(0) & w_5(0) & w_6(0) \\ w_1'(0) & w_2'(0) & w_3'(0) & w_4'(0) & w_5'(0) & w_6'(0) \\ w_1''(0) & w_2''(0) & w_3''(0) & w_4''(0) & w_5''(0) & w_6''(0) \\ w_1'''(0) & w_2'''(0) & w_3'''(0) & w_4'''(0) & w_5'''(0) & w_6'''(0) \\ w_1^{(4)}(0) & w_2^{(4)}(0) & w_3^{(4)}(0) & w_4^{(4)}(0) & w_5^{(4)}(0) & w_6^{(4)}(0) \\ w_1^{(5)}(0) & w_2^{(5)}(0) & w_3^{(5)}(0) & w_4^{(5)}(0) & w_5^{(5)}(0) & w_6^{(5)}(0) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{37}$$

can be obtained by using the method of Frobenius and extending the works done by Stafford and Giurgiutiu [20,21] and Lee and Kuo [22]. Otherwise, the approximated solutions can be obtained by following the algorithm developed by Lee [23].

The six independent solutions are assumed to be in the form of

$$w_i(\xi) = \sum_{n=0}^{\infty} k_{i,n}\xi^n, \ i=1,2,\ldots,6 \tag{38a}$$

where

$$\begin{aligned} &\text{for } w_1(\xi): k_{1,0}=1, k_{1,1}=k_{1,2}=k_{1,3}=k_{1,4}=k_{1,5}=0 \\ &\text{for } w_2(\xi): k_{2,1}=1, k_{2,0}=k_{2,2}=k_{2,3}=k_{2,4}=k_{2,5}=0 \\ &\text{for } w_3(\xi): k_{3,2}=1/2, k_{3,0}=k_{3,1}=k_{3,3}=k_{3,4}=k_{3,5}=0 \\ &\text{for } w_4(\xi): k_{4,3}=1/6, k_{4,0}=k_{4,1}=k_{4,2}=k_{4,4}=k_{4,5}=0 \\ &\text{for } w_5(\xi): k_{5,4}=1/24, k_{5,0}=k_{5,1}=k_{5,2}=k_{5,3}=k_{5,5}=0 \\ &\text{for } w_6(\xi): k_{6,5}=1/120, k_{6,0}=k_{6,1}=k_{6,2}=k_{6,3}=k_{6,4}=0 \end{aligned} \tag{38b}$$

Upon substituting Eqs. (36) and (38) into Eq. (32) and collecting the coefficients of like powers of ξ, the following recurrence formula can be obtained as

$$\begin{aligned} k_{i,m+6} = \frac{-1}{(m+6)(m+5)\cdots(m+1)} &\times \Bigg\{ \left[\sum_{j=1}^{m} (m-j+6)\cdots(m-j+1)k_{i,m-j+6} \right] \\ &+ \left[\sum_{j=0}^{m} (m-j+4)\cdots(m-j+1)h_j k_{i,m-i+4} \right] \\ &+ \left[\sum_{j=0}^{m} (m-j+3)\cdots(m-j+1)g_j k_{i,m-j+3} \right] \\ &+ \left[\sum_{j=0}^{m} (m-j+2)(m-j+1)f_j k_{i,m-j+2} \right] \\ &+ \left[\sum_{j=0}^{m} (m-j+1)e_j k_{i,m-j+1} \right] + \left[\sum_{j=0}^{m} d_j k_{i,m-j} \right] \Bigg\} \end{aligned} \tag{39}$$

$$m=0,1,\cdots\to\infty$$

With this recurrence formula, the six exact normalized fundamental solutions (38) can be generated. Therefore, the general solution of the system is

$$\bar{V}(\xi)=\sum_{i=1}^{6}c_i w_i(\xi) \tag{40}$$

where $\{c_i\}$ are the constants to be determined.

Substituting the general solution into the associated boundary conditions, it yields a set of equations

$$[B_{ij}]\{c_i\}=0,\ i,j=1:6 \tag{41}$$

where

$$B_{ij}=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -(\Lambda^2+\alpha^2\sin^2\psi) & -\alpha^2\overline{N_p}' & -\alpha^2\overline{N_p} & 0 & 1 & 0 \\ w_1''(1) & w_2''(1) & w_3''(1) & w_4''(1) & w_5''(1) & w_6''(1) \\ w_1'''(1) & w_2'''(1) & w_3'''(1) & w_4'''(1) & w_5'''(1) & w_6'''(1) \\ s_1 & s_2 & s_3 & s_4 & s_5 & s_6 \end{bmatrix}, \tag{42}$$

in which $s_i, i=1-6$ are

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \end{bmatrix}=\begin{bmatrix} w_1'(1) & w_1''(1) & w_1'''(1) & w_1^{(5)}(1) \\ w_2'(1) & w_2''(1) & w_2'''(1) & w_2^{(5)}(1) \\ w_3'(1) & w_3''(1) & w_3'''(1) & w_3^{(5)}(1) \\ w_4'(1) & w_4''(1) & w_4'''(1) & w_4^{(5)}(1) \\ w_5'(1) & w_5''(1) & w_5'''(1) & w_5^{(5)}(1) \\ w_6'(1) & w_6''(1) & w_6'''(1) & w_6^{(5)}(1) \end{bmatrix}$$

$$\times\begin{bmatrix} -(\Lambda^2+\alpha^2\sin^2\psi+\alpha^2\overline{N_p}'') \\ -2\alpha^2\overline{N_p}' \\ -\alpha^2\overline{N_p} \\ 1 \end{bmatrix} \tag{43}$$

As a result, the natural frequencies of the rotating inclined beam can be obtained from the following frequency equation,

$$|B_{ij}|=0 \tag{44}$$

It should be mentioned that the proposed power series solutions converge when the radius of convergence is less than/equal to 1. In this paper, the problem is described in terms of dimensionless variables. The dimensionless variable ξ used to express the fundamental solutions and the coefficients of the governing differential equation is confined within the interval [0, 1]. In addition, the governing differential equation is a regular differential equation. The coefficients of the differential equation are analytic functions. All the physical quantities are finite at the boundaries. Therefore, the solutions converge. The accuracy of the numerical results will be checked and compared with those in the existing literature in the later section.

6 Frequency Relations

The natural frequencies of the system can be numerically determined by the method revealed in the previous section. However, most of the numerical results can only provide partial qualitative conclusions. In addition, it requires a tremendous computer calculation. In this section, several qualitative relations are explored without a numerical analysis.

6.1 Frequency Relations For the Systems With Different Inclination Angle and Hub Radius. Consider two dynamic systems with the same physical parameters, except the inclination angle θ and the hub radius r_h. In terms of dimensionless quantities, they are θ and μ. To specify two different systems, subscripts "a" and "b" are added to the associated physical parameters.

It can observed that if

$$\mu_a\cos\theta_a=\mu_a\cos\theta_b \tag{45}$$

then the governing characteristic differential equations (26)–(28) and the associated boundary conditions, Eqs. (30) and (31), of two dynamic systems will be the same. Therefore, the fundamental solutions and the natural frequencies of two systems will be the same. In addition, it is well known that if one increases the hub radius of the rotating beam, the induced centrifugal force will be increased. Consequently, the natural frequencies are increased [13].

Based on the two facts, the following two conclusions can be drawn.

(1) For two dynamic systems with the same physical parameters except the inclination angle and the hub radius, if $\mu_a\cos\theta_a=\mu_b\cos\theta_b$, the natural frequencies of two systems will be the same.

(2) If the hub radius is not zero and the inclination angle of the dynamic system is increased, the natural frequencies decrease.

It should be mentioned that the two conclusions are valid for both systems with and without considering the Coriolis force effect.

6.2 Frequency Relations For the Systems Without Coriolis Force Effect. If the Coriolis force is not considered, the flexural and the longitudinal motions of the beam are independent. The

Table 1 Dimensionless fundamental flexural natural frequencies of a rotating beam (L_z=70, θ=0, ψ=90)

μ	α	Present[a]	Present[b]	Ref. [9][b]	Ref. [13][b]
0	2	3.6220	3.6195	3.6196	3.6196
	10	5.1362	4.9588	4.9700	4.9703
1	2	4.4009	4.3977	4.3978	4.3978
	10	13.342	13.044	13.048	13.049
5	2	6.6481	6.6429	6.6430	6.6430
	10	27.862	27.306	27.266	27.276

[a]Without the Coriolis force effect.
[b]With the Coriolis force effect.

Table 2 First two natural frequencies of rotating beams with different μ and θ, those satisfying the frequency relation $\mu_a \cos \theta_a = \mu_b \cos \theta_b (\psi = 90°, L_z = 70)$

		$\mu=1, \cos\theta=1$		$\mu=2, \cos\theta=1/2$	
α	Λ	With Coriolis force effect	Without Coriolis force effect	With Coriolis force effect	Without Coriolis force effect
0.5	Λ_1	3.5780	3.5782	3.5780	3.5782
	Λ_2	22.1143	22.1145	22.1143	22.1145
1.0	Λ_1	3.7575	3.7581	3.7575	3.7581
	Λ_2	22.3521	22.3527	22.3521	22.3527
1.5	Λ_1	4.0376	4.0392	4.0376	4.0392
	Λ_2	22.7427	22.7441	22.7427	22.7441
2.0	Λ_1	4.3977	4.4009	4.3977	4.4009
	Λ_2	23.2784	23.2807	23.2784	23.2807
3.0	Λ_1	5.2826	5.2921	5.2826	5.2921
	Λ_2	24.7433	24.7487	24.7433	24.7487

governing characteristic differential equation for the flexural motion can be reduced from Eq. (26*a*) by taking off the Coriolis force term. If all the physical parameters of two different systems are the same except the setting angle and the natural frequencies, it can be observed that if the relation

$$\alpha^2 \sin^2 \psi_a + \Lambda_a^2 = \alpha^2 \sin^2 \psi_b + \Lambda_b^2 \quad (46)$$

exists, then the reduced governing characteristic differential equations and the associated boundary conditions of two dynamic systems will be the same. This relation reveals the following.

(1) If all the physical parameters of two different vibration systems are the same except the setting angle, when the setting angle is less than 90°, the larger the setting angle is, the smaller the natural frequencies are.
(2) The influence of the setting angle on the flexural natural frequencies of a beam rotating at high speed is greater than that of a beam rotating at low speed.

The frequency relations revealed from Eqs. (45) and (46) will also be verified numerically in the next section.

The governing characteristic differential equation for the longitudinal motion can be reduced from Eq. (26*b*) by taking off the last term, the Coriolis force term, on the left hand side of the equation. It can be observed that if the following relation exists, the reduced governing characteristic differential equations and the associated boundary conditions of two dynamic systems are the same

$$\frac{(\Lambda_a^2 + \alpha_a^2)}{{L_{z,a}}^2} = \frac{(\Lambda_b^2 + \alpha_b^2)}{{L_{z,b}}^2} \quad (47)$$

or

$$(\omega^2 + \Omega^2)\frac{\rho L^2}{E} = \text{const} \quad (48)$$

This relation reveals the following.

(1) For the longitudinal vibration of rotating beams with constant material and geometric properties, $L_{z,a}=L_{z,b}$, the natural frequencies will decrease if the rotating speed is increased. This conclusion is converse to that for the flexural vibration of rotating beams.
(2) At a constant rotating speed, the longitudinal natural frequencies decrease as the length of the beam is increased.

The relation (47) can also be obtained from the eigenfunction solution of the dynamic system. The eigenfunctions that satisfy the differential equation (26*b*) and the associated boundary conditions (30*c*) and (31*c*) are

$$\bar{U} = \sin k_n \xi \quad (49)$$

where $k_n = \sqrt{(\Lambda_n^2 + \alpha^2)/L_z^2} = n\pi/2$, $n=1,3,5,$ L are the wavenumbers. It can be found that the results are consistent.

7 Numerical Results

To illustrate the previous analysis and the accuracy of the numerical analysis, several numerical results are presented and discussed.

Table 3 First three natural frequencies of rotating beams with different ψ and Λ, those satisfying the frequency relation $\alpha^2 \sin^2 \psi_a + \Lambda_a^2 = \alpha^2 \sin^2 \psi_b + \Lambda_b^2$ $(\theta=0, \mu=1, L_z=70)$

				$\psi_b=90°$			
		$\psi_a=30°$		Present		Prediction via Eq. (46)	
α	Λ	With Coriolis force effect	Without Coriolis force effect	With Coriolis force effect	Without Coriolis force effect	With Coriolis force effect	Without Coriolis force effect
1	Λ_1	3.8564	3.8566	3.7575	3.7581	3.7579	3.7581
	Λ_2	22.3693	22.3695	22.3521	22.3527	22.3525	22.3527
	Λ_3	62.0409	62.0411	62.0345	62.0350	62.0349	62.0351
2	Λ_1	4.7283	4.7295	4.3977	4.4009	4.3996	4.4009
	Λ_2	23.3445	23.3451	23.2784	23.2807	23.2802	23.2808
	Λ_3	63.0596	63.0601	63.0343	63.0363	63.0358	63.0363
3	Λ_1	5.8915	5.8955	5.2826	5.2921	5.2877	5.2922
	Λ_2	24.8835	24.8847	24.7433	24.7487	24.7475	24.7487
	Λ_3	64.7175	64.7187	64.6619	64.6665	64.6653	64.6665

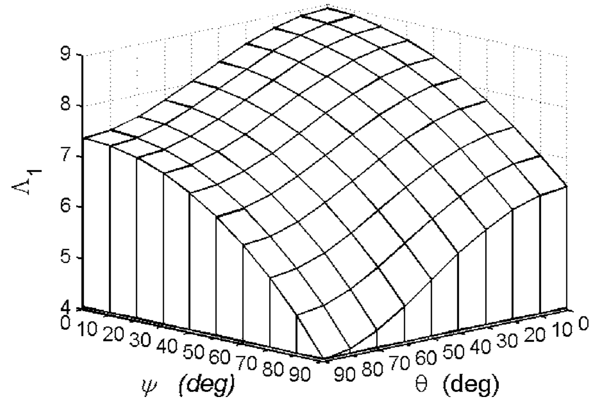


Fig. 2 Influence of the setting angle ψ and the inclination angle θ on the fundamental natural frequencies of a rotating cantilever beam with hub radius ($\mu=1,\alpha=5,L_z=70$)

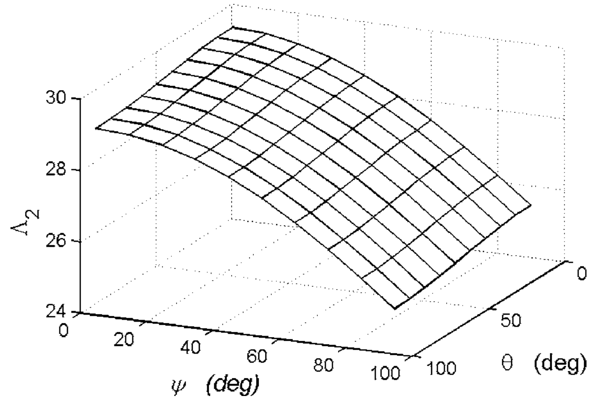


Fig. 4 Influence of setting angle ψ and inclination angle θ on the second natural frequencies of a rotating cantilever beam with hub radius ($\mu=1,\alpha=5,L_z=70$)

In Table 1, the fundamental flexural natural frequencies of the rotating beam without inclination angle, obtained by the present numerical method, are compared with those in the existing literature. It can be observed that the numerical results are in good agreement.

The numbers in the third and fourth columns of Table 2 are the first two natural frequencies, determined by the present numerical method, of system "a," considering and without considering the Coriolis force effect, respectively. The numbers in the fifth and the sixth columns are the corresponding natural frequencies, determined by the present numerical method, of system "b." The physical parameters of these two systems satisfy the frequency relation, Eq. (45). It can be found that the two sets of natural frequencies are all the same. This verifies the first conclusion revealed from the frequency relation, Eq. (45).

Table 3 is given to illustrate the fourth conclusion revealed from the frequency relation, Eq. (46). The numbers in the third, fourth, fifth, and sixth columns of Table 3 are the first three natural frequencies, determined by the present numerical method, of two different systems considering and without considering the Coriolis force effect, respectively. The seventh and eighth columns are the corresponding natural frequencies, determined directly by using the frequency relation, Eq. (46), based on the data given in the third and the fourth columns. It can be found that the last two sets of natural frequencies are completely consistent.

From Tables 1–3, it can also be found that the fundamental natural frequencies of the rotating beam considering the Coriolis force effect are higher than those of the beam without considering the Coriolis force effect.

Figures 2 and 3 show the influence of the setting angle and the inclination angle on the fundamental natural frequencies of rotating cantilever beams with and without a hub radius, respectively. From Fig. 2, it can be observed that when the hub radius is not zero and the inclination angle of the dynamic system is increased, the natural frequencies decrease. The numerical results are consistent with the second conclusion revealed from the frequency relation, Eq. (45).

From Fig. 3, it can also be found that when the hub radius is zero, the inclination angle will have no influence on the natural frequencies of the dynamic system. It is due to the fact that when $\mu=0$, the coefficients in the governing characteristic differential equation (32) will not contain the parameter of inclination angle.

Both Figs. 2 and 3 also show that the natural frequencies decrease as the setting angle is increased. This result is the same as that for the rotating beam without inclination angle [17].

Both Figs. 4 and 5 show that the influences of the setting angle and the inclination angle on the second natural frequencies of rotating cantilever beams with and without the hub radius are the same as those on the fundamental natural frequencies of the rotating beams. As a matter of fact, the two conclusions revealed from the frequency relation, Eq. (45), are valid for all the natural frequencies.

Figures 6 and 7 show that the influence of the slenderness ratio L_z, the dimensionless rotating speed α, and the inclination angle θ on the fundamental natural frequencies of rotating cantilever beams with and without considering the Coriolis force effect, respectively. The numerical results show that the Coriolis force ef-

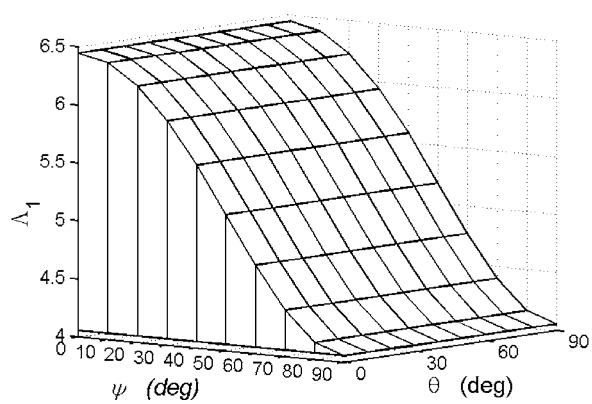


Fig. 3 Influence of setting angle ψ and inclination angle θ on the fundamental natural frequencies of a rotating cantilever beam without hub radius ($\mu=0,\alpha=5,L_z=70$)

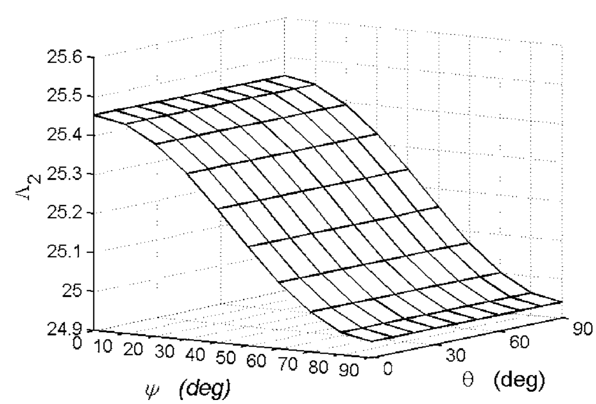


Fig. 5 Influence of setting angle ψ and inclination angle θ on the second natural frequencies of a rotating cantilever beam without hub radius ($\mu=0,\alpha=7,L_z=70$)

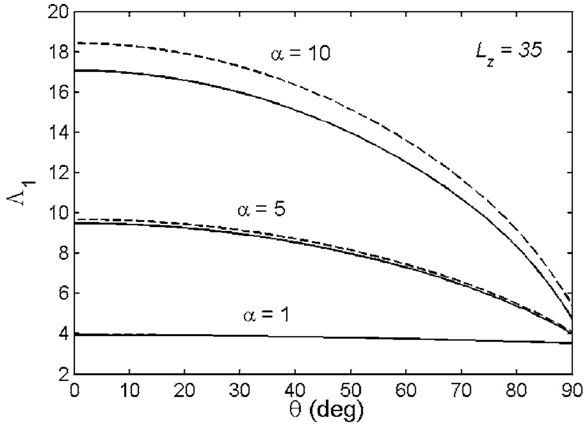


Fig. 6 Influence of the rotating speed and the inclination angle θ on the fundamental natural frequencies of a rotating cantilever beam with a hub radius ($\mu=2, \psi=90°, L_z=35$; dashed lines: the Coriolis force effect ignored; real line: the Coriolis force effect considered)

fect on the natural frequencies will increase when the slenderness ratio is decreased and the rotating speed is increased.

8 Conclusions

By utilizing the Hamilton principle and the consistent linearization of the fully nonlinear beam theory, two coupled governing differential equations for a rotating inclined beam are successfully derived. Both the extensional deformation and the Coriolis force effect are considered. The exact series solution of the system is developed. The frequency relations and analytical numerical results reveal the following general qualitative conclusions.

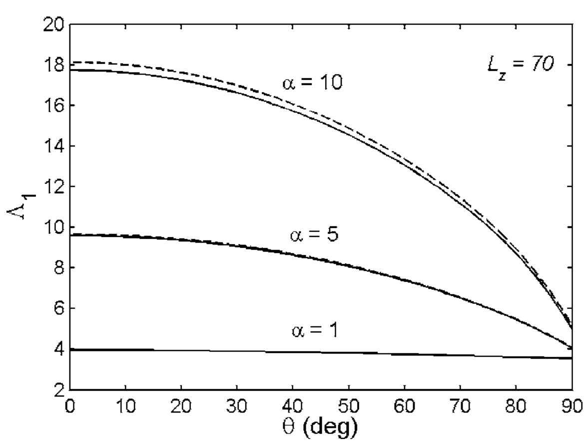


Fig. 7 Influence of the rotating speed and the inclination angle θ on the fundamental natural frequencies of a rotating cantilever beam with hub radius ($\mu=2, \psi=90°, L_z=70$; dashed lines: the Coriolis force effect ignored; real line: the Coriolis force effect considered)

(1) For two dynamic systems with the same physical parameters except the inclination angle and the hub radius, if $\mu_a \cos\theta_a = \mu_b \cos\theta_b$, the natural frequencies of two systems will be the same.
(2) If the hub radius is not zero and the inclination angle of the dynamic system is increased, the natural frequencies decrease.
(3) When the hub radius is zero, the inclination angle will have no influence on the natural frequencies of the dynamic system.
(4) If the Coriolis force effect is ignored and all the physical parameters of two different flexural vibration systems are the same except the setting angle, when the setting angle is less than 90°, the larger the setting angle, the smaller the natural frequencies.
(5) If the Coriolis force effect is ignored, the influence of the setting angle on the flexural natural frequencies of a beam rotating at high speed is greater than that of a beam rotating at low speed.
(6) For the longitudinal vibration of rotating beams with constant material and geometric properties, $L_{z,a}=L_{z,b}$, the natural frequencies will decrease if the rotating speed is increased. This conclusion is converse to that for the flexural vibration of rotating beams.
(7) At a constant rotating speed, the longitudinal natural frequencies decrease, as the length of the beam is increased.
(8) The Coriolis force effect on the natural frequencies will increase when the slenderness ratio is decreased and the rotating speed is increased.

Acknowledgment

This research work was supported by the National Science Council of Taiwan, Republic of China, under Grant No. NSC92-2751-B-006-012 and is gratefully acknowledged.

References

[1] Leissa, A., 1981, "Vibrational Aspects of Rotating Turbomachinery Blades," Appl. Mech. Rev., **34**, pp. 629–635.
[2] Ramamurti, V., and Balasubramanian, P., 1984, "Analysis of Turbomachine Blades: A Review," Shock Vib. Dig., **16**, pp. 13–28.
[3] Rao, J. S., 1987, "Turbomachine Blade Vibration," Shock Vib. Dig., **19**, pp. 3–10.
[4] Young, T. H., and Lin, T. M., 1998, "Stability of Rotating Pretwisted, Tapered Beams With Randomly Varying Speeds," ASME J. Vibr. Acoust., **120**, pp. 784–790.
[5] Rao, J. S., and Carnegie, W., 1972, "Nonlinear Vibrations of Rotating Cantilever Blades Treated by the Ritz Averaging Process," Aeronaut. J., **76**, pp. 566–569.
[6] Ko, C. L., 1989, "Dynamic Analysis for Free Vibrations of Rotating Sandwich Tapered Beams," AIAA J., **27**, pp. 1425–1433.
[7] Boyce, W. E., 1956, "Effect of Hub Radius on the Vibrations of a Uniform Bar," J. Appl. Mech., **23**, pp. 287–290.
[8] Lin, S. M., Wu, C. T., and Lee, S. Y., 2003, "Analysis of Rotating Nonuniform Pretwisted Beams With an Elastically Restrained Root and a Tip Mass," Int. J. Mech. Sci., **45**, pp. 741–755.
[9] Chung, J., and Yoo, H. H., 2002, "Dynamic Analysis of a Rotating Cantilever Beam by Using the Finite Element Method," J. Sound Vib., **249**, pp. 147–164.
[10] Fung, E. H. K., and Yau, D. T. W., 1999, "Effects of Centrifugal Stiffening on the Vibration Frequencies of a Constrained Flexible Arm," J. Sound Vib., **224**, pp. 808–841.
[11] Banerjee, J. R., 2000, "Free Vibration of Centrifugally Stiffened Uniform and Tapered Beams Using the Dynamic Stiffness Method," J. Sound Vib., **233**, pp. 857–875.
[12] Park, J. H., and Kim, J. H., 1999, "Dynamic Analysis of Rotating Curved Beam with a Tip Mass," J. Sound Vib., **228**, pp. 1017–1034.
[13] Yoo, H. H., and Shin, S. H., 1998, "Vibration Analysis of Rotating Cantilever Beams," J. Sound Vib., **212**, pp. 807–828.
[14] Rao, S. S., and Gupta, R. S., 2001, "Finite Element Analysis of Rotating Timoshenko Beams," J. Sound Vib., **242**, pp. 103–124.
[15] Schilhansl, M. J., 1958, "Bending Frequency of a Rotating Cantilever Beam," ASME J. Appl. Mech., **25**, pp. 28–30.

[16] Pnueli, D., 1972, "Natural Bending Frequency Comparable to Rotational Frequency in Rotating Cantilever Beam," ASME J. Appl. Mech., **39**, pp. 602–604.
[17] Lee, S. Y., and Kuo, Y. H., 1991, "Bending Frequency of a Rotating Beam With an Elastically Restrained Root," ASME J. Appl. Mech., **58**, pp. 209–214.
[18] Simo, J. C., and Quoc, L. V., 1987, "The Role of Non-Linear Theories in Transient Dynamic Analysis of Flexible Structures," J. Sound Vib., **119**, pp. 487–508.
[19] Lin, S. C., and Hsiao, K. M., 2001, "Vibration Analysis of a Rotating Timoshenko Beam," J. Sound Vib., **240**, pp. 303–322.
[20] Stafford, R. O., and Giurgiutiu, V., 1975, "Semi-Analytic Methods for Rotating Timoshenko Beams," Int. J. Mech. Sci., **17**, pp. 719–727.
[21] Giurgiutiu, V., and Stafford, R. O., 1977, "Semi-Analytic Methods For Frequencies and Mode-Shapes of Rotor Blades," Vertica, **1**, pp. 291–306.
[22] Lee, S. Y., and Kuo, Y. H., 1992, "Exact Solutions for the Analysis of General Elastically Restrained Nonuniform Beams," ASME J. Appl. Mech., **59**, pp. 205–212.
[23] Lee, S. Y., and Kuo, Y. H., 1992, "Bending Vibrations of a Rotating Nonuniform Beam With an Elastically Restrained Root," J. Sound Vib., **154**, pp. 441–451.

On the Vibration Isolation of Flexible Structures

Y. Q. Tu
Department of Civil Engineering,
Beijing University of Aeronautics and Astronautics,
Beijing, China
e-mail: yongqingtu@263.net

G. T. Zheng[1]
School of Aerospace Engineering,
Tsinghua University,
Beijing, China
e-mail: gtzhengtu@yahoo.co.uk

Although the study of vibration isolation has a very long history, when an isolated structure is so flexible that it cannot be properly approximated with a rigid body or a single-degree-of-freedom model, its vibration isolation brings about some new questions and problems. By transforming the dynamic equation of motion of the coupled structure formed by the isolator and the isolated structure into the modal space and following the tradition of studying features of the vibration transmissibility across the isolator, questions and problems associated with the flexible structure vibration isolation are studied. It is found from the study that a lower isolation frequency and a higher damping level can both increase the isolation effectiveness, the isolated structure is a vibration absorber to the isolator, and a combination of the vibration isolation and the vibration attenuation can be more effective in mitigating the vibration. A numerical example of the whole spacecraft vibration isolation has proved the above conclusions.
[DOI: 10.1115/1.2201882]

1 Introduction

Vibration isolation is an effective way of mitigating the vibration of a structure or reducing the vibration transmission of a vibration source to its surroundings. Research into vibration isolation has a very long history, some research results have already been included in many textbooks [1,2], and there are also many books devoted specially to the study of vibration isolation [3]. Applications of the vibration isolation technique have also been from machinery systems, vehicles, and buildings such traditional areas extended to many new areas, for example spacecrafts [4].

Traditionally, for studying the theory, explaining phenomena, etc., the structure being isolated is often simplified as a rigid body or with only one degree of freedom. With these simplified models, the concept of vibration transmissibility has been developed for defining the effectiveness of isolation, and thus the transmissibility becomes the basis of the research from which such concepts and conclusions as isolation frequency and the function of damping are derived. However, when the flexibility of an isolated structure is so high that it cannot be approximated with these simplified models, the question about whether those concepts and conclusions thus derived are still valid or not, should be answered. As a direct consequence, the suitability of the transmissibility for describing the effectiveness of the isolation has been taken as a question in some researches. Because of the coupling between the isolator and the flexible structure, an expression for the transmissibility is so complicated that usually the analysis has to rely on numerical calculations, and hence it usually does not provide explicit relations among dynamic characteristics of the isolated structure and the parameters of the isolator. If there is any analytical solution, the isolated structure should be simple, such as beam or plate [5]. This difficulty not only makes the concepts and conclusions derived based on the simplified models obscure, but sometimes also causes perplexity in practical applications.

An alternative approach of solving this problem is to drop the transmissibility and seek other ways of defining the isolation effectiveness, such as power flow [2]. Nevertheless, mainly for historical reasons, the transmissibility has been accepted widely in nearly all engineering sections and engineers are accustomed to the concept. Furthermore, vibration at the interface of two connected structure sections, i.e., isolator and isolated structure, is meaningful for many engineering applications, such as the whole spacecraft vibration isolation where the vibration at the interface is used as the input to the spacecraft for examining its suitability for launching. More importantly, for the sake of ensuring the completeness and coherence of the vibration isolation theory, it is also required that the transmissibility and consequently those concepts derived from it should be extended to the isolation of flexible structures.

In the present paper, the study will be focused on problems and questions produced by the flexibility of isolated structures. The transmissibility will still be used as the basis for the study. Features of the transmissibility on the amplitude-frequency plane will be related to dynamic characteristics of the isolated structures and parameters of the isolator. With the relations such established, the concept of isolation frequency and the function of damping are reexamined. To facilitate the analysis, dynamic equations of motion will be transformed into the modal space, in which the influence of the flexibility represented by the modal parameters of the isolated structure on the isolation effectiveness can be explicitly expressed with analytical formulas.

2 Dynamic Equations of Motion in Modal Space

The dynamic equation of motion, which is in the form of the finite element model, of the coupled structure formed by an isolated structure on the top of an isolation structure (isolator), for example a flexible satellite on the top of a vibration isolation platform as shown in Fig. 1, is

$$\begin{bmatrix} M_1 & & & \\ & M_2 & & \\ & & M_b & \\ & & & M_3 \end{bmatrix}\begin{Bmatrix} \ddot{X}_1 \\ \ddot{X}_2 \\ \ddot{X}_b \\ \ddot{X}_3 \end{Bmatrix} + \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & C_{bb} & C_{b3} \\ & & C_{3b} & C_{33} \end{bmatrix}\begin{Bmatrix} \dot{X}_1 \\ \dot{X}_2 \\ \dot{X}_b \\ \dot{X}_3 \end{Bmatrix} + \begin{bmatrix} K_{11} & K_{12} & & \\ K_{21} & K_{22} & K_b & \\ & K_b^T & K_{bb} & K_{b3} \\ & & K_{3b} & K_{33} \end{bmatrix}\begin{Bmatrix} X_1 \\ X_2 \\ X_b \\ X_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ C_u\dot{U} + K_u U \end{Bmatrix} \tag{1}$$

[1]Author to whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received May 1, 2005; final manuscript received March 28, 2006. Review conducted by N. Sri Namachchivaya. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

Here the damping of the isolated structure is supposed to be zero for simplifying the study. In Eq. (1), M, C, and K represent the mass, damping, and stiffness matrix, respectively, subscripts 1 and 2 denote the inner and boundary coordinates of the isolated structure, respectively, subscripts 3 and b denote the inner and boundary coordinates of the isolator, respectively. In the equation, K_b is the connection stiffness between the isolated structure and the isolator, K_{22} and K_{bb} also contain the elements of connection stiffness. In the equation, U is the excitation from the bottom of the isolator.

To study the transmissibility, which can be defined in the frequency domain as the function of the frequency, with the Fourier transform, the previous equation is transformed into an equation in the frequency domain as

$$\begin{bmatrix} -\omega^2 M_1 + K_{11} & K_{12} & & \\ K_{21} & -\omega^2 M_2 + K_{22} & K_b & \\ & K_b^T & -\omega^2 M_b + j\omega C_{bb} + K_{bb} & j\omega C_{b3} + K_{b3} \\ & & j\omega C_{3b} + K_{3b} & -\omega^2 M_3 + j\omega C_{33} + K_{33} \end{bmatrix} \begin{Bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_b \\ \bar{X}_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ (j\omega C_u + K_u)\bar{U} \end{Bmatrix} \tag{2}$$

Since the flexibility of a structure means that the structure has many natural frequencies, to discuss the influence of the structural flexibility, the modal transformation will be introduced in the following part of this section. With the modal transformation [8], the transmissibility can be explicitly expressed as a function of the natural frequencies both of the isolated structure and the isolator. Let

$$\begin{Bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{Bmatrix} = \Phi \bar{Q}_s = \begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix} \bar{Q}_s, \tag{3}$$

and Φ is the modal matrix of the isolated structure with which the physical parameter matrices of the structure can be diagonalized, such that

$$D_s = \Phi^T \begin{bmatrix} -\omega^2 M_1 + K_{11} & K_{12} \\ K_{21} & -\omega^2 M_2 + K_{22} \end{bmatrix} \Phi = \underset{n=1,2,\ldots,N}{\mathrm{diag}} \{k_{sn} - \omega^2 m_{sn}\}, \tag{4}$$

Substitute this coordinate transformation into Eq. (2); there is

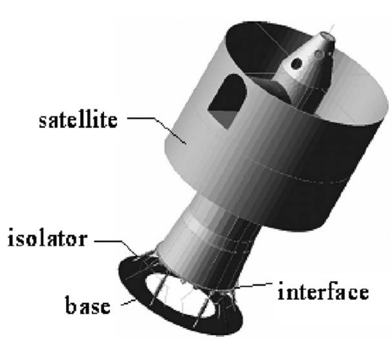


Fig. 1 The finite element model of a satellite on the top of an isolator

$$\begin{bmatrix} D_s & \Phi_2^T K_b & \\ K_b^T \Phi_2 & -\omega^2 M_b + j\omega C_{bb} + K_{bb} & j\omega C_{b3} + K_{b3} \\ & j\omega C_{3b} + K_{3b} & -\omega^2 M_3 + j\omega C_{33} + K_{33} \end{bmatrix} \begin{Bmatrix} \bar{Q}_s \\ \bar{X}_b \\ \bar{X}_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ (j\omega C_u + K_u)\bar{U} \end{Bmatrix} \tag{5}$$

where $\bar{Q}_s$ is the vector of modal coordinates and N is the number of modes of the isolated structure.

With another modal transformation,

$$\begin{Bmatrix} \bar{X}_b \\ \bar{X}_3 \end{Bmatrix} = \Psi \bar{Q}_I = \begin{bmatrix} \Psi_b \\ \Psi_3 \end{bmatrix} \bar{Q}_I \tag{6}$$

and by supposing that the physical parameter matrices of the isolator can be diagonalized with this modal matrix Ψ, i.e.,

$$D_I = \Psi^T \begin{bmatrix} -\omega^2 M_b + j\omega C_{bb} + K_{bb} & j\omega C_{b3} + K_{b3} \\ j\omega C_{3b} + K_{3b} & -\omega^2 M_3 + j\omega C_{33} + K_{33} \end{bmatrix} \Psi = \underset{r=1,2,\ldots,R}{\mathrm{diag}} \{k_{Ir} - \omega^2 m_{Ir} + j\omega c_{Ir}\} \tag{7}$$

Eq. (5) can be further transformed into

$$\begin{bmatrix} D_s & \Phi_2^T K_b \Psi_b \\ \Psi_b^T K_b^T \Phi_2 & D_I \end{bmatrix} \begin{Bmatrix} \bar{Q}_s \\ \bar{Q}_I \end{Bmatrix} = \begin{Bmatrix} 0 \\ \Psi_3^T (j\omega C_u + K_u)\bar{U} \end{Bmatrix} \tag{8}$$

This is an equation explicitly constituted from the modal parameters of the isolated structure and the isolator. In Eq. (7), R is the number of modes of the isolator with a fixed bottom boundary.

From Eq. (8), modal responses of the isolator and the isolated structure can be expressed as

$$\bar{Q}_I = [E + D_I]^{-1} \Psi_3^T (j\omega C_u + K_u)\bar{U} \tag{9}$$

and

$$\bar{Q}_s = -D_s^{-1} \Phi_2^T K_b \Psi_b \bar{Q}_I = -D_s^{-1} \Theta \Psi_b [E + D_I]^{-1} \Psi_3^T (j\omega C_u + K_u)\bar{U} \tag{10}$$

In Eqs. (9) and (10),

$$E = -\Psi_b^T K_b^T \Phi_2 D_s^{-1} \Phi_2^T K_b \Psi_b = -\sum_{n=1}^{N} (k_{sn} - \omega^2 m_{sn})^{-1} \Psi_b^T \theta_n^T \theta_n \Psi_b \tag{11}$$

where

$$\Theta = \begin{Bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{Bmatrix}_{N\times N_2}, \quad \theta_n = \phi_n^{2T} K_b \quad (n = 1,2, \ldots, N),$$

ϕ_n^2 is the part of modal vector corresponding to the connection that is defined as the following,

$$\Phi_2 = [\phi_1^2, \phi_2^2, \ldots, \phi_N^2]_{N_2\times N},$$

N_2 is the number of degrees of freedom of the isolated structure that connect to the isolator. Obviously, θ_n defines the connection condition.

A direct conclusion from Eqs. (9) and (10) is that the flexibility of the isolated structure will introduce more resonant frequencies to the coupled structure, and also the flexibility of the isolator.

3 "Anti-Resonance"

The phenomenon of "anti-resonance" usually appears when some vibration absorbers are attached onto a structure, and at any resonant frequency of these absorbers the vibration amplitude of the main structure reaches to its minimum value in the neighborhood of this frequency [6]. As a structure on the top of an isolator is an attached structure to the isolator, if its damping level is low, the "anti-resonance" phenomenon can also appear.

Equation (9) and Eq. (10) can be rewritten as the polynomial of the frequency variable ω, which are

$$\bar{Q}_I = G_s\left[-\sum_{n=1}^{N} \Psi_b^T \theta_n^T \theta_n \Psi_b \prod_{\substack{i=1\\ i\neq n}}^{N} (k_{si} - \omega^2 m_{si}) + D_I G_s\right]^{-1} \Psi_3^T (j\omega C_u + K_u)\bar{U} \tag{12}$$

and

$$\bar{Q}_s = -P\Theta\Psi_b\left[-\sum_{n=1}^{N} \Psi_b^T \theta_n^T \theta_n \Psi_b \prod_{\substack{i=1\\ i\neq n}}^{N} (k_{si} - \omega^2 m_{si}) + D_I G_s\right]^{-1} \Psi_3^T (j\omega C_u + K_u)\bar{U} \tag{13}$$

where

$$G_s = \prod_{n=1}^{N} (k_{sn} - \omega^2 m_{sn})$$

$$P = \operatorname*{diag}_{n=1,2,\ldots,N}\left\{\prod_{\substack{i=1\\ i\neq n}}^{N} (k_{si} - \omega^2 m_{si})\right\}$$

It can be seen from Eq. (12) that at any natural frequency of the isolated structure, i.e.,

$$\omega = \omega_{sn} = \sqrt{k_{sn} m_{sn}^{-1}},$$

all modal responses of the isolator are zero. In other words, the transmissibility of the vibration from the bottom boundary of the isolator to the interface with the isolated structure is zero. Nevertheless, the modal responses of the isolated structure are not all zero, and they are

$$\bar{q}_{si} = \begin{cases} -\phi_{2n}^T K_b \Psi_b [\Psi_b^T K_b^T \phi_n^2 \phi_n^{2T} K_b \Psi_b]^{-1} \Psi_3^T (j\omega C_u + K_u)\bar{U} & i = n \\ 0 & i \neq n \end{cases} \tag{14}$$

This is a typical "anti-resonance" situation. In other words, without the damping, the isolated structure becomes a flexible vibration absorber of the isolator. This is also true for a chain of structures connected together; the top one always is the vibration absorber of others.

This "anti-resonance" phenomenon raises a question about the evaluation of the isolation effect of an isolator. From the viewpoint of the transmissibility, which is defined by the absolute value of the ratio of the bottom boundary vibration of the isolator to the vibration at the interface with the isolated structure, the vibration transmitted to the structure is zero at this frequency. However, because of the flexibility, the vibration of the isolated structure is not zero. Therefore, for a flexibility structure, vibrations at those concerned points should be used for evaluating the isolation effect.

4 Isolation Frequency

If the flexibility of the isolated structure is not taken into account, an important index to the isolator is the isolation frequency, which is defined as the square root of the stiffness of the isolator divided by the mass of the isolated object. When only the first natural frequency of the isolated structure is considered, a requirement on this defined isolation frequency is that it should be at least smaller than $\sqrt{0.5}$ times of this natural frequency.

To study the case that all natural frequencies of the isolated structure are involved, for the sake of obtaining an analytical solution, the following assumptions will be made: all degrees of freedom at the interface have same displacement and the isolator only has one degree of freedom. With reference to Eq. (8), these assumptions can be mathematically expressed with the following definitions

$$\Psi_b = \begin{Bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{Bmatrix}_{N_2\times 1}, \Psi_3 = 1,\ R = 1,\ D_I = k_I - \omega^2 m_I + j\omega c_I,$$

$$k_I = K_u,\ c_I = C_u,\ \bar{Q}_I = \bar{q}_I,\ \bar{U} = \bar{u}.$$

Here $\bar{q}$ and $\bar{u}$ are scalar. With the above conditions and equations, Eqs. (12) and (13) can be simplified as

$$\bar{\mathrm{q}} = \left[-\sum_{n=1}^{N} M^{-1} m_{sn}^{-1} (\omega_{sn}^2 - \omega^2)^{-1} \lambda_n^2 + (\omega_I^2 - m_I M^{-1}\omega^2 + 2j\varsigma_I \omega\omega_I)\right]^{-1} (2j\varsigma_I \omega\omega_I + \omega_I^2)\bar{u} \tag{15}$$

$$\bar{Q}_s = -\{m_{s1}^{-1}(\omega_{s1}^2 - \omega^2)^{-1}\lambda_1, m_{s2}^{-1}(\omega_{s2}^2 - \omega^2)^{-1}\lambda_2, \ldots, m_{sN}^{-1}(\omega_{sN}^2 - \omega^2)^{-1}\lambda_N\}^T \bar{q} \tag{16}$$

where

$$\lambda_n = \phi_{2n}^T K_b \Psi_b, \quad \omega_I = \sqrt{\frac{k_I}{M}}, \quad \varsigma_I = \frac{c_I}{2\omega_I M},$$

and M is the total mass of the isolated structure. According to the traditional definition, ω_I is the isolation frequency.

With this simplification, the transmissibility across the isolator from the bottom boundary to the interface in the frequency domain can be written explicitly as

$$\left|\frac{\bar{q}}{\bar{u}}\right| = \sqrt{\frac{4\omega^2\omega_I^2\varsigma_I^2 + \omega_I^4}{\left[(\omega_I^2 - \omega^2 m_I M^{-1}) - \sum_{n=1}^{N}(m_{sn}M)^{-1}(\omega_{sn}^2-\omega^2)^{-1}\lambda_n^2\right]^2 + 4\omega^2\omega_I^2\varsigma_I^2}} \tag{17.1}$$

From equation (16), it can be found that the transmissibility to a modal coordinate is

$$\left|\frac{\bar{q}_{sn}}{\bar{u}}\right| = \frac{|\lambda_n|}{m_{sn}|\omega_{sn}^2-\omega^2|}\sqrt{\frac{4\omega^2\omega_I^2\varsigma_I^2 + \omega_I^4}{\left[(\omega_I^2 - \omega^2 m_I M^{-1}) - \sum_{n=1}^{N}(m_{sn}M)^{-1}(\omega_{sn}^2-\omega^2)^{-1}\lambda_n^2\right]^2 + 4\omega^2\omega_I^2\varsigma_I^2}} \tag{17.2}$$

With the traditional definition of isolation, it is required that

$$\left|\frac{\bar{q}}{\bar{u}}\right| < 1 \tag{18}$$

When the excitation frequency is not equal to any resonant frequency of the isolated structure and the coupled structure, this equals to the requirement that

$$\left[m_I M^{-1}\omega^2 + \sum_{n=1}^{N}(m_{sn}M)^{-1}(\omega_{sn}^2-\omega^2)^{-1}\lambda_\mathrm{n}^2\right]^2 > 2\omega_I^2\left[m_I M^{-1}\omega^2 + \sum_{n=1}^{N}(m_{sn}M)^{-1}(\omega_{sn}^2-\omega^2)^{-1}\lambda_\mathrm{n}^2\right] \tag{19}$$

Because

$$\frac{df(\omega)}{d\omega} = 2\omega\left[m_I M^{-1} + \sum_{n=1}^{N}(m_{sn}M^{-1})(\omega_{sn}^2-\omega^2)^{-2}\lambda_n^2\right] > 0, \tag{20}$$

function

$$f(\omega) = m_I M^{-1}\omega^2 + \sum_{n=1}^{N}(m_{sn}M)^{-1}(\omega_{sn}^2-\omega^2)^{-1}\lambda_n^2 \tag{21}$$

is a incremental function of the frequency variable ω on condition that the excitation frequency is not equal to any resonant frequency of the isolated structure. Therefore, after a certain value of the frequency, the inequality (19) will be satisfied. This frequency can be considered as the effective frequency of the isolator.

If the value of function $f(\omega)$ is negative, inequality (19) is naturally satisfied. When its value is positive, after it is larger than $2\omega_I^2$, i.e.,

$$\sqrt{f(\omega)} = \sqrt{m_I M^{-1}\omega^2 + \sum_{n=1}^{N}(m_{sn}M)^{-1}(\omega_{sn}^2-\omega^2)^{-1}\lambda_n^2} > \sqrt{2}\omega_I, \tag{22}$$

the transmissibility will be less than 1. Because $f(\omega)$ is the function of not only the frequency variable, but also all natural frequencies of the isolated structure, the isolation frequency does not mean that the isolator will be effective after the frequency variable is larger than its $\sqrt{2}$ times. Instead, since $f(\omega)$ is an incremental function of the frequency variable, the isolation frequency ω_I defines, in fact, a critical value for the existence of the inequality (22), which is also the effective frequency equation when the symbol of greater is replaced with the symbol of equal. A lower isolation frequency means that the isolator will be effective earlier in the frequency domain, or in other words, the isolator has a lower effective frequency.

It can be seen from Eq. (16) that the quality of the isolation is not only decided by the isolator itself, but also by the dynamic characteristics of the isolated structure and the connection condition between them.

If it is also required that

$$\left|\frac{\bar{q}_{sn}}{\bar{u}}\right| < 1, \tag{23}$$

in addition to those requirements on the effectiveness of the isolator, the following condition should be satisfied, too,

$$|\omega_{sn}^2-\omega^2| > \frac{|\lambda_n|}{m_{sn}} \tag{24}$$

This indicates that the excitation frequency should be outside the neighborhood of the "anti-resonance" frequency, whose radius is defined by $|\lambda_n|/m_{sn}$.

5 Resonance of Coupled Structure

Factors that determining the performance of an isolator at the resonant frequency of the coupled structure are also important to the design and selection of the isolator. As a part of the coupled structure, properties of the isolated structure will inevitably also have a significant influence on the performance of the isolator.

If the damping of the isolated structure is neglected, at any resonant frequency of the coupled structure, i.e., in Eq. (17.1)

$$(\omega_I^2 - \omega^2 m_I M^{-1}) - \sum_{n=1}^{N}(m_{sn}M)^{-1}(\omega_{sn}^2-\omega^2)^{-1}\lambda_n^2 = 0, \tag{25}$$

there is

$$\left|\frac{\bar{q}}{\bar{u}}\right| = \sqrt{1 + \frac{\omega_I^2}{4\omega^2\varsigma_I^2}}, \tag{26.1}$$

or

$$\left|\frac{\bar{q}}{\bar{u}}\right| = \sqrt{1 + \frac{k_I^2}{\omega^2 c_I^2}} \tag{26.2}$$

This means if the isolated structure is a flexible structure and without any damping, at the resonant frequency of the coupled structure, the transmissibility across the isolator cannot be less than 1. Equation (26) also indicates that a lower resonant frequency means a higher transmissibility. Apparently, a softer isolator, i.e., a lower isolation frequency, and/or a high damping isolator can reduce this transmissibility.

If the isolated structure has damping, the transmissibility can be less than 1 at some resonant frequencies as

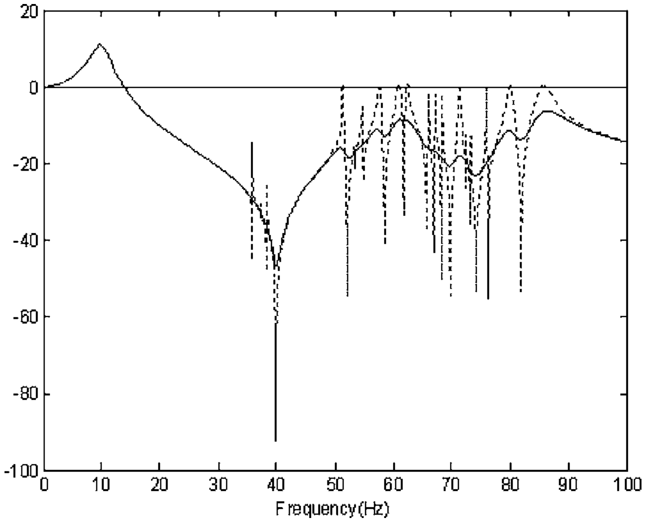


Fig. 2 Isolation frequency 10 Hz (dotted line $\eta=0$, solid line $\eta=0.02$)

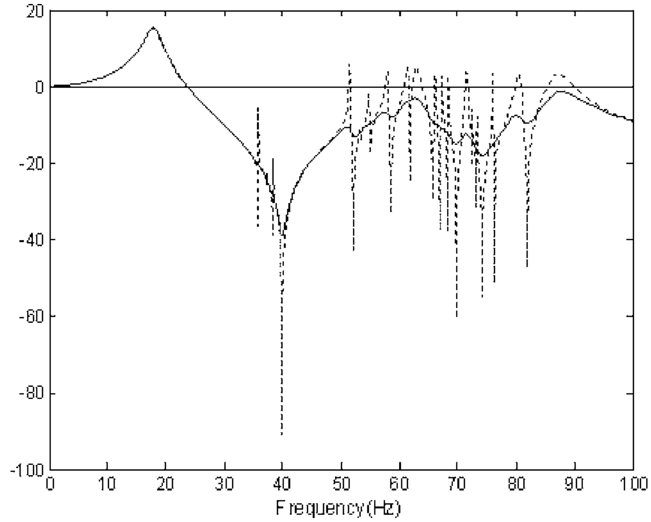


Fig. 3 Isolation frequency 18 Hz (dotted line $\eta=0$, solid line $\eta=0.02$)

$$\left|\frac{\bar{q}}{\bar{u}}\right| = \frac{1}{1+\sum_{n=1}^{N}\frac{\lambda_n^2 \varsigma_{sn}\varsigma_I^{-1}\omega_{sn}\omega_I^{-1}m_{sn}M}{(\omega_{sn}^2-\omega^2)^2+4\omega^2\omega_{sn}^2\varsigma_{sn}^2}}\sqrt{1+\frac{\omega_I^2}{4\omega^2\varsigma_I^2}} \qquad (27)$$

This result shows the importance of combining together the vibration isolation and the vibration attenuation, which is realized by adding damping or increasing the damping level.

At the resonant frequency, the transmissibility from the bottom of the isolator to a modal coordinate of the isolated structure is

$$\left|\frac{\bar{q}_{sn}}{\bar{u}}\right| = \frac{|\lambda_n|}{m_{sn}|\omega_{sn}^2-\omega^2|}\sqrt{1+\frac{k_I^2}{\omega^2 c_I^2}} \qquad (28)$$

It can be seen that a larger deviation of the resonant frequency from the natural frequencies of the isolated structure can decrease the transmissibility more effectively.

6 Numerical Example

Figure 1 is the finite element model of a satellite on the top of a whole spacecraft vibration isolation platform [7]. This is a highly flexible satellite with the first order natural frequency as low as about 14 Hz. Here, as the major purpose is to demonstrate those conclusions obtained in the above sections, for the sake of simplicity, only the vibration isolation in the vertical direction is discussed. The interface surface between these two structures is defined to be rigid.

In the vertical direction, the first order natural frequency of the satellite is about 40 Hz. Figures 2 and 3 are the transmissibility in dB across the isolation platform when the isolation frequency is chosen as 10 and 18 Hz, respectively. Here, η denotes the damping ratio of the isolated structure. According to Eqs. (26) and (27), the isolation effect of a lower isolation frequency is better than that of a higher isolation frequency. When the damping of the isolated structure is set to zero, the "anti-resonance" phenomenon can be clearly identified from these figures. The most obvious one is at the frequency about 40 Hz, which is the natural frequency of the satellite. As predicated by Eq. (27), the damping in the isolated structure can significantly reduce the transmissibility at the resonant frequencies of the coupled structure. With the damping, the "anti-resonance" phenomenon is not so obvious and even disappears at some natural frequencies of the isolated structure.

When the isolation frequency is chosen as 50 Hz, which is higher than the first order natural frequency of the satellite in the vertical direction, instead of 50 Hz, as shown in Fig. 4, the frequency of the first peak in the transmissibility becomes 33 Hz, which is neither the isolation frequency nor a natural frequency of the isolated structure. The isolation frequency loses its original meaning here. If the results of 18 Hz isolation frequency and 50 Hz isolation frequency are put on one diagram of Fig. 5, two major differences can be found, i.e., the frequency of the first peak and the effective frequency of the isolator.

7 Discussion and Conclusion

As components of a coupled structure, dynamic characteristics of both the isolator and the isolated structure will inevitably influence each other's behavior. Comparing with the rigid body model and the single-degree-of-freedom model, the flexibility model basically brings more resonant frequencies into the coupled structure. Reflected on the transmissibility, this fact means more peaks in the curve of the transmissibility on the amplitude-frequency plane. To reduce the vibration transmission across the isolator at these resonant frequencies, a higher damping in the isolator is helpful. Also, the damping in the isolated structure is important

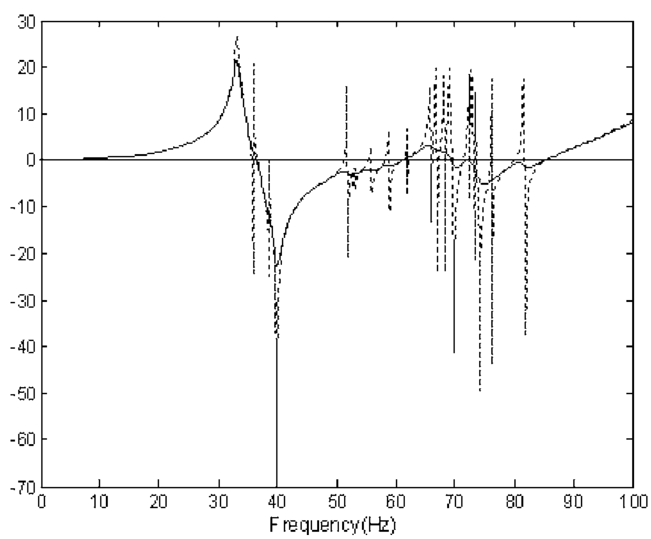


Fig. 4 Isolation frequency 50 Hz (dotted line $\eta=0$, solid line $\eta=0.02$)

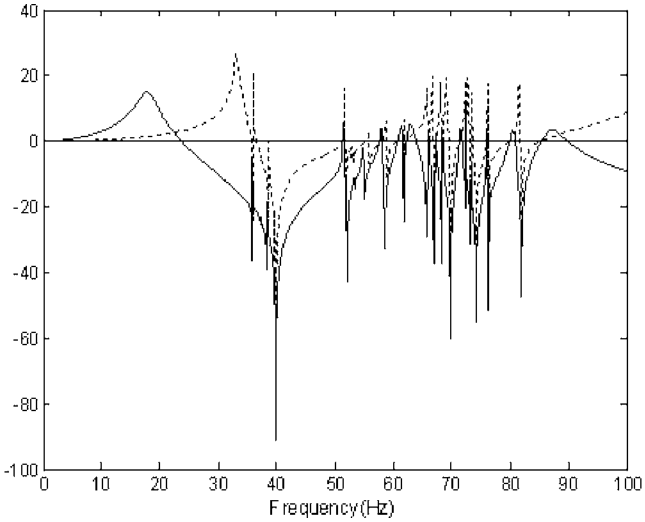


Fig. 5 Transmissibility of 50 Hz isolation frequency (dotted line) and 18 Hz isolation frequency (solid line) ($\eta=0$)

for further reducing the vibration transmission. In other words, for achieving a better result, the vibration isolation should be combined with the vibration attenuation.

Although the flexibility introduces many more resonant frequencies, and because of the coupling between the isolator and the isolated structure, the isolation frequency generally is not the same as or very close to the frequency of the first peak in the transmissibility curve along the frequency axis, a lower isolation frequency still means that the isolation effect is better than that of a higher one.

At the natural frequencies of the isolated structure (including the connection stiffness to the isolator), the isolated structure acts as a vibration absorber to the isolator. If the damping level of the isolated structure is low enough, there will be many "anti-resonance" points on the transmissibility curve. Nevertheless, at other frequencies, a lower transmissibility across the isolator still indicates that the vibration at any point of the isolated structure is lower.

In summary, to the vibration isolation of flexibility structures, the transmissibility across the isolator is still a proper way of defining the effectiveness of the isolation. The old conclusion, which is deduced with the rigid-body model, about the function of the damping in the isolation effectiveness is not valid in the case of flexible structure isolation.

References

[1] Inman, D. J., 1994, *Engineering Vibration*, 1st ed., Prentice–Hall, Englewood Cliffs, NJ.
[2] Mead, J. D., 1999, *Passive Vibration Control*, Wiley, UK.
[3] Rivin, I. E., 2003, *Passive Vibration Isolation*, ASME, New York.
[4] Zheng, G. T., 2003, "Parametric Studies of the Whole Spacecraft Vibration Isolation," AIAA J., **41**, pp. 1839–1841.
[5] Sciulli, D., and Inman, D. J., 1998, "Isolation Design for a Flexible System," J. Sound Vib., **216**, pp. 251–267.
[6] Nagaya, K., Kurusu, A., Ikai, S., and Shitani, Y. 1999, "Vibration Control of a Structure by Using a Tunable Absorber and an Optimal Vibration Absorber Under Auto-Tuning Control," J. Sound Vib., **228**, pp. 773–792.
[7] Liu, L. K., Liang, L., Zheng, G. T., and Huang, W. H., 2005, "Whole-Spacecraft Passive Vibration Isolation Using Octo-Strut Platform," J. Spacecr. Rockets, **42**, pp. 654–662.
[8] Meirovitch, L. 1997, *Principles and Techniques of Vibrations*, Prentice–Hall, Englewood Cliffs, NJ.

A Variational Principle Governing the Generating Function for Conformations of Flexible Molecules

L. B. Freund
Division of Engineering,
Brown University,
Providence, RI 02912-9104
e-mail: freund@brown.edu

The generation of a random walk path under the action of an external potential field has been of interest for decades. The motivation derives largely from the prospect of incorporating the nonlocal excluded volume effect through such a potential in characterizing the statistical behavior of a long flexible polymer molecule. In working toward a continuum mean-field model, a central feature is a partial differential equation incorporating the influence of the potential and governing the generating function for the dependence of end to end separation distance of the molecule on its pathlength. The purpose here is to describe an approach in which the differential equation is recast as a global minimization of a functional. The variational approach is illustrated by an application to familiar configurations, the first of which is a molecule attached at one end to a noninteracting plane barrier in the presence of a uniform potential field. As a second illustration, the generating function is sought for a free molecule for the case in which conformations must be consistent with the excluded volume condition. This is accomplished by adapting a local form of the Flory approach to the phenomenon and extracting estimates of the expected end to end separation distance, the entropy and other statistical features of behavior. By means of the variational principle, the problem is recast into a form that admits a direct, noniterative analysis of conformations within the context of the self-consistent field theory. [DOI: 10.1115/1.2201888]

1 Introduction

An attempt at a detailed description of the geometrical configuration of a linear polymer macromolecule in space invariably meets with significant complexity. Adjacent monomers in the polymer typically have considerable freedom of relative motion across shared bonds. As a result, the number of possible configurations of a long molecule in space is enormous. If such a molecule is in a dilute solution or is otherwise free of external constraints, all such configurations are accessible. Understanding the thermodynamic properties of polymers derives from a characterization of the accessible configurations through the principles of statistical mechanics.

In those cases in which the behavior of each monomer in the polymer chain is influenced only by its immediate neighbors along the chain, the theory of random walk paths has provided an effective tool for representing distributions of conformations. A principal shortcoming of the classical random path description is that it cannot account for the fact that two monomers in a chain cannot occupy the same place in space at the same time, no matter what their relative position along the chain might be. This is the so-called excluded volume effect of polymer molecules, an effect that produces a nonlocal constraint on admissible configurations and thereby influences thermodynamic properties. The first successful treatment of the effect was provided by Flory in 1949 and was summarized in a monograph that appeared a few years later [1]. The basic idea is that the polymer chain is assumed to exist as a gas of its own monomers, and the real (nonideal) gas interactions that arise lead to an estimate of the swelling of the molecule due to this additional constraint.

In an effort to include the excluded volume effect within a continuum mean field theory of polymer conformations, Edwards [2] proposed the idea of studying random path generation in the presence of an externally imposed potential field, with the eventual goal being the representation of nonlocal interactions within the chain by means of an appropriately chosen potential. The work led to a partial differential equation governing a probability distribution function for end to end separation distance of the molecule in the presence of an external potential field.

This partial differential equation is difficult to solve, even in simple cases. Therefore, as an aid in extracting approximate solutions and as a basis for numerical simulation studies, the formulation is recast as a minimum principle through a variational approach. Representative applications of the principle are illustrated. The first example is the case of a molecule attached at one end to an impenetrable noninteracting plane surface with the other in free, a case for which the exact solution of the governing partial differential equation is known. The second example is the case of a free molecule including the excluded volume effect, a configuration for which exact solutions are not known.

2 Background

The notion of a random walk path has served as a cornerstone concept in statistical mechanics of large flexible polymer molecules, as well as in other branches of science, and its features continue to be of active interest [3]. The success of the concept derives largely from its simplicity, and it is most easily described in terms of a sequence of straight line segments of fixed length, say b, strung out in a head-to-tail fashion in space. Starting at some point O, the first line segment extends a distance b from O in an arbitrary direction. The next segment extends from the end of the first a distance b in an arbitrary direction in space, and so on

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received March 27, 2006; final manuscript received March 31, 2006. Review conducted by R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California – Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

until n such segments have been pieced together, ending at point P. The result is a random walk path. The statistical distribution of vertex positions and the probability distribution of end-to-end distance from O to P have been key to understanding Brownian motion of small particles in a fluid or the statistics of conformations of long flexible polymer molecules.

Denoting the position of P by the vector $\mathbf{r}$ from O to P, let $p(\mathbf{r},n)$ be the probability of finding the end of the path at point P in space after n steps, located with respect to origin at O. This is a scalar valued function of path end position with the property that $p(\mathbf{r},n)dR$ is the fraction of all possible paths that end within an infinitesimal volume dR at P. Alternatively, it is the number of path ends per unit volume at P. Because every path must end somewhere, it is necessary that

$$\int_{R_3} p(\mathbf{r},n)dR = 1 \tag{1}$$

where R_3 is all of three-dimensional space accessible to the configuration, that is, $p(\mathbf{r},n)$ is a probability distribution function.

In a pioneering paper by Chandrasekhar [4], it was demonstrated that, in the limit of very small b and very large n with nb held fixed, the probability distribution function satisfies the linear partial differential equation governing isotropic gradient-driven diffusion, namely,

$$\partial_n p = \frac{b^2}{6}\nabla^2 p \tag{2}$$

with pathlength (as represented by n for fixed b) replacing the role of time in the classical counterpart. With this interpretation, the parameter n becomes a continuous variable. The "diffusivity," $b^2/6$ in the case of (2), depends to some degree on the details of the model on which the derivation of the partial differential equation is based. Chandrasekhar used a lattice model, whereby all step ends necessarily coincide with the points in a regular lattice. With the interpretation of $p(\mathbf{r},n)$ as a spatial density, Eq. (2) admits the usual interpretation as a conservation condition. According to this interpretation, the local rate of change of density $\partial_n p$ is equal to the negative of the local divergence of the flux $\mathbf{h} = -(b^2/6)\nabla p$. The flux lines are not the paths themselves but, instead, are the trajectories of path ends as n increases.

The partial differential equation (2) must be augmented by an initial condition at $n=0$, and

$$p(\mathbf{r},0) = \delta(\mathbf{r}) \tag{3}$$

is commonly adopted as this condition in order to satisfy (3) for arbitrarily small length. The right side of (3) is understood to be the appropriate form of the Dirac delta function for whichever coordinate system is adopted. If there is no spatial constraint, then both g and $|\nabla g|$ are normally assumed to become vanishingly small in the limit as $r\to\infty$. The solution of (2) subject to the initial condition (3) without boundary constraints is the classical spherically symmetric free space Green's function,

$$p(r,n) = \left(\frac{3}{2\pi nb^2}\right)^{3/2} e^{-3r^2/2nb^2} \tag{4}$$

which satisfies the normalization condition (1).

The number of conformations ending within the infinitesimal volume dR at radial distance r from O in a *particular* direction is $p(r,n)dR$. On the other hand, the number terminating between radial distance r and $r+dr$ in *any* direction is $4\pi r^2 p(r,n)dr$. If dr is replaced by b (or any other small length parameter), the partition function of conformations ending between r and $r+b$ is $\mathcal{Z}(r)=4\pi r^2 bp(r,n)$, so the entropy is $\mathcal{S}(r)=k\ln\mathcal{Z}(r)$. The free energy $\mathcal{F}$ is entirely entropic, so

$$\mathcal{F}(r) = -T\mathcal{S}(r) = kT\left(\frac{3r^2}{2nb^2} - \ln\frac{r^2}{b^2}\right) \tag{5}$$

to within an additive constant, where k is the Boltzmann constant and T is absolute temperature. The most probable end to end separation distance $\langle r\rangle$ is then

$$\langle r\rangle = \frac{\int_0^\infty re^{-\mathcal{F}(r)/kT}dr}{\int_0^\infty e^{-\mathcal{F}(r)/kT}dr} = 2b\sqrt{\frac{2n}{3\pi}}. \tag{6}$$

The quantity $\mathcal{F}'(r)$ can be viewed as a generalized force acting at the end of the molecule to maintain a separation distance r. The expected value $\langle\mathcal{F}'(r)\rangle$ of this force is zero. The implication is that a radial outward force is required to maintain conformations with end to end distance greater than $\langle r\rangle$, whereas a radial inward force is required to maintain conformations at smaller radii.

The most probable radius $\langle r\rangle$ calculated in (6) can be compared to the estimate $r_{\text{eq}}=b\sqrt{2n/3}$ obtained as an equilibrium condition, that is, the value of r for which $\mathcal{F}'(r)=0$ with $\mathcal{F}''(r)>0$. The latter is slightly smaller than $\langle r\rangle$. In either case, the mean end to end separation position (as opposed to distance) is zero in light of spherical symmetry.

3 Effect of a Potential Field

In the preceding section, the process of generating a random path was described without reference to interactions of the chain links with their surroundings. The only factor controlling a particular step in the random walk was the endpoint of the previous step. Beyond that, the direction of the next step was completely arbitrary. In this section, the issue of generating a random walk path in the presence of a potential field is reviewed.

Suppose that the scalar valued function of position $kTu(\mathbf{r})$ defines a potential energy field throughout the region of interest. With potential energy defined in this way, u is dimensionless. Its physical origin is immaterial for the time being. Before proceeding, it is necessary to be more specific about the nature of the free energy of interaction of the chain links with this potential. The simplest point of view to adopt is that each link is insensitive to the potential except at its endpoint. If this is so, then the potential energy of the mth link with endpoint position $\mathbf{r}_m$ is $kTu(\mathbf{r}_m)$. Consequently, this point experiences a force $-kT\,\nabla u|_{\mathbf{r}=\mathbf{r}_m}$ acting along the potential gradient.

On the other hand, if all points on the link are equally sensitive to the potential field, then each link senses a net force at its midpoint that tends to push that link down the potential gradient. It also senses a torque that tends to rotate the link into the tangent plane of an equipotential surface in space, and so on. Still other characterizations of the nature of interactions can be imagined. The endpoint interactions mentioned first are the simplest to consider, so that viewpoint is adopted for the remainder of the discussion. Conceptually, there is no difficulty in accommodating other points of view. In general, the interaction energy per unit volume equals the number of links per unit volume times the local strength of the field. For example, consider all conformations ending at some end to end distance r. Let $\rho(\mathbf{r}_*,r)$ be the density of links of these conformations per unit volume at point $\mathbf{r}_*$. Then the local contribution to the free energy is $\rho(\mathbf{r}_*,r)u(\mathbf{r}_*)dR_*$, where dR_* is a local volume element at point $\mathbf{r}_*$.

The generation of a random walk in a potential field has been discussed often in the literature on polymer molecule conformations. Among the earliest treatments are those by Edwards [2] (see Eq. 3.1), de Gennes [5] (see Eq. 1.10) and Freed [6] (see Eq. 6.25). In these cases, the influence of the potential energy field was included in the formulation through a Boltzmann factor associated with each link, and a particular partial differential equation governing the generating function $g(\mathbf{r},n)$, rather than a probability distribution function, was extracted. This equation has the form of a Schrödinger equation with imaginary time, that is, the so-called Feynman-Katz equation. Since then, many papers have appeared in the literature which present results on the basis of this modified diffusion equation.

The derivation of the modified field equation proposed by Edwards, de Gennes, and Freed has been reviewed by Grosberg and Khokhlov [7] (see Eq. 6.28) and the result has the form

$$\partial_n g = \frac{b^2}{6}\nabla^2 g - u(\mathbf{r})g. \tag{7}$$

The solution of this equation provides the generating function $g(\mathbf{r},n)$ of conformations of length n beginning at an arbitrary origin and ending at point $\mathbf{r}$ with respect to that origin. This particular generating function is commonly designated as the Green's function or the propagator of the molecule. Physically, it represents the partition function of the chain with fixed endpoints. While some solutions of the differential equation in one space dimension are well known, solutions for three-dimensional cases are elusive, in general. In situations for which u is intended to represent the excluded volume effect, the situation is more complicated. Then, the potential depends on the spatial distribution of monomers in the chain, but this distribution is unknown before the equation is solved. Consequently, the process of analysis becomes convoluted. As a basis for studying (7) when u is specified, a variational approach is described in the section that follows. The approach is then applied to a case of a molecule attached to a noninteracting plane barrier in the presence of the simple potential field. Then, as a second illustration, it is also applied to the case of an isolated molecule for which conformations are restricted by the excluded volume effect.

4 Variational Formulation

For a given state of the system at some value of n as represented by the fields g and u at that value of n, the essential feature of the partial differential equation is that it prescribes how the system leaves that state as n increases. A variational principle is sought that has the same characteristic feature. Similar partial differential equations have been analyzed in modeling the evolution of material microstructure due to mass transport, and these have been successfully addressed via a variational approach [8–10].

To this end, each term in the partial differential equation is multiplied by an arbitrary scalar function of position, say $h(\mathbf{r},n)$, and then integrated over the spatial domain of the differential equation, say R. The result, following application of the divergence theorem, is

$$\int_R h\,\partial_n g\,dR = -\int_R\left(\frac{b^2}{6}\nabla h\cdot\nabla g + ugh\right)dR + \frac{b^2}{6}\int_S h\,\nabla g\cdot\mathbf{n}\,dS, \tag{8}$$

where S is the boundary surface of R. This provides the so-called weak form of the partial differential equation. The differential equation can be recovered by appeal to the arbitrariness of h.

Next, the function h is assigned a particular role, without compromising its arbitrariness, that permits the interpretation of (8) as the vanishing of the first variation of a particular functional. Recall that the goal is to determine how the system leaves a certain configuration represented by g and u as n increases from its current value. If the rate of departure is denoted by $f(\mathbf{r},n)$ (which is $\partial_n g$ for the exact solution) then the obvious choice for h is as an arbitrary variation in f, say δf in the standard notation of the calculus of variations. If f is an admissible rate of departure, then $f+\delta f$ is another admissible rate of departure. With that interpretation, the form of (8) is equivalent to the vanishing of the first variation of a functional, that is,

$$\delta\Phi = 0, \tag{9}$$

where the functional is defined by

$$\Phi[f] = \int_R\left(\frac{1}{2}f^2 + \frac{b^2}{6}\nabla f\cdot\nabla g + ugf\right)dR. \tag{10}$$

The variation operator δ can be moved outside the integral sign because neither g, u nor the domain R are altered when the rate of departure is perturbed. Furthermore, in all cases of interest here either δf or ∇g vanishes on S so that the boundary integral term in (8) makes no contribution. When written in this way, it is evident that the functional $\Phi[f]$ is stationary under arbitrary perturbations in f when f is the actual rate of departure, that is, when $f\equiv\partial_n g$. Furthermore, it can readily be verified that the second variation of the functional is positive definite, so the functional is not only stationary for the actual rate of change of the generating function but it is also an absolute minimum there.

In light of the property of the functional as possessing an absolute minimum value among all admissible conformations for the exact solution, the variational form of the governing equation is ideal as a basis for numerical simulation of the evolution process by means of the finite element method. The objective here is more modest, namely, to illustrate the variational approach toward obtaining approximate generating functions for a couple of nontrivial configurations by exploiting its properties of being both stationary and minimum for the exact generating function.

The variational principle (10) can be applied in the following way to determine an approximate generating function. Suppose it is assumed at the outset that g has the form

$$g(\mathbf{r},n) = G[\mathbf{r},\varphi_1(n),\ldots,\varphi_M(n)] \tag{11}$$

where G is an explicit function of its arguments but the functions $\varphi_m(n)$, $m=1,\ldots,M$ themselves are unknown. Then

$$f(\mathbf{r},n) = \sum_{m=1}^{M}\frac{\partial G}{\partial\varphi_m}\varphi_m'(n) \tag{12}$$

where the prime denotes differentiation with respect to n. Next, the functional Φ can be evaluated in terms of the unknown functions $\varphi_m(n)$ and $\varphi_m'(n)$. In these circumstances, minimization of Φ under variations in the rate f is equivalent to minimization with respect to the M variables $\varphi_m'(n)$ in the usual sense of functions of several variables, that is,

$$\frac{\partial\Phi}{\partial\varphi_m'} = 0, \quad m=1,\ldots,M. \tag{13}$$

This condition yields a set of M ordinary differential equations, in general nonlinear, for the M functions $\varphi_m(n)$. The set of differential equations must be augmented by a suitable set of initial conditions chosen to be compatible with the initial condition (3) in any particular case. The solution of the differential equations provides an approximate generating function for the configuration of interest. From the variational principle, it is known that the quality of the approximation improves (or, at least, it does not diminish) if the set of admissible functions, (11) is made richer in content. This procedure is next demonstrated in a particular case for which an exact solution is known.

5 A Simple Application of $\delta\Phi=0$

The configuration considered is a flexible molecule of length nb with one end bonded to an impenetrable, noninteracting plane surface while the motion of the other end is unconstrained. Furthermore, the region of space that is accessible to the molecule is exposed to a spatially uniform potential field $u(\mathbf{r})=u_0$. The objective is to determine the generating function g.

A convenient coordinate system in this case is the $r\theta z$ cylindrical system with $z=0$ being the barrier plane, $r=z=0$ being the attachment point on the barrier plane and the molecule being confined to the half-space $z \geq 0$. The generating function is subject to the boundary condition $g(r,0,n)=0$ for $n>0$. At remote points, g is expected to decay exponentially to zero in magnitude. Finally, the configuration is invariant under rotation about the z axis so g is expected to be axisymmetric, that is, independent of θ.

An admissible form of the generating function with these properties is

$$g(r,z,n) = \frac{z}{b^4}\varphi(n)e^{-\psi(n)(r^2+z^2)/b^2} \tag{14}$$

where the powers of b have been chosen on dimensional grounds and $\varphi(n)$, $\psi(n)$ are unknown dimensionless functions of n. To be compatible with the initial conditions (3), these functions must have the asymptotic characteristics

$$\varphi(n) \to \frac{3}{n}\left(\frac{3}{2\pi n}\right)^{3/2}, \quad \psi(n) \to \frac{3}{2n} \tag{15}$$

as $n \to 0^+$.

The rate of change of the generating function is then calculated as in (12), the functional Φ is evaluated, and the requirement that it be minimum with respect to variations in $\varphi'(n)$ and $\psi'(n)$ is imposed to obtain ordinary differential equations for the two unknown functions. In the present case, these equations are found to be

$$12\psi(n)\varphi'(n) + \varphi(n)[12u_0\psi(n) + 10\psi(n)^2 - 15\psi'(n)] = 0$$

$$4\psi(n)\varphi'(n) + \varphi(n)[4u_0\psi(n) + 2\psi(n)^2 - 7\psi'(n)] = 0. \tag{16}$$

The possibility that $\varphi(n) \equiv 0$ and/or $\psi(n) \equiv 0$ was excluded in extracting these ordinary differential equations. The solution of (16) consistent with (15) is

$$\varphi(n) = \frac{3e^{-nu_0}}{n}\left(\frac{3}{2\pi n}\right)^{3/2}, \quad \psi(n) = \frac{3}{2n}. \tag{17}$$

It is readily verified by direct substitution that this provides the actual solution to the initial-boundary value problem. This was relatively easy to achieve in the present case because the exact solution is known; the assumed form is consistent with the spatial dependence of the exact solution and the minimum principle then naturally selected the optimal functions $\varphi(n)$ and $\psi(n)$ to provide the corresponding dependence on n.

The expected value of the distance of the endpoint of the molecule from its attachment point can be calculated in any of several ways. With $g(r,z,n)$ understood as the partition function (or statistical weight) for a molecule ending at a particular point with coordinates (r,θ,z) for any $0 \leq \theta < 2\pi$, the expected value of the end to end distance is

$$\langle\sqrt{r^2+z^2}\rangle = \frac{2\pi\int_0^\infty\int_0^\infty \sqrt{r^2+z^2}\,g(r,z,n)r\,dr\,dz}{2\pi\int_0^\infty\int_0^\infty g(r,z,n)r\,dr\,dz} = b\sqrt{\frac{3\pi n}{8}}. \tag{18}$$

This expected value of end to end separation distance is about 18% larger than the expected end to end separation distance in the absence of a confining barrier, as given in (6). Because the free energy due to the background energy field of uniform strength u_0 is the same for all conformations, the expected value of end to end distance is independent of the potential field.

A slightly different measure of the range of the bound molecule is obtained in the following way. The partition function for all conformations of the bound molecule that end on a particular plane $z=\text{const}$ is

$$\mathcal{Z}(z,n) = 2\pi\int_0^\infty g(r,z,n)r\,dr. \tag{19}$$

In this case, the expected value of the distance of the free end of the molecule from the barrier plane is

$$\langle z\rangle = b\sqrt{\pi n/6}. \tag{20}$$

6 Example With Excluded Volume

In this case, the equation governing the generating function is again (7) with the potential u given by the product of the excluded volume parameter $v \sim b^3$ times the local density ρ of monomers in the chain. This form of the potential follows from the viewpoint that the excluded volume effect can be modeled by representing the chain as a real (nonideal) gas of its links and retaining only the first two terms of the virial expansion for energy per unit volume in powers of spatial density of links. The task of calculating this density for all conformations with a certain end to end separation distance has been addressed by Dolan and Edwards [11].

To determine this density by largely heuristic reasoning, suppose that the notation for generating function is generalized for the moment to reflect the positions of both ends of the chain. For example, the generating function for a molecule starting at point $\mathbf{r}_0$ and ending at point $\mathbf{r}$ at length n, the generating function is $g(\mathbf{r},\mathbf{r}_0,n)$. Then, the number of conformations that start at $\mathbf{r}_0$, pass through the point $\mathbf{r}_*$ at a particular length m, and end at the point $\mathbf{r}$ at total length n is

$$g(\mathbf{r}_*,\mathbf{r}_0,m)g(\mathbf{r},\mathbf{r}_*,n-m). \tag{21}$$

Therefore, the number of such confirmations that pass through $\mathbf{r}_*$ at *any* length m in the range $0<m<n$ is

$$\int_0^n g(\mathbf{r}_*,\mathbf{r}_0,m)g(\mathbf{r},\mathbf{r}_*,n-m)dm. \tag{22}$$

If the ends of the chain of length n would be fixed, then this quantity would be the chain density at point $\mathbf{r}_*$ in space. To account for the arbitrariness in the endpoint position $\mathbf{r}$ with respect to the origin, it is necessary to sum (22) over the full range of $\mathbf{r}$ and to normalize by the total number of conformations that are accessible, yielding the result

$$\rho(r_*) = \frac{\int_0^n g(\mathbf{r}_*,\mathbf{r}_0,m)\int_R g(\mathbf{r},\mathbf{r}_*,n-m)dR\,dm}{\int_R g(\mathbf{r},\mathbf{r}_0,n)dR}. \tag{23}$$

The density ρ can depend on $\mathbf{r}_*$ only through its magnitude r_* due to the spherical symmetry of the configuration.

The quantity ρ is not a physical density. Instead, it is the probability that, among all conformations, a link will lie within an infinitesimal volume at point $\mathbf{r}_*$. Because all n links in the chain must be somewhere, it is necessary that

$$4\pi\int_0^\infty \rho(r_*)r_*^2\,dr_* = n. \tag{24}$$

It is readily verified that (23) satisfies this constraint. The density ρ is used to determine the influence of the excluded volume effect on conformations.

In terms of the density $\rho(r_*)$, the potential field u is

$$u(r_*) = v\rho(r_*) \tag{25}$$

where $v \sim b^3$ is the excluded volume parameter. This is the potential energy of interaction per link at radial distance r_* in a particular direction due to repulsion by a molecule in any admissible configuration.

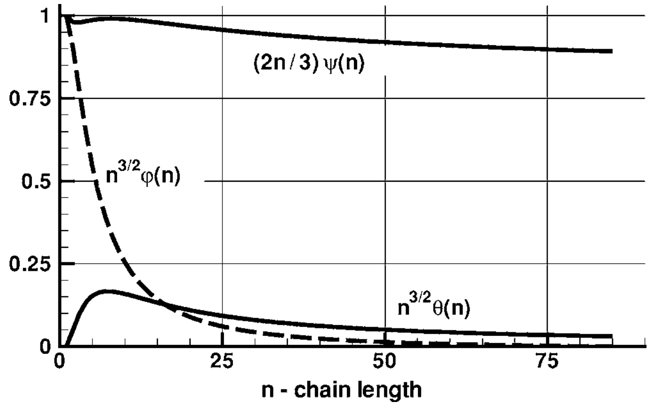


Fig. 1 The dependence of the unknown functions $\varphi(n)$ and $\psi(n)$ appearing in (26) as determined by means of the variational principle. The deviations from their initial values shows the influence of the excluded volume effect for the case when $v=b^3$

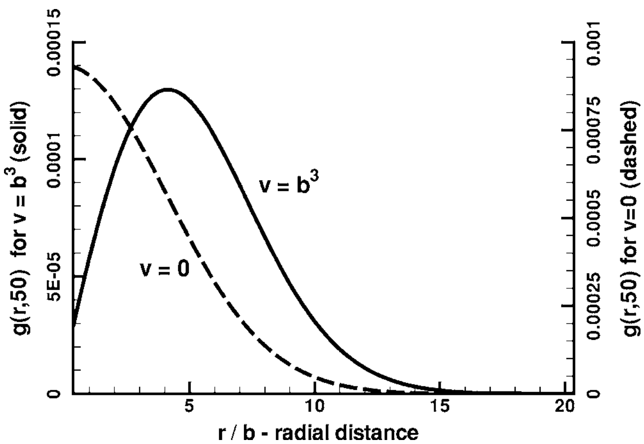


Fig. 2 The influence of the excluded volume of fact on the shape of the generating function for an isolated molecule. The curve shows the generating function $g(r,50)$ for the case in which the excluded volume parameter is $v=b^3$. For purposes of comparison, the corresponding function for the case in which the excluded volume effect is neglected is shown as a dashed curve plotted against the right hand scale, the range of which is approximately seven times the range of the left hand scale

For purposes of applying the variational principle,

$$g(r,z) = \left(\frac{3}{2\pi b^2}\right)^{3/2}\left(\varphi(n) + \frac{r}{b}\theta(n)\right)e^{-\psi(n)r^2/b^2} \tag{26}$$

is selected as an admissible generating function for this illustration. The quantities $\varphi(n)$, $\theta(n)$, and $\psi(n)$ are unknown dimensionless functions of n. To satisfy the initial condition, these functions must have the asymptotic properties that

$$\varphi(n) \to \frac{1}{n^{3/2}}, \quad \psi(n) \to \frac{3}{2n}, \quad \theta(n) \to 0 \tag{27}$$

as $n \to 0^+$. The functional $\Phi[\varphi'(n), \theta'(n), \psi'(n)]$ can then be evaluated. The requirement that the functional must be minimized under variations in $\varphi'(n)$, $\theta'(n)$, and $\psi'(n)$ leads to three ordinary differential equations for the unknown functions of length n. The asymptotic conditions are interpreted as initial conditions at some small but nonzero value of n. The differential equations are complicated in form and no closed form solution is apparent. Recall that the link density $\rho(r_*)$ at current length n depends on the history of the generating function $g(\mathbf{r},m)$ for all lengths in the range $0<m<n$, the derivatives such as $\varphi'(n)$ depend not only on the current values of the unknown functions at length n but also on the "histories" of these functions for all links up to n. However, these equations are readily integrated numerically.This was done for several choices of the initial value of n to ensure that there was no sensitivity to the particular choice. For the case when $v=b^3$, graphs of the functions are shown in Fig. 1 for n in the range $1 \leq n \leq 85$. For large n, $n\psi(n)$ flattens toward a value somewhat below 3/2 whereas $n^{3/2}\varphi(n)$ continues to decline toward zero.

The shape of the generating function for $n=50$ is illustrated in Fig. 2, also for the case when $v=b^3$. While the one term representation of the generating function used in the variational principle is not sufficiently rich in content to produce that function precisely, the result can be expected to qualitatively capture the influence of the excluded volume potential.

The expected value of the end to end distance is given as a function of n by

$$\langle r\rangle = \frac{\int_0^\infty r^3 g(r,n)dr}{\int_0^\infty r^2 g(r,n)dr}. \tag{28}$$

The measure $\langle r\rangle$ of statistical size of the molecule with $v=b^3$ is illustrated in Fig. 3 for n in the range $1 \leq n \leq 85$. For comparison, the corresponding result for $v=0$ as given in (4) is also shown by the dashed line in the same figure.

The size of the configuration of the molecule can also be estimated by means of equilibrium considerations, in the spirit of Flory's original estimates of the excluded volume effect, but this issue is not pursued here.

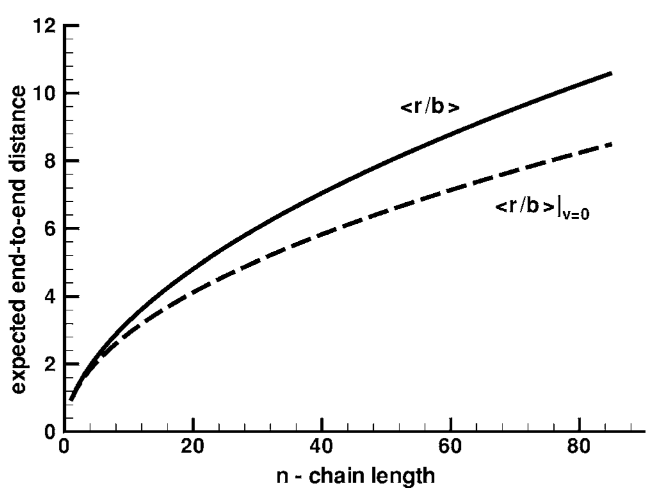


Fig. 3 Estimate of the expected value of the end to end distance $\langle r\rangle$ and the square root of the expected value of the square of the distance $\langle r^2\rangle^{1/2}$ for an isolated molecule as determined by (28) for excluded volume $v=b^3$. Also shown (dashed) for comparison is the expected value of radius when $v=0$

7 Concluding Observations

The objective here has been to introduce a variational approach for computing the generating function for a long flexible molecule and to illustrate this approach in analyzing simple cases. The first configuration was a molecule attached at one end to a plane noninteracting barrier in the presence of a uniform potential field, a case for which a simple closed form solution is known. Not surprisingly, the variational approach reproduced the exact solution when the possibility of doing so was included among the admissible generating functions used in the minimum principle.

The second configuration studied was concerned with the influence of the excluded volume effect on conformations of a long flexible molecule. The calculation was based on the physical ideas introduced by Flory, but the analysis differed from the global approach in two fundamental ways. First, the interactions giving rise to the change in conformation due to excluded volume were considered locally in the mean field analysis rather than globally. Second, the estimate of molecule swelling due to excluded volume was based on the statistical expected value of end to end separation distance rather than on the equilibrium estimate of size based on minimizing global free energy.

The variational approach may have significant potential as a basis for numerical analysis of such configurations. For example, for the spherically symmetric configuration considered above, if $G_m(r)$ is a complete set of functions spanning the space of admissible generating functions then $g(r,n)$ can be assumed in the form

$$g(r,n) = \sum_m \varphi_m(n) G_m(r). \qquad (29)$$

The variational approach then leads to a set of ordinary differential equations for the amplitudes $\varphi_m(n)$.

Acknowledgments

The research support of the National Science Foundation through the Brown University MRSEC Program (No. DMR-0079964) is gratefully acknowledged. I am grateful to my Brown University colleague Tom Powers for helpful discussions.

References

[1] Flory, P. J., 1966, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY.
[2] Edwards, S. F., 1965, "The Statistical Mechanics of Polymers With Excluded Volume," Proc. Phys. Soc., **85**, pp. 613–624.
[3] Keller, J. B., 2004, "Diffusion at Finite Speed and Random Walks," Proc. Natl. Acad. Sci. U.S.A. **101**, pp. 1120–1122.
[4] Chandrasekhar, S., 1943, "Stochastic Problems and Physics and Astronomy," Rev. Mod. Phys. **15**, pp. 1–89.
[5] de Gennes, P. G., 1969, "Some Conformation Problems of Long Macromolecules," Rep. Prog. Phys. **32**, pp. 187–205.
[6] Freed, K. F., 1972, "Functional Integrals and Polymer Statistics," Adv. Chem. Phys. **22**, pp. 1–128.
[7] Grosberg, A. Yu., and Khokhlov, A. R., 1994, *Statistical Physics of Macromolecules*, pp. American Institute of Physics, Woodbury, NY.
[8] Freund, L. B., and Suresh, S., 2003, *Thin Film Materials*, Cambridge University Press, Cambridge, UK.
[9] Shenoy, V. B., Ramasubramanian, A., and Freund, L. B., 2003, "A Variational Approach to Nonlinear Dynamics of Nanoscale Surface Undulations," Surf. Sci. **592**, pp. 365–383.
[10] Suo, Z., 1997, "Motions of Microscopic Surfaces in Materials," Adv. Appl. Mech. **33**, pp. 193–293.
[11] Dolan, A. K., and Edwards, S. F., 1975, "The Effect of Excluded Volume on Polymer Dispersant Action," Proc. R. Soc. London, Ser. A **343**, pp. 427–442.

Robert Pietzsch
e-mail: r.pietzsch@fh-sm.de

Miroslaw Brzoza
e-mail: miroslaw.brzoza@vst.uni-magdeburg.de

Yalçin Kaymak
e-mail: yalcin.kaymak@student.uni-magdeburg.de

Eckehard Specht
e-mail: eckehard.specht@vst.uni-magdeburg.de

Institut für Strömungstechnik und Thermodynamik,
Universitätsplatz 2,
39106 Magdeburg, Germany

Albrecht Bertram
Institut für Mechanik,
Universitätsplatz 2,
39106 Magdeburg, Germany
e-mail: Bertram@mb.uni-magdeburg.de

Simulation of the Distortion of Long Steel Profiles During Cooling

A complex thermomechanical model for simulating the transient fields of the temperature, microstructure, stress, strain, and displacement during quenching of steel profiles is introduced. The thermoplastic material model is formulated on the basis of J_2-plasticity theory with a temperature- and phase fraction-dependent yield limit. Coupling effects such as dissipation, phase transformation enthalpy, and transformation-induced plasticity are considered. The validity of the model is verified by comparing the simulation results with available experimental measurements. The introduced model serves as a basis for optimizing the cooling conditions for reducing residual stresses and distortions. The simulation results for T and L profiles of two different types of steel are described.
[DOI: 10.1115/1.2338050]

1 Introduction

Quenching processes are applied to long steel profiles after the rolling process to produce profiles with reliable service properties. The typical problem is the distortion that is occurring upon residual stresses. Profile distortion should be eliminated as it leads to rework expenditure and creates problems associated with bearing and bundling. Moreover, the residual stresses should be controlled to prevent the crack initiation. The modeling of the cooling process is difficult due to the complicated couplings among different physical and mechanical processes [1]. The main reason for the distortion is the inhomogeneity of the temperature profile during the cooling process which results in residual stresses, varying phase fractions, etc. [2]. The cooling distortion of railway profiles has already been extensively investigated by experiments. An overview of the works can be found in the literature [3], where the experimental results for the residual stresses and deformations after quenching are presented. The distortion and residual stresses in carburized thin strips are investigated by Prantil et al. [4].

The thermoplasticity and metallurgical transformations establish the fundamental part of the problem. The distortion is due to transformation-induced stresses and thermal stresses. The phase transitions, temperature, and stresses must be accurately simulated to determine the distortion. If the temperature and cooling conditions are constant in the axial direction of the profile, all the other fields are also constant. Consequently, a 2D mathematical model with fewer degrees of freedom is enough for the simulations. This reduction of the dimension simplifies not only the visualization of the fields, but also the determination of the optimal cooling conditions. This work indicates that the distortion can be significantly reduced by an intensified cooling of the mass lumped parts of the cross section. The atomized spray quenching technique [5,6], where the spraying water is atomized into fine droplets by means of compressed air, is suitable to intensify the cooling of the mass lumped regions, since the vapor film, which occurs in other quenching techniques, is avoided. In this way the mass lumped regions of the workpieces can be cooled more intensively than the edges. Consequently, a uniform temperature distribution on the surface and a reduced stress distribution inside the body with reduced distortion can be obtained.

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received September 28, 2005; final manuscript received May 21, 2006. Review conducted by Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California–Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME Journal of Applied Mechanics.

2 Mathematical Modeling

In the following sections, the details of the mathematical model for the temperature, microstructure, and stress computations are given. Moreover, the distortion measure and the solution algorithm used for computation are briefly described.

2.1 Temperature Field. The temperature field is modeled by Fourier's law of heat conduction with two additional heat source terms. The first additional term accounts for the transformation enthalpies, and the second term includes the heat generation by mechanical energy dissipation. During the quenching processes, the heat generated due to the mechanical work is negligibly small as compared to the transformation enthalpies. The Fourier's law is

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right)+q_v=\rho c_p\frac{\partial T}{\partial t} \tag{1}$$

with boundary condition

$$k\frac{\partial T}{\partial \mathbf{n}_S}+q_S=0. \tag{2}$$

In Eqs. (1) and (2) k is the heat conductivity, q_v is the heat source, ρ is the mass density, c_p is the specific heat, q_S is the heat flux through the body surface, and $\mathbf{n}_S$ is the surface unit normal. All of the material constants are assumed to be functions of the temperature T and the converted phase fractions ξ_i. The dependency of the material properties on the phase fractions ξ_i can be expressed by the arithmetic, geometric, harmonic, or other means.

The heat flux q_S through the boundary and internal heat source q_v are

$$q_S = \psi \cdot e \cdot \sigma(T^4 - T_\infty^4) + \alpha(T - T_\infty) = \alpha'(T - T_\infty)$$

$$q_v = \chi \sigma_y \dot{\varepsilon}_{\mathrm{eff}}^p + \Sigma_{i=1}^n L_i \cdot \dot{\xi}_i \tag{3}$$

The heat flux q_S is due to thermal radiation and heat convection, where α' is the equivalent convection coefficient, T is the temperature of the body surface, T_∞ is the ambient temperature, ψ is the view factor, e is the emissivity, and σ is the Stefan Boltzmann constant. The internal heat source q_v is composed of latent heat generation and dissipation of mechanical energy into heat, where L_i are the latent heats of transformation, $\dot{\xi}_i$ are the rates of phase conversion, χ is the fraction of mechanical energy converted to heat energy, σ_y is the yield strength, and $\dot{\varepsilon}_{\mathrm{eff}}^p$ is the rate of effective plastic strain.

The finite element method is used to solve Eq. (1). For the case of simplified (plane stress, plane strain, axis symmetric, and beam case) 2D simulations, the six-node isoparametric triangular element is implemented. The temperature field in the element is approximated by

$$T^e \cong \Sigma_{i=1}^N H_i^e T_i^e, \tag{4}$$

where T_i^e are the element nodal temperatures, and N is the number of nodes in the element. The quadratic element interpolation functions are

$$\mathbf{H}^e = \begin{bmatrix} 1-3(r+s)+2(r+s)^2 \\ 4r(1-r-s) \\ r(2r-1) \\ 4 \\ s(2s-1) \\ 4s(1-r-s) \end{bmatrix}. \tag{5}$$

The domain is subdivided into finite elements and the temperature of each element is represented by using element nodal temperatures T_i^e and interpolation functions H_i^e. With this substitution, the differential equation and the boundary condition are not identically satisfied, but are instead equal to an error term. The interpolation functions are used as weight functions and the Galerkin's method is applied to minimize this error term and to get the system of linear algebraic equations

$\mathbf{KT} + \mathbf{C}\dot{\mathbf{T}} = \mathbf{F}$ where

$$\begin{aligned} K_{ij} &= \Sigma_{e=1}^{Nel} \int_{A^e} \nabla H_i^e k \nabla H_j^e h dA + \Sigma_{e=1}^{Nel} \int_{S^e} H_i^e \alpha' H_j^e h dS, \\ C_{ij} &= \Sigma_{e=1}^{Nel} \int_{A^e} H_i^e \rho c_p H_j^e h dA, \\ F_i &= \Sigma_{e=1}^{Nel} \int_{A^e} H_i^e q_v h dA + \Sigma_{e=1}^{Nel} \int_{S^e} H_i^e \alpha' T_\infty h dS \end{aligned} \tag{6}$$

The surface integrals are calculated only for those elements that lay on the boundary. The summation symbol stands for an assembly operator to construct the finite element system matrices.

The material properties of the phase mixture are calculated by using the phase fractions and material properties of constituting phases. Different homogenization methods are applied for different material properties. The generalized homogenization equation is given by

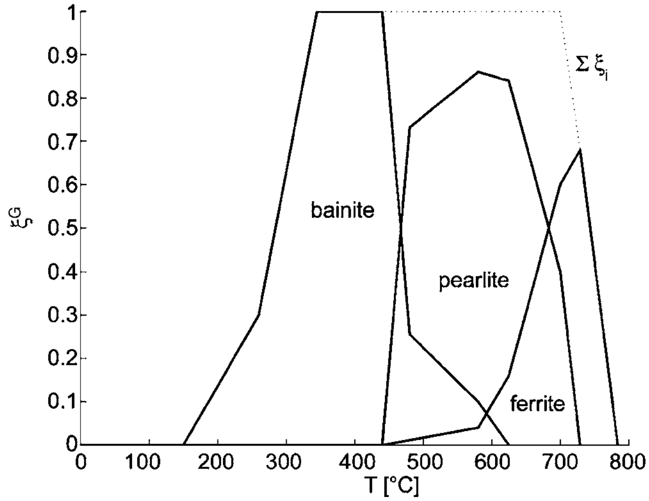


Fig. 1 Equilibrium microstructure fraction for C45 steel

$$Z_{\mathrm{mix}} = \left(\Sigma_{i=1}^n \xi_i Z_i^N\right)^{\frac{1}{N}}, \tag{7}$$

where Z_{mix} stands for the mixture material property estimate and Z_i for the material property of ith phase. The arithmetic mean is assumed for the density ρ, heat conductivity k, or heat capacity c_p with a value of $N=1$. The harmonic mean is assumed for the flow stress σ_y with a value of $N=-1$. The geometric mean, $Z_{\mathrm{mix}} = \exp[\Sigma_{i=1}^n \xi_i \ln(Z_i)]$, is assumed for the bulk modulus κ or the shear modulus μ with a value of $N=0$.

2.2 Microstructure Field. The mathematical theory of the phase transformation kinetics is still not well established. In general, the time-temperature-transformation (TTT) diagrams which are derived from experiments are used. Isothermal transformation (IT) diagrams are convenient, since continuous cooling transformation (CT) diagrams for an arbitrary cooling process can be calculated from IT diagrams by the methods described in literature [7–11].

There are two different types of phase transformation, namely diffusional and martensitic transformations. For diffusional transformations, the formation of the new phase is described by the Kolmogorov-Johnson-Mehl-Avrami [12] equation,

$$\xi_i = \xi_i^{eq}[1 - \exp\{-a_i \cdot (t^{\mathrm{iso}})^{b_i}\}], \tag{8}$$

under isothermal conditions with temperature-dependent transformation parameters a_i and b_i for each phase. After an infinitely long time, the phases achieve their equilibrium fractions ξ_i^{eq}. These equilibrium fractions are shown in Fig. 1 for the steel C45.

The transformation limit curves in the IT diagrams contain two C-shaped curves for each phase transition. One is for the beginning of the transformation and the other is for its end. Each curve is approximated by the equation

$$\tau(T) = d_0 \cdot \exp \frac{d_1 + d_2(T_U - T)^{-n}}{T - T_L}, \tag{9}$$

which is proposed by Hougardy [13]. Here, T_L and T_U are the lower and upper asymptotes of the curve, respectively. The other parameters d_0, d_1, and d_2 are uniquely defined if three points $\{(t_1,T_1),(t_2,T_2),(t_3,T_3)\}$ on the limit curve and the exponent n are given

$$
\begin{aligned}
d_0 &= t_1 \exp\left[\frac{(T_3-T_L)(\varphi_{13}-\varphi_{23})\ln\left(\frac{t_1}{t_3}\right)+(T_L-T_2)(\varphi_{12}-\varphi_{23})\ln\left(\frac{t_1}{t_2}\right)}{(T_1-T_2)\varphi_{12}+(T_3-T_1)\varphi_{13}+(T_2-T_3)\varphi_{23}}\right],\\
d_1 &= \frac{\varphi_{23}[T_2T_3-T_L(T_2+T_3)+T_L^2]-\varphi_{12}[T_1T_2-T_L(T_1+T_2)+T_L^2]}{(T_1-T_2)\varphi_{12}+(T_3-T_1)\varphi_{13}+(T_2-T_3)\varphi_{23}}\ln\left(\frac{t_1}{t_2}\right)\\
&\quad+\frac{\varphi_{13}[T_1T_3-T_L(T_1+T_3)+T_L^2]-\varphi_{23}[T_2T_3-T_L(T_2+T_3)+T_L^2]}{(T_1-T_2)\varphi_{12}+(T_3-T_1)\varphi_{13}+(T_2-T_3)\varphi_{23}}\ln\left(\frac{t_1}{t_2}\right),\\
d_2 &= \frac{(T_3-T_L)(T_2-T_1)\ln\left(\frac{t_1}{t_2}\right)+(T_3-T_1)(T_L-T_2)\ln\left(\frac{t_1}{t_2}\right)}{(T_1-T_2)\varphi_{12}+(T_3-T_1)\varphi_{13}+(T_2-T_3)\varphi_{23}}\varphi_{123},
\end{aligned}
\tag{10}
$$

where

$$
\begin{aligned}
\varphi_{12} &= \{(T_1-T_U)(T_2-T_U)\}^n,\\
\varphi_{13} &= \{(T_1-T_U)(T_3-T_U)\}^n,\\
\varphi_{23} &= \{(T_2-T_U)(T_3-T_U)\}^n,\\
\varphi_{123} &= \{-(T_1-T_U)(T_2-T_U)(T_3-T_U)\}^n.
\end{aligned}
\tag{11}
$$

The transformation beginning curve gives the beginning time t_B^{iso} when 1% of the new phase fraction is formed. The other curve gives the transformation end time t_E^{iso} when 99% of the equilibrium phase fraction ξ^{eq} is formed at a constant temperature T

$$
\begin{aligned}
1-\exp\{-a\cdot(t_B^{\text{iso}})^b\} &= 0.01\\
1-\exp\{-a\cdot(t_E^{\text{iso}})^b\} &= 0.99\xi^{eq}.
\end{aligned}
\tag{12}
$$

The transformation parameters a and b are obtained by solving Eq. (12)

$$
\begin{aligned}
a &= -\frac{1}{(t_E^{\text{iso}})^b}\ln(1-0.99\xi^{eq})\\
b &= \frac{1}{\ln\left(\frac{t_S^{\text{iso}}}{t_E^{\text{iso}}}\right)}\ln\left[\frac{\ln(1-0.01)}{\ln(1-0.99\xi^{eq})}\right],
\end{aligned}
\tag{13}
$$

at a constant temperature T, if the beginning time t_B^{iso} and the end time t_E^{iso} are known. The variations of these factors with respect to the temperature are shown in Fig. 2 for C45.

In the continuous cooling case, the transformation begins at the incubation time t_B^{inc} and it ends after time t_E^{inc}, which are obtained by Scheil's sum [14]

$$
\int_0^{t_{S,E}^{\text{inc}}}\frac{1}{\tau_{B,E}[T(t)]}dt = 1,
\tag{14}
$$

on discrete basis

$$
\Sigma_{i=1}^{n}\frac{\Delta t}{\tau_{B,E}(T_i)} = 1,
\tag{15}
$$

for an arbitrary cooling curve $T(t)$.

The calculations are much easier for martensitic transformations that do not depend directly on the time but only on the temperature. The martensite phase fraction is calculated according to Koistinen-Marburger [15] equation

$$
\xi_M = \xi_A\{1-\exp[-k_M(T_{Ms}-T)]\},
\tag{16}
$$

where ξ_A is the austenite phase fraction at the beginning of martensitic transformation, k_M is a stress-dependent transformation constant, and T_{Ms} is the temperature at which the martensitic transformation starts. The effects of the stress state on the transformation kinetics are discussed by Denis et al. [16].

2.3 Stress/Strain Field. An isotropic thermoplastic material behavior with a temperature and phase fraction dependent constitutive relation is employed. The basic assumptions are as follows:

A1: total deformations are small,

$$
\mathbf{E} = 1/2(\text{grad}(\mathbf{u})+\text{grad}^T(\mathbf{u})) \text{ and } \mathbf{E} = \mathbf{E}^{T,\xi}+\mathbf{E}^p+\mathbf{E}^{\text{trip}}+\mathbf{E}^e;
\tag{17}
$$

A2: the material is isotropic,

$$
\mathbf{T} = \mathbf{C}[\mathbf{E}^e] = \kappa\,\text{tr}(\mathbf{E}-\mathbf{E}^{T,\xi})\mathbf{I}+2\mu(\mathbf{E}'-\mathbf{E}^{\text{trip}}-\mathbf{E}^p);
\tag{18}
$$

A3: plastic deformations are incompressible,

$$
\text{tr}(\mathbf{E}^p) = 0, \quad \mathbf{E}'^p = \mathbf{E}^p;
\tag{19}
$$

A4: the material behavior is rate independent,
A5: the flow condition is of von Mises type,

$$
\phi(\mathbf{T}',\varepsilon^p,T,\xi_i) = \|\mathbf{T}'\| - \sqrt{2/3}\sigma_y(\varepsilon^p,T,\xi_i)
\tag{20}
$$

A6: an associative plastic model is used,

$$
\dot{\mathbf{E}}^p = \lambda\frac{\partial\varphi}{\partial T} = \lambda\frac{\mathbf{T}'}{\|\mathbf{T}'\|} = \lambda\mathbf{N}'; \text{and}
\tag{21}
$$

A7: linear isotropic hardening behavior is assumed,

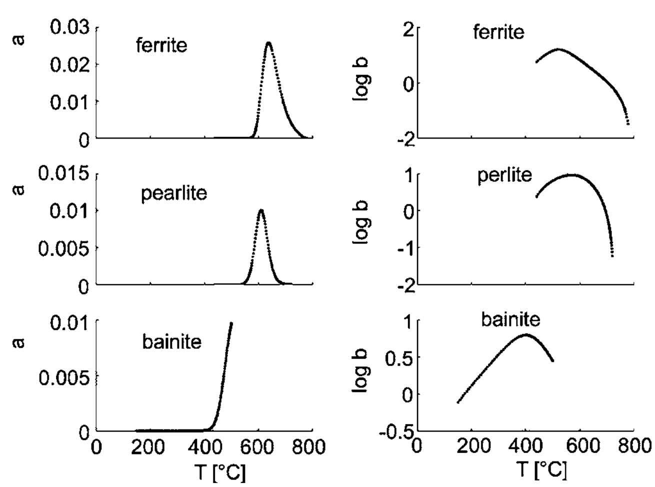


Fig. 2 Avrami constant of transformation kinetics

$$\sigma_y = \sigma_{yvmix}^{HM} + H_{mix}^{HM}\varepsilon^p, \dot{\varepsilon}^p = \sqrt{2/3}\lambda. \quad (22)$$

The prime "$\mathbf{N}'$" over a tensor represents its deviator, "$\mathrm{tr}(\mathbf{E})$" is the trace operator.

Material equations

$$\text{loading condition: } \dot{\phi}|_{\varepsilon^p=\text{const.}} \frac{\partial \phi}{\partial \mathbf{T}'} \cdot \dot{\mathbf{T}}' + \frac{\partial \phi}{\partial T}\dot{T} + \Sigma_{i=1}^{n}\frac{\partial \phi}{\partial \xi_i}\dot{\xi}_i = \begin{cases} >0 \text{ loading} \\ =0 \text{ neutral} \\ <0 \text{ unloading} \end{cases} \quad (23)$$

$$\text{consistency condition: } \dot{\phi} = \frac{\partial \phi}{\partial \mathbf{T}'} \cdot \dot{\mathbf{T}}' + \frac{\partial \phi}{\partial \varepsilon^p}\dot{\varepsilon}^p + \frac{\partial \phi}{\partial T}\dot{T} + \Sigma_{i=1}^{n}\frac{\partial \phi}{\partial \xi_i}\dot{\xi}_i = 0. \quad (24)$$

For a plastic loading increment, the unknown plastic multiplier λ can be obtained from the consistency condition Eq. (24). The first term in the consistency condition is computed from the elastic law Eq. (18)

$$\frac{\partial \varphi}{\partial \mathbf{T}'} \cdot \dot{\mathbf{T}}' = \mathbf{N}' \cdot \{2\dot{\mu}(\mathbf{E} - \mathbf{E}^{\text{trip}} - \mathbf{E}^p) + 2\mu(\dot{\mathbf{E}} - \dot{\mathbf{E}}^{trip} - \dot{\mathbf{E}}^p)\}. \quad (25)$$

Equations (20)–(22) are used to calculate the last three terms in the consistency condition. Finally, the unknown plastic multiplier is

$$\lambda = \frac{2\dot{\mu}\mathbf{N}' \cdot (\mathbf{E} - \mathbf{E}^{\text{trip}}) + 2\mu\mathbf{N}' \cdot (\dot{\mathbf{E}} - \dot{\mathbf{E}}^{\text{trip}}) - \sqrt{\frac{2}{3}}\left(\frac{\partial \sigma_y}{\partial T}\dot{T} + \Sigma_{i=1}^{n}\frac{\partial \sigma_y}{\partial \xi_i}\dot{\xi}_i\right)}{2\mu + \frac{2}{3}H}. \quad (26)$$

A linear relation between temperature and thermal strain is used in general. However, this method contains a linearization [17] and sometimes requires a conversion of the reference temperatures. An alternative method is based on the temperature dependence of the density [18]. The thermal and transformation induced strains are combined in an isotropic tensor

$$\mathbf{E}^{T,\xi} = \left(\sqrt[3]{\frac{\rho_R}{\rho}} - 1\right)\mathbf{I}, \quad (27)$$

where ρ_R denotes the reference density and ρ is the mixture density. For an n-phase mixture, the thermal and transformation induced strain rate becomes

$$\dot{\mathbf{E}}^{T,\xi} = \frac{d\mathbf{E}^{T,\xi}}{d\rho}\left(\frac{\partial \rho}{\partial T}\dot{T} + \Sigma_{i=1}^{n}\frac{\partial \rho}{\partial \xi_i}\dot{\xi}_i\right). \quad (28)$$

The transformation induced plasticity (TRIP) is an additional irreversible deformation, which occurs during the phase transition. The TRIP flow occurs in the direction of the stress deviator $\mathbf{T}'$, even if the global stress does not exceed the yield limit of a single phase. The flow rate is proportional to the phase transition rate and the stress deviator

$$\dot{\mathbf{E}}^{\text{trip}} = -\frac{3}{2}\Lambda \ln(\xi)\dot{\xi}\mathbf{T}'. \quad (29)$$

This equation characterizes the macroscopic material behavior, which is determined by the micromechanical processes. Moreover, the proportionality factor depends on the fraction of the transformed phase and Greenwood Johnson coefficient Λ [19], which must be determined experimentally. The equation for the factor is given as

$$\Lambda_i = \frac{5(\rho_A - \rho_i)}{6\rho_A\sigma_{yA}}, \quad (30)$$

where the density ratio is a measure for the volume ratio of the converting phases (e.g. austenite, pearlite) and σ_{yA} the yield limit of the softer phase (mostly austenite) at transformation temperature.

The relationship between the stress tensor $\mathbf{T}$ and the elastic strain tensor $\mathbf{E}^e$ is given by the elastic law Eq. (18). The tangential elastoplastic material operator is

$$\mathbf{C}^{ep} = \left.\frac{\partial \mathbf{T}}{\partial \mathbf{E}^{ep}}\right|_{T,\xi=\text{const.}} = \mathbf{C}^e - \frac{2\mu}{1 + \frac{H}{3\mu}}\mathbf{N}' \otimes \mathbf{N}' = 3\kappa\mathbf{I} \otimes \mathbf{I} + 2\mu\left(\mathbf{I} - \tfrac{1}{3}\mathbf{I} \otimes \mathbf{I}\right) - \frac{2\mu}{1 + \frac{H}{3\mu}}\mathbf{N}' \otimes \mathbf{N}', \quad (31)$$

where $\mathbf{I}$ is fourth order identity tensor and $\mathbf{I} \otimes \mathbf{I}$ the dyadic product of second order identity tensors. Quadratic convergence is provided for plastic deformations with the elastoplastic tangential operator.

2.4 Description of the Distortion. The distortion of long steel profiles is intended to be simulated by the model described above. The distortion cannot be adequately described by the deflection of the profile as it depends on the length of the profile. Hence, the distortion is defined in terms of curvature of profile. The deflection Δ from the neutral axis is approximated by a parabola

$$\Delta = p \cdot z^2. \quad (32)$$

The curvature related factor p serves as a measure for the distortion.

2.5 Solution Algorithm. A time integration scheme that involves equilibrium iterations at each time step is applied to solve the coupled field problem described above. In an iteration step, first the temperature, phase transitions and stress field are calculated, respectively. The phase transitions are calculated by using the current iteration temperature and the temperature from the previous time step. When the values of temperature, microstructure fractions and stresses have converged, the iteration for the next time step is started.

3 Experimental Verification of Mathematical Model

The simulation results were compared with the experimental measurements in order to validate the mathematical model introduced above. As no proper measurements for profiles are known, we used the available experimental results for notched shafts and

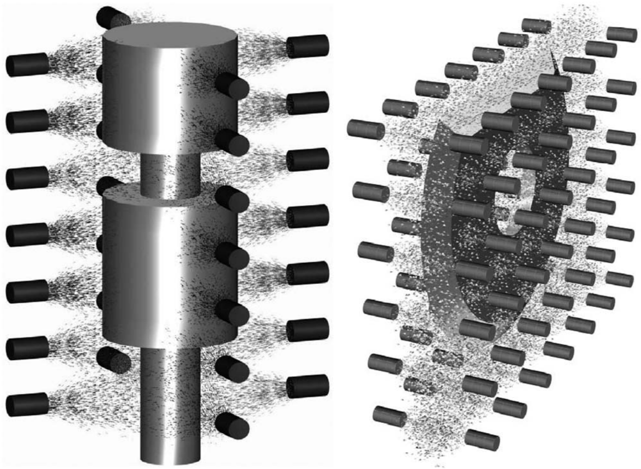


Fig. 3 The shaft (α=450 W/m^2/K) and the disk (α=200 W/m^2/K) with the quenching nozzle field

cutting disks as shown in Fig. 3. The samples were homogeneously heated in a roller-hearth kiln up to a temperature value of 850°C in a nitrogen atmosphere. After the heating, the samples fell into a catching cage, in which they were precisely positioned and quenched in a nozzle field. The nozzle spacing, flow rate, and distance from the sample could be adjusted. The heat transfer coefficient could be controlled in this way and was well known from other specific measurements. Samples were available in two geometries, i.e., shafts; 18 mm in diameter with 60 mm of length, and 36 mm in diameter with 120 mm of length, and disks; 90 mm and 120 mm of outer diameter. As a material, 100Cr6 was used because the material properties are well known and available with high accuracy as given in Table 1. Further details of the experiments can be found in the literature [20].

The temperatures were measured with thermocouples of type K and class 1, which have an accuracy of 1.5°C for 0–375°C and 0.004T °C for 375–1000°C. The thermocouples were placed in the drilled holes as shown in Fig. 4. The regions 2 and 3 have the highest and lowest temperatures, respectively. As an example, the measured temperatures at these two points are compared to the calculated ones for the bigger shaft. As expected, the curves coincide for an appropriate heat transfer coefficient.

The calculated stresses are compared to the measured stresses for the bigger shaft in Fig. 5. The measurements are performed at five sampling points in longitudinal direction with a computer-controlled diffractometer. The simulation result matches well the measurements as it can be seen in Fig. 5.

Table 2 presents the calculated and measured microstructure fractions for the shafts and disks. The probes were taken from the positions at the surface given in the upper-left corner of Table 2. The microstructure fractions were measured by the point analysis method [21] with an accuracy of 1% fraction. The calculated martensite fraction values were higher for the shafts but smaller for the disks than those measured. The calculated bainite fraction was slightly lower than the measured values for the small shaft, and for all other cases it was slightly higher. Calculated and measured pearlite fractions occurred only for the shafts and these values are similar. The discrepancies with the experimental measurements are less than 10% mostly and they are acceptable for such simulations. Such a behavior might be due to small fluctuations of chemical compositions which are listed in Table 3.

The distortion of the cutting disk was measured by the change in the radius of a hole with accuracy 1.3 μm. Figure 6 shows the radial displacement of the hole in the angular position. As can be seen from Fig. 6, the simulation results fit the measured distortion of the disks.

The overall performance of the model is quite satisfactory, and the finite element simulation results are in agreement with the measurements. The mesh convergence tests have been performed and it is observed that further refinements of the mesh have neg-

Table 1 Material properties for 100Cr6

Pearlite & Martensite							
Temperature (°C)	Elasticity Modulus (GPa)	Poisson's Ratio	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Conductivity (W/m/K)	Heat Capacity (J/kg/K)
0	212	0.283	479- 1116	21- 63	7820	40.0	410
200	202	0.291	472- 914	26- 40	7740	39.0	550
400	186	0.299	400- 712	22- 17	7770	35.5	620
600	166	0.307	277- 510	19- 0	7620	29.0	790
800	141	0.315	118- 0	11- 0	7530	20.0	1160
1000	112	0.323	0- 0	0- 0	7460	8.5	1850
Austenite							
Temperature (°C)	Elasticity Modulus (GPa)	Poisson's Ratio	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Conductivity (W/m/K)	Heat Capacity (J/kg/K)
0	212	0.283	300	12.0	7980	14.5	450
200	202	0.291	270	10.4	7900	17.2	510
400	186	0.299	240	8.8	7820	19.9	570
600	166	0.307	210	7.2	7720	22.6	600
800	141	0.315	180	5.6	7620	25.3	630
1000	112	0.323	150	4.0	7530	28.0	650

Pearlite & Martensite							
Temperature (°C)	Elasticity Modulus (GPa)	Poisson's Ratio	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Cunductivity (W/m/K)	Heat Capacity (J/kg/K)
0	215	83	550-1000	5.1-12.5	7850	48.8	410
200	196	75	420-940	5.8-10.0	7790	46.5	550
400	177	67	470-800	4.9-7.5	7710	41.0	620
600	158	59	470-0	2.6-0	7630	35.0	790
800	140	52	180-0	0-0	7530	26.0	1160
1000	121	44	0-0	0-0	7430	17.0	1850
Austenite							
Temperature (°C)	Elasticity Modulus (GPa)	Poisson's Ratio	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Cunductivity (W/m/K)	Heat Capacity (J/kg/K)
0	200	80	250	4.2	7940	14.8	450
200	160	64	200	4.0	7860	17.3	510
400	120	48	150	3.7	7790	19.6	570
600	80	32	100	3.1	7700	21.9	600
800	40	16	50	2.3	7600	24.3	630
1000	0	0	0	1.4	7500	26.7	650

Fig. 4 Calculated and measured temperature profiles for the bigger shaft

ligibly small influence on the simulation results. Hence, this model can be applied to other samples such as long profiles.

4 Profile Cooling Simulations

The model is now used to investigate the distortion behavior of the profiles. *T* and *L* profile simulations have been carried out to investigate their cooling distortion behaviors. The geometrical dimensions, finite element (FE) mesh, and the observation points are shown in Fig. 7. The profile is made of C45 steel, which is a typical material for such profiles. The temperature-dependent flow curves can be found in the literature [22], and the temperature-dependent thermophysical material properties are given in Table 4. This steel, unlike high-strength steels, completely loses its elastic characteristics above a temperature of about 600°C. In order to demonstrate the plasticity effects, the cooling conditions have to be much more intensive as compared to the natural cooling conditions. The heat transfer coefficient and the emissivity are taken as 1000 $\mathrm{W/m^2/K}$ and 0.7, respectively. Radiation constitutes approximately 10% of the heat transfer. The cooling process starts at 600°C. Therefore, there is no phase transformation effect.

The residual axial stress in the cross section after the cooling is shown in Fig. 8. The outer region is subjected to compressive stresses and the core is subjected to tensile ones. The evolution of the plastic zones can be observed in Fig. 9 during different stages of cooling. At the beginning, the external region shortens faster due to the temperature difference between the core and surface, and the external regions yields under tension. However, it returns later to the elastic range, while the compressive stress in the core region reaches the yield limit. Finally, the exterior region comes under compression, while the compressive stress in the core becomes tensile stress.

In practice, 75–95% of the plastic work is converted into heat. The unconverted part of the dissipated energy is stored in the form of micromechanical stresses in crystal structure. The dissipation of the mechanical energy significantly affects the temperature field only in special cases such as metal forming. The dissipation process and the corresponding heat generation are shown as functions of the cooling time in Fig. 10 for the points described in Fig. 7. It is obvious that the heat generation effect is negligibly small as compared to the convection and radiation in the case of quenching.

The second simulation is the cooling of a standard *L* profile with the dimensions of $L120\times 12$ according to DIN 1028. The profile cross section is simply symmetric and prone to distortions. The geometrical dimensions, FE mesh, and observation points are shown in Fig. 11. The behavior of the state variables at the observation points during cooling is demonstrated in Fig. 12. The left column represents plots for the hypoeutectic steel C45, the material properties of which are listed in Table 4, and the right ones for the eutectic steel C80, listed in Table 5. The heat transfer coefficient is $\alpha_0=10\ \mathrm{W/m^2/K}$, and the value of emissivity is taken as 0.7 for all surfaces. It is obvious that the temperature at the outer

Pearlite & Martensite							
Temperature (°C)	Elasticity Modulus (GPa)	Poisson's Ratio	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Cunductivity (W/m/K)	Heat Capacity (J/kg/K)
0	220	83	650-1000	4.0-60	7860	50	410
200	200	75	625-945	4.7-50	7790	45	550
400	175	67	450-775	4.5-40	7720	38	620
600	133	59	375-505	1.8-0	7650	32	790
800	80	52	150-150	0-0	7570	29	1160
1000	13	44	0-0	0-0	7490	27	1850
Austenite							
Temperature (°C)	Elasticity Modulus (GPa)	Poisson's Ratio	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Cunductivity (W/m/K)	Heat Capacity (J/kg/K)
0	220	80	250	1.4	7740	6	450
200	176	64	200	1.4	7690	10	510
400	132	48	150	1.5	7640	14	570
600	88	32	100	1.5	7590	18	600
800	44	16	50	1.4	7540	22	630
1000	0	0	0	1.3	7490	26	650

Fig. 5 Calculated and measured internal stresses for the bigger shaft

Table 2 Microstructure phase fractions from measurements (*m*) and simulations (*s*)

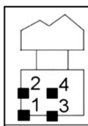
	small shaft						big shaft					
	pearlite [%]		bainite [%]		martensite [%]		pearlite [%]		bainite [%]		martensite [%]	
	m	s	m	s	m	s	m	s	m	s	m	s
point 1	-	-	33	31	67	69	94	86	-	5	6	9
point 2	-	-	33	31	67	69	97	86	-	9	3	5
point 3	-	-	22	18	78	82	100	90	-	10	-	-
point 4	-	-	26	22	74	78	100	89	-	11	-	-
	small disk						big disk					
	pearlite [%]		bainite [%]		martensite [%]		pearlite [%]		bainite [%]		martensite [%]	
	m	s	m	s	m	s	m	s	m	s	m	s
point 1	-	-	4	12	96	88	-	-	12	22	88	78
point 2	-	-	5	18	95	82	-	-	16	30	84	70

Table 3 Chemical compositions of shaft steel, disk steel, and steel in the model

	C (%wt.)	Si (%wt.)	Mn (%wt.)	P (%wt.)	S (%wt.)	Cr (%wt.)	Ni (%wt.)	Cu (%wt.)
Shaft	0.97	0.20	0.29	0.01	0.02	1.46	-	0.22
Disc	0.92	0.23	0.35	0.01	0.01	1.48	0.05	0.09
Model	0.95-1.10	0.10-0.35	0.20-0.40	<0.030	<0.025	1.35-1.60	<0.30	<0.30

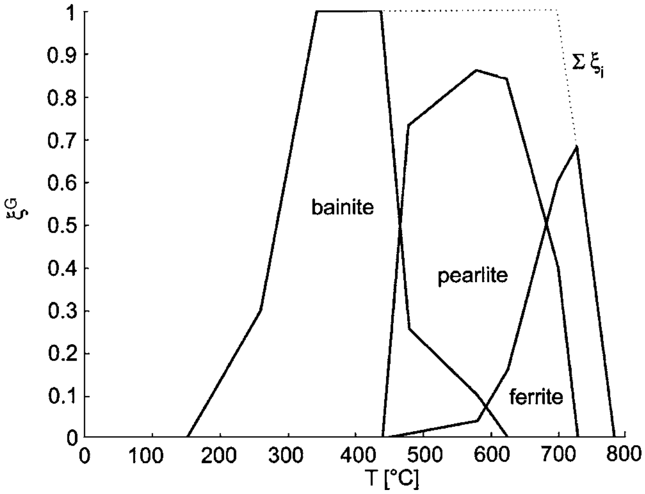


Fig. 6 Measured and calculated distortion profiles for the bigger disk

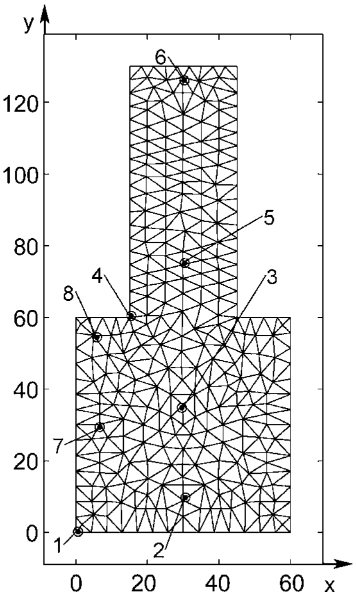


Fig. 7 Finite-element mesh and geometric dimension in (mm) of *T* profile

observation points 1 and 4 drops at the fastest rate, and for the interior point 3 at the slowest rate. Accordingly, the regions lying on the outer parts experience the phase transition earlier than the core. In the third row of diagrams the axial stress is shown. This stress usually changes its direction during the phase transition and achieves a maximum value at the end of the phase transition. The TRIP strains in axial direction are shown in the fourth row. It is noticed that the external regions are subjected to the largest plastic deformations. The distortion of the profiles which is defined in Eq. (32) is shown in the fifth row. The distortion changes direction during cooling, and finally a positive plastic distortion remains. It is also noticed that the deformations are higher for C80 than that of C45.

Figure 13 shows the residual axial stress in the *L* profile. The bold line is the border between the tension region and the compression region. The thermal contraction in the core results in a high tensile stress in this region. The distortion is obviously determined by the cooling of the mass-lumped region. In order to reduce the distortion, the cooling rate around the point 2 should be increased as schematically described in Fig. 14. Such a heat transfer process can be achieved if the point 1 is cooled with an air nozzle from the outside. The distortions for different heat transfer coefficients are shown in Fig. 15. The maximum heat transfer should occur at the vertex in order to reduce the distortion. In some extreme cases, even a distortion in the opposite direction can be obtained [18]. The same effect can be achieved by placing the air nozzle towards point 2 from inside.

5 Conclusions

The experiments and examples demonstrate that:

Table 4 Material properties for C45

Pearlite & Martensite							
Temperature (°C)	Elasticity Modulus (GPa)	Shear Modulus (GPa)	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Conductivity (W/m/K)	Heat Capacity (J/kg/K)
0	215	83	550- 1000	5.1-12.5	7850	48.8	410
200	196	75	420- 940	5.8-10.0	7790	46.5	550
400	177	67	470- 800	4.9- 7.5	7710	41.0	620
600	158	59	470- 0	2.6- 0.0	7630	35.0	790
800	140	52	180- 0	0.0- 0.0	7530	26.0	1160
1000	121	44	0- 0	0.0- 0.0	7430	17.0	1850
Austenite							
Temperature (°C)	Elasticity Modulus (GPa)	Shear Modulus (GPa)	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Conductivity (W/m/K)	Heat Capacity (J/kg/K)
0	200	80	250	4.2	7940	14.8	450
200	160	64	200	4.0	7860	17.3	510
400	120	48	150	3.7	7790	19.6	570
600	80	32	100	3.1	7700	21.9	600
800	40	16	50	2.3	7600	24.3	630
1000	0	0	0	1.4	7500	26.7	650

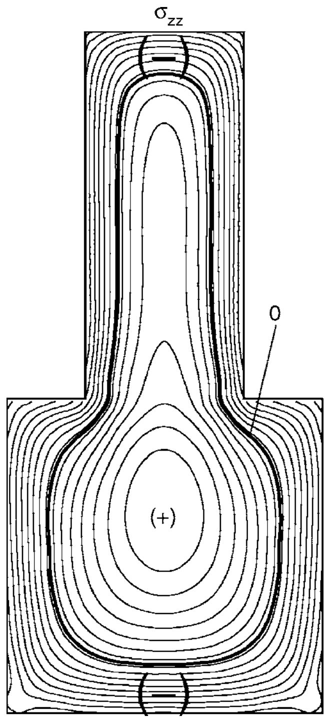


Fig. 8 The residual axial stress in the *T* profile after cooling

- The Avrami equation together with Scheil's additivity rule estimates well the phase evaluations.
- The cooling of steel workpieces can be effectively controlled in a nozzle field.
- The temperature field can be accurately simulated when an appropriate heat transfer coefficient is used.
- The plastic deformation always starts from the surface and penetrates to the core henceforth.
- The heat generation by plastic dissipation is negligible as compared to the latent heat of the phase transition.
- In general the core remains under tensile and the shell remains under compressive residual stresses at the end.

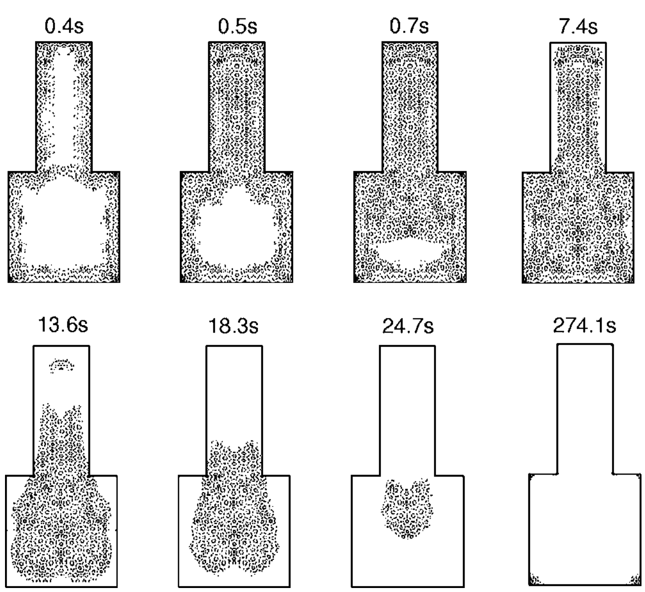


Fig. 9 Evolution of plastic zones during cooling of *T* profile

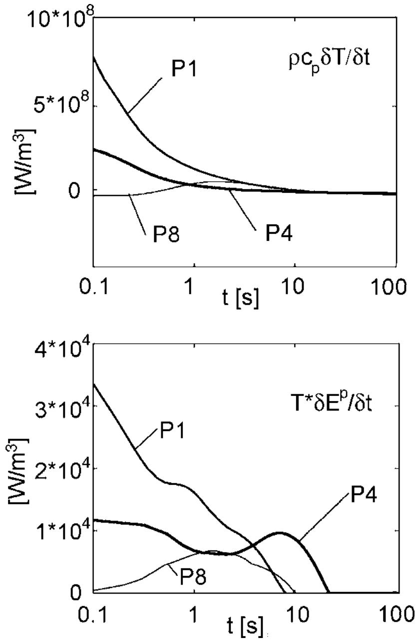


Fig. 10 Dissipation (bottom) and corresponding internal heat generation (top)

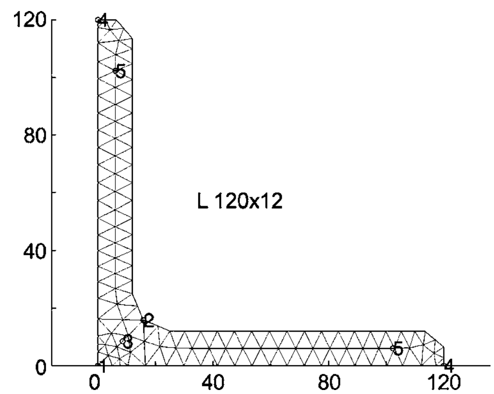


Fig. 11 Finite-element mesh and geometric dimension in (mm) of *L* profile

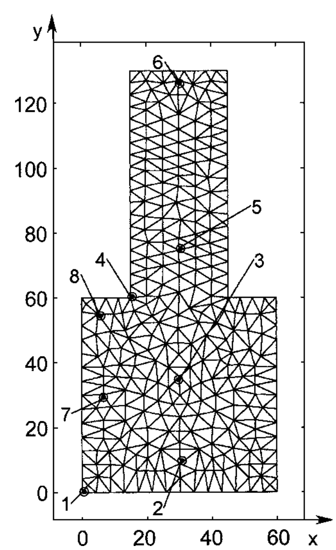


Fig. 12 Behavior of state variables during the cooling of *L* profiles

Table 5 Material properties for C80

Pearlite & Martensite							
Temperature (°C)	Elasticity Modulus (GPa)	Shear Modulus (GPa)	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Conductivity (W/m/K)	Heat Capacity (J/kg/K)
0	220	83	650-1000	4.0- 60	7860	50	410
200	200	75	625- 945	4.7- 50	7790	45	550
400	175	67	450- 775	4.5- 40	7720	38	620
600	133	59	375- 505	1.8- 0	7650	32	790
800	80	52	150- 150	0.0- 0	7570	29	1160
1000	13	44	0- 0	0.0- 0	7490	27	1850
Austenite							
Temperature (°C)	Elasticity Modulus (GPa)	Shear Modulus (GPa)	Yield Strength (MPa)	Hardening Modulus (GPa)	Density (kg/m^3)	Heat Conductivity (W/m/K)	Heat Capacity (J/kg/K)
0	220	80	250	1.4	7740	6	450
200	176	64	200	1.4	7690	10	510
400	132	48	150	1.5	7640	14	570
600	88	32	100	1.5	7590	18	600
800	44	16	50	1.4	7540	22	630
1000	0	0	0	1.3	7490	26	650

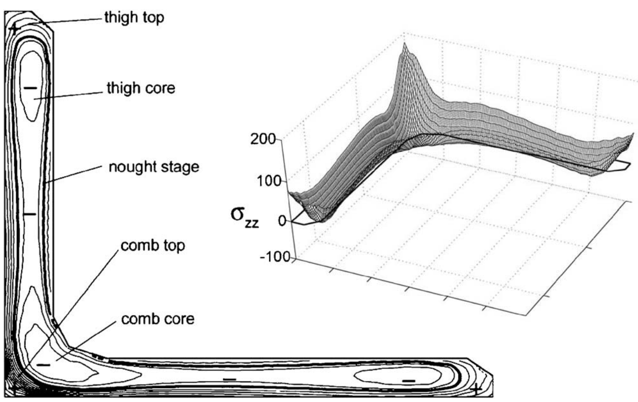


Fig. 13 Residual axial stress in the *L* profile after the cooling

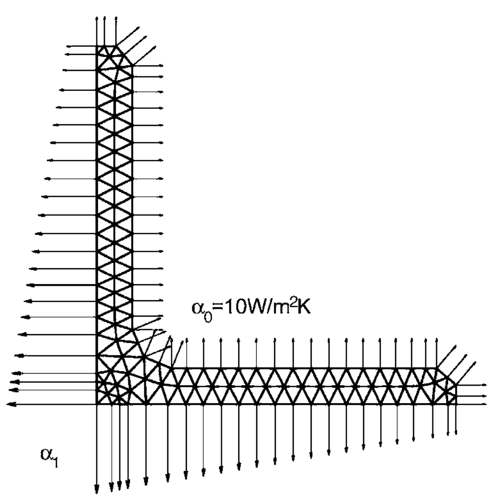


Fig. 14 Schematic representation of cooling strategy

- The distortion changes direction and reaches its maximum value during the phase transition.
- The stresses and strains also reach their maximum value during the phase transition.
- The distortion is mainly due to TRIP strains, which are related to phase transition phenomena.
- The profiles with smaller section sizes have more distortion than those with bigger section sizes, which has been already ascertained in practice.
- The results support the idea that distortion can be reduced by increasing the cooling at mass lumped regions.

Nomenclature

$\mathbf{C}$ = capacitance matrix for heat transfer
$\mathbf{C}, \mathbf{C}^{ep}$ = fourth order elastic and elastoplastic constitutive tensors, respectively
$\mathbf{E}$ = linearized strain tensor
$\mathbf{F}$ = force vector for heat transfer
$\mathbf{H}^e$ = element interpolation functions
$\mathbf{I}, \mathbf{I}$ = fourth and second order identity tensor, respectively
$\mathbf{K}$ = conductivity matrix for heat transfer (Eq. (6)) stiffness matrix for displacement solution

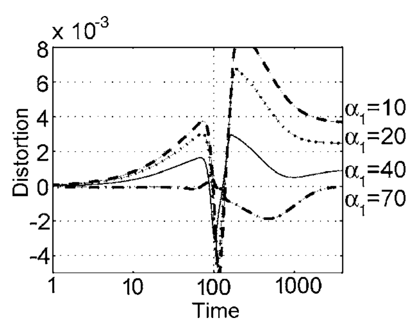


Fig. 15 The distortion of *L* profiles for different cooling conditions (according to Eq. (32))

$\mathbf{N}$ = flow surface unit normal
$\mathbf{n}_S$ = surface unit normal
$\mathbf{q}_S$ = heat flux vector
$\mathbf{u}$ = displacement vector
$\mathbf{T}$ = nodal temperature vector (Eq. (6)) Cauchy stress tensor
a, b = Avrami parameters
A^e = cross section area of an element
c_p = specific heat capacity
d_0, d_1, d_2 T_U, and T_L = Hougardy parameters
e = emissivity
H = hardening parameter
H_i^e = element interpolation function for ith node
h = element thickness in the third dimension
k_M = martensitic transformation constant
k = heat conductivity
L_i = latent heat of ith phase conversion
n = exponent in an expression, Hougardy parameter (Eqs. (9) and (11)) number of phases in mixture (Eqs. (3) and (8))
N = exponent in the expression to define mixture rule (Eq. (7)) number of nodes per element (Eq. (4))
p = distortion defined by curvature
q_v = heat generation rate in the body
r, s = local (natural) coordinates
S^e = convective boundary line for a boundary finite element
t = time
T_i^e = temperature at ith node
T = temperature
T_∞ = ambient temperature
T_{Ms} = martensitic transformation start temperature
x, y, z = global cartesian coordinates
Z_{mix}, Z_i = material property estimate for mixture and its value for ith phase, respectively
α = convection coefficient
α' = equivalent convection coefficient (including radiation)
Δ = profile deflection
$\varepsilon_{\text{eff}}^p$ = equivalent plastic strain
ε = strain
φ = dummy variable in Hougary parameters calculation
ϕ = flow condition
ψ = radiation view factor
κ = bulk modulus
Λ = Greenwood Johnson factor
λ = plastic multiplier
μ = shear modulus
ρ = mass density
σ = Stefan-Boltzmann constant (Eq. (3)) stress
σ_y = yield stress
σ_{eff} = effective stress
ξ_i = phase fraction of ith phase
τ = TTT-curve defining function
χ = fraction of mechanical energy converted to heat energy

References

[1] Fischer, F. D., Oberaigner, E. R., Tanaka, K., and Nishimura, F., 1998, "Transformation Induced Plasticity Revised an Updated Formulation," Int. J. Solids Struct., **38**(18), pp. 2209–2227.
[2] Berns, H., 1977, "Distortion of Steel due to Heat Treatment," Verzug von Stählen infolge Wärmebehandlung, Werkstofftechnik, **8**, pp. 149–157.
[3] Hinteregger, E., 1990, "Internal Stresses and Deformations of Railway Profiles after Rolling Before Bundling," (German: Eigenspannungen und Verformungen in Schienen nach dem Walzen vor dem Richten), Ph.D. thesis, Montanuniversität Leoben, Austria.
[4] Prantil, V. C., Callabresi, M. L., Lathrop, J. F., Ramaswamy, G. S., and Lusk, M. T., 2003, "Simulating Distortion and Residual Stresses in Carburized Thin Strips," ASME J. Eng. Mater. Technol., **125**, pp. 116–124.
[5] Puschmann, E., and Specht, E., 2004, "Atomized Spray Quenching as an Alternative Quenching Method for Defined Adjustment of Heat Transfer," Steel Res., **75**, pp. 283–288.
[6] Puschmann, E., and Specht, E., 2004, "Transient Measurement of Heat Transfer for Metal Quenching with Atomized Sprays," Exp. Therm. Fluid Sci., **28**, pp. 607–615.
[7] Fernandes, F. M. B., Denis, S., and Simon, A., 1985, "Mathematical Model Coupling Phase Transformation and Temperature Evolution During Quenching of Steels," Mater. Sci. Technol., **1**(10), pp. 805–815.
[8] Song, R., Dhatt, G., and Cheikh, A. B., 1990, "Thermo-mechanical Finite Element Model of Casting Systems," Int. J. Numer. Methods Eng., **30**, pp. 579–599.
[9] Gür, C. H., and Tekkaya, A. E., 1996, "Finite Element Simulation of Quench Hardening," Steel Res., **67**, pp. 298–306.
[10] Inoue, T., and Wang, Z., 1985, "Coupling Between Stress, Temperature and Metallic Structures During Processes Involving Phase Transformations," Mater. Sci. Technol., **1**(10), pp. 845–850.
[11] Leblond, J. B., Mottet, G., Devaux, J., and Devaux, J. C., 1985, "Mathematical Model of Anisothermal Phase Transformations in Steels, and Predicted Plastic Behavior," Mater. Sci. Technol., **1**, pp. 815–822.
[12] Kolmogorov, A. E., 1937, Izv. Akad. Nauk SSSR, Ser. Mat., **1**, 355; Johnson, W. A., and Mehl, R. F., 1939, Trans. Am. Inst. Min., Metall. Pet. Eng., **135**, p. 416; Avrami, M., 1939, J. Chem. Phys., **7**, p. 103.
[13] Hougardy, H. P., 1990, *Phase Transition and Microstructure of Unalloyed Steel*, (German: Umwandlung und Gefüge unlegierter Stähle), Steel Research Press, Dusseldorf, Germany.
[14] Scheil, E., 1935, Steel Res., **12**, p. 565.
[15] Koistinen, D. P., and Marburger, R. E., 1959, "A General Equation Prescribing the Extent of the Austenite-martensite Transformation in Pure Iron-Carbon Alloys and Plain Carbon Steels," Acta Metall., **7**, pp. 59–60.
[16] Denis, S., Gautier, E., Sjöström, S., and Simon, A., 1987, "Influence of Stress on the Kinetics of Pearlitic Transformation during Continuous Cooling," Acta Metall., **35**, pp. 1621–1632.
[17] Betten, J., 1993, *Continuum Mechanics: Elasto- Plasto- and Creep Mechanics*, (German: Kontinuumsmechanik: Elasto- Plasto- und Kriechmechanik), Springer Press, New York.
[18] Pietzsch, R., 2000, *Simulation and Minimization of the Distortion of Steel Profiles with the Cooling*, (German: Simulation und Minimierung des Verzuges von Stahlprofilen bei der Abkühlung), Ph.D. thesis, Shaker Press, Aachen.
[19] Greenwood, G. W., and Johnson, R. H., 1965, "The Deformation of Metals Under Small Stress during Phase Transformation," Proc. R. Soc. London, Ser. A, **283**, pp. 403–422.
[20] Brzoza, M., Specht, E., Ohland, J., Luebben, T., Belkessam, O., Fritsching, U., and Mayr, P., 2005, "Adjustment of Nozzle Field During Gas Quenching," (German: Dusenfeldanpassung bei der flexiblen Gasabschreckung), Harterei-Technische Mitteilungen, München, **60**, no. 3, pp. 166–172.
[21] Exner, H. J., and Hougardy, H. P., 1986, "Introduction to the Quantitative Microstructure Analysis," (German: Einführung In die Quantitative Gefügeanalyse), DGM Informationsgesellschaft Oberursel, pp. 8–18.
[22] Radaj, D., 1992, *Heat Effects of Welding*, Springer Press, New York.

Elastic Analysis for Defects in an Orthotropic Kirchhoff Plate

Kyeong-Jin Yang
e-mail: kjyang@kaeri.re.kr

Ki-Young Lee

Jong-Hwa Chang

Nuclear Hydrogen Project,
Korea Atomic Energy Research Institute,
Yuseong, Daejeon 305-353,
Republic of Korea

Defects such as inhomogeneities, inclusions with eigenstrains, and dislocations in an infinite orthotropic Kirchhoff plate are analyzed. These results could be applied to thin plate problems regardless of whether the plate is homogeneous or inhomogeneous in the direction of a thickness. An orthotropic laminated plate with a symmetric plane normal to the direction of the thickness is included as a special case. The eigenstrain is assumed to vary throughout the direction of the thickness. Thus, a bending of the plate due to the eigenstrain is considered. Employing Green's functions, which are expressed in explicit compact forms in a Cartesian coordinates system and were recently obtained by using a Stroh-type formalism, the elastic fields for defects are obtained by way of Eshelby's inclusion method. The general solutions for the extension and bending deformations due to the mid-plane eigenstrain and eigencurvature are expressed in quasi-Newtonian potentials and their derivatives, which appear in a closed form for the elliptic inclusion. For the bending problem of an inclusion with uniform eigencurvature, the curvature inside the inclusion becomes uniform, corresponding to that from Eshelby's analysis of an isotropic solid. Edge dislocation and elliptic inclusions with polynomial eigenstrains are also discussed in this work. [DOI: 10.1115/1.2338051]

1 Introduction

Since the pioneering works of Eshelby [1–3] on the elastic analysis of inclusions and inhomogeneities, extensive research on this subject has been reported, including the study of inclusions and inhomogeneities with nonuniform eigenstrains, cracks, dislocations, and effective material properties. Following Mura [4], a homogeneous inclusion is defined as a bounded sub-region contained in a material with identical material properties but with a finite stress-free transformation strain (or eigenstrain) within the sub-region. An inhomogeneous inclusion is defined as a bounded sub-region with material properties different from those of the surrounding region or matrix and with an eigenstrain. Various examples of naturally occurring eigenstrains are those due to a thermal expansion, lattice parameter mismatch, phase transformation, inelastic deformation, electromechanical strains, compositional differences, etc.

A number of two-dimensional inclusion problems in an infinite solid have been solved, mostly by the Green's function method and a complex-variables approach. For example, using Eshelby's inclusion method, Jaswon and Bhargava [5] obtained explicit solutions for elliptic inclusions under conditions of plane strain and generalized plane stress. Their methods were extended, independently, by Willis [6] and Bhargava and Radhakrishna [7], to cubic and orthotropic media, respectively. Hwu and Ting [8] solved a two-dimensional elliptic inhomogeneity problem in an infinite anisotropic solid by using the Stroh formalism. Recently, Beom [9] and Beom et al. [10,11] successfully expanded Eshelby's methodology to the problem of a plate containing an inclusion or inhomogeneity with a uniform eigenstrain. The results have been applied by Duong and Yu to the analysis of thermal stresses due to a low uniform operating temperature in a one-sided composite bonded repair [12,13]. The works done by Beom [9] and Beom et al. [10,11] extended to an inclusion with polynomial eigenstrains by Yang et al. [14]. However, these works are concerned with isotropic plates. No explicit solutions for the anisotropic plates have yet been reported in the literature.

Considered in the present study is an infinite orthotropic Kirchhoff plate containing an elliptic cylindrical inclusion with a uniform eigenstrain. The eigenstrain is assumed to vary throughout the direction of a thickness. Thus, the bending of the plate due to an eigenstrain has to be considered. According to the Kirchhoff plate theory, thickness-direction-varying eigenstrain generates a mid-plane eigenstrain and an eigencurvature, which are, respectively, related to the extension and bending problems of an orthotropic Kirchhoff plate. By way of the Green's function method, integral-type general solutions of the in-plane and out-of-plane displacements are expressed in terms of the mid-plane eigenstrain and eigencurvature, respectively. Real-form Green's functions explicitly expressed in a Cartesian coordinates system, recently obtained by using a Stroh-like formalism, are employed here [15]. The elastic fields are obtained by evaluating the integrals. An infinite orthotropic plate containing an elliptic cylindrical inhomogeneity with a prescribed eigenstrain is also considered. The elastic fields caused by the inhomogeneity are determined by the equivalent eigenstrain method, which is an extension of the equivalent inclusion method proposed by Eshelby [1] for an ellipsoidal inhomogeneity problem. An edge dislocation and an inclusion with polynomial eigenstrains are also discussed. The inclusion problem with polynomial eigenstrains proposed by Sendeckyj [16] is extended to the problem of a bending deformation of a plate due to an inclusion with polynomial eigencurvatures.

2 Formulation

Among the numerous two-dimensional models for studying deformations of a thin plate, the most classic and celebrated model is the Kirchhoff plate model, in which a transverse normal plane remains normal to the mid-plane of the plate during a deformation. Consider a deformation of an infinite orthotropic Kirchhoff plate containing an elliptic subregion Ω in which a uniform eigenstrain ε_{ij}^{*} is prescribed, as shown in Fig. 1. The traction vanishes at infinity. The major and minor semi-axes of the ellipse are a_1 and a_2, respectively. The eigenstrain ε_{ij}^{*} is assumed to vary throughout the direction of the thickness. Thus, the bending deformation of the plate due to the eigenstrain must be considered. The plate is

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received June 27, 2005; final manuscript received April 17, 2006. Review conducted by G. C. Buscaglia. 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.

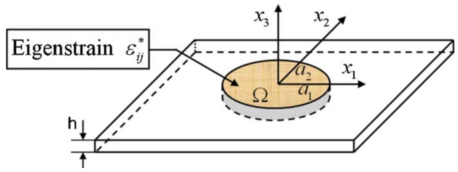


Fig. 1 Infinite plate containing an elliptic cylindrical inclusion with principal half axes a_1 and a_2, in which uniform eigenstrain is prescribed

composed of an orthotropic, linearly elastic material that can be inhomogeneous in the thickness direction. Accordingly, the plate includes a special case of an orthotropic laminated plate with a symmetric plane normal to the direction of the thickness.

According to the Kirchhoff plate theory, the displacements at any point of a plate are written as [17]

$$u_i = u_i^0 - x_3 w_{,i}(x_1\ x_2), \quad (i=1,2) \tag{1a}$$

$$u_3 = w(x_1,x_2) \tag{1b}$$

where u_i $(i=1,2)$ and u_3 are the in-plane and out-of-plane displacements, respectively, u_i^0 is the in-plane displacement on the mid-plane, and the subscript comma denotes a partial derivative with respect to the in-plane Cartesian coordinates, x_1 and x_2. Total strain ε_{ij} is regarded as the sum of an elastic strain e_{ij} and an eigenstrain ε^*_{ij}

$$\varepsilon_{ij} = e_{ij} + \varepsilon^*_{ij} \quad (i,j=1,2) \tag{2}$$

and the total strain must be compatible

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

The elastic strain e_{ij} is related to stress σ_{ij} by Hooke's law

$$\sigma_{ij} = \tilde{C}_{ijkl}e_{kl} = \tilde{C}_{ijkl}(\varepsilon^0_{kl} + x_3\kappa_{kl} - \varepsilon^*_{kl}) \quad (i,j,k,l=1,2) \tag{4}$$

where $\tilde{C}_{ijkl}$ is the reduced elasticity tensor related to the elasticity tensor C_{ijkl} as

$$\tilde{C}_{ijkl} = C_{ijkl} - \frac{C_{ij33}C_{33kl}}{C_{3333}} \tag{5}$$

and ε^0_{kl} and κ_{kl} are the mid-plane strain and curvature defined as $\varepsilon^0_{kl}=(u^0_{k,l}+u^0_{l,k})/2$ and $\kappa_{kl}=-w_{,kl}$, respectively. In this paper, the repetition of an index in a term denotes a summation with respect to that index over its range of 1 to 2 for a lowercase Roman letter unless indicated otherwise, and boldfaced symbols represent vectors or matrices. By integrating Eq. (4) through the plate thickness and making use of the attribute that the resultant stress N_{ij} and bending moment M_{ij} are, respectively, defined as

$$N_{ij} = \int_{x_3} \sigma_{ij}\, dx_3 \tag{6a}$$

$$M_{ij} = \int_{x_3} \sigma_{ij}x_3\, dx_3 \tag{6b}$$

the constitutive relations for the orthotropic plate containing an inclusion with the mid-plane eigenstrain ε^{0*}_{ij} and the eigencurvature κ^*_{ij} can be written as

$$N_{ij} = A_{ijkl}(\varepsilon^0_{kl} - \varepsilon^{0*}_{kl}) \quad (i,j,k,l=1,2) \tag{7a}$$

$$M_{ij} = D_{ijkl}(\kappa_{kl} - \kappa^*_{kl}) \quad (i,j,k,l=1,2) \tag{7b}$$

where A_{ijkl} and D_{ijkl} are the extensional and bending stiffness tensors, respectively, given by

$$A_{ijkl} = \int_{x_3} \tilde{C}_{ijkl}\, dx_3 \tag{8a}$$

$$D_{ijkl} = \int_{x_3} \tilde{C}_{ijkl}x_3^2\, dx_3 \tag{8b}$$

and the mid-plane eigenstrain and the eigencurvature are, respectively, related to the prescribed eigenstrain ε^*_{ij} as

$$A_{ijkl}\varepsilon^{0*}_{kl} = \int_{x_3} \tilde{C}_{ijkl}\varepsilon^*_{kl}\, dx_3 \tag{9a}$$

$$D_{ijkl}\kappa^*_{kl} = \int_{x_3} \tilde{C}_{ijkl}\varepsilon^*_{kl}x_3\, dx_3 \tag{9b}$$

It is briefly noted here how to determine the mid-plane eigenstrain ε^{0*}_{ij} and eigencurvature κ^*_{ij} from the prescribed eigenstrain ε^*_{ij} for an orthotropic laminated plate with a symmetric plane normal to the x_3-axis, a laminated plate layered by isotropic and transversely isotropic materials with a symmetric plane along the x_3-axis, and for an orthotropic plate. Equations (9a) and (9b) provide the expressions for a system of simultaneous equations with unknown mid-plane eigenstrains and eigencurvatures for the extension and bending problems, respectively. However, if the inclusion is also an orthotropic solid with the same crystalline direction to the matrix, then the tensile and shear components will be decoupled. Hence, the tensile components of the eigenstrain only generate the tensile ones of the mid-plane eigenstrain and eigencurvature, and the shear component of the eigenstrain also generates its counterpart of the mid-plane eigenstrain and eigencurvature. The tensile components of the mid-plane eigenstrain $\bar{\varepsilon}^{0*}=\{\varepsilon^{0*}_{11},\varepsilon^{0*}_{22}\}$ and eigencurvature $\bar{\boldsymbol{\kappa}}^*=\{\kappa^*_{11},\kappa^*_{22}\}$ can be, respectively, expressed in a compact matrix form from Eqs. (9a) and (9b) as

$$\bar{\boldsymbol{\varepsilon}}^{0*} = \int_{x_3} \bar{\mathbf{A}}^{-1}\bar{\mathbf{C}}\bar{\boldsymbol{\varepsilon}}^* dx_3 \tag{10a}$$

$$\bar{\boldsymbol{\kappa}}^* = \int_{x_3} \bar{\mathbf{D}}^{-1}\bar{\mathbf{C}}\bar{\boldsymbol{\varepsilon}}^* x_3 dx_3 \tag{10b}$$

where the components of the 2×2 stiffness matrices, $\bar{\mathbf{A}}$, $\bar{\mathbf{D}}$, and $\bar{\mathbf{C}}$ are, respectively,

$$\bar{\mathbf{A}} = \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix} \quad \bar{\mathbf{D}} = \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & C_{22} \end{bmatrix} \quad \bar{\mathbf{C}} = \begin{bmatrix} \tilde{C}_{11} & \tilde{C}_{12} \\ \tilde{C}_{12} & \tilde{C}_{22} \end{bmatrix} \tag{11}$$

In this paper, the transformations between the extension and bending stiffness, A_{ijkl} and D_{ijkl}, and A_{mn} and D_{mn}, respectively, are accomplished by replacing the subscripts ij (or kl) by m (or n) using the following role:

ij (or kl)$=11,22,12$ (or 21)$\leftrightarrow m$(or n)$=1,2,6$.

The shear components of the mid-plane eigenstrain and the eigencurvature can be, respectively, expressed as

$$\varepsilon^{0*}_{12} = \frac{1}{A_{66}}\int_{x_3} \tilde{C}_{66}\varepsilon^*_{12}\, dx_3 \tag{12a}$$

$$\kappa^*_{12} = \frac{1}{D_{66}}\int_{x_3} \tilde{C}_{66}\varepsilon^*_{12}x_3\, dx_3 \tag{12b}$$

For a laminated plate layered by isotropic or transversely isotropic materials with a symmetric plane normal to the x_3-axis, the mid-plane eigenstrain and the eigencurvature are, respectively, reduced to

$$\varepsilon_{ij}^{0*}=\frac{1}{A_{11}-A_{12}}\int_{x_3}\left[\frac{A_{11}C_{12}-A_{12}C_{11}}{A_{11}+A_{12}}\varepsilon_{kk}^{*}\delta_{ij}+(C_{11}-C_{12})\varepsilon_{ij}^{*}\right]dx_3 \tag{13a}$$

$$\kappa_{ij}^{*}=\frac{1}{D_{11}-D_{12}}\int_{x_3}\left[\frac{D_{11}C_{12}-D_{12}C_{11}}{D_{11}+D_{12}}\varepsilon_{kk}^{*}\delta_{ij}+(C_{11}-C_{12})\varepsilon_{ij}^{*}\right]x_3\,dx_3 \tag{13b}$$

For an orthotropic plate, the mid-plane eigenstrain ε_{ij}^{0*} and eigencurvature κ_{ij}^{*} in Eqs. (9*a*) and (9*b*), respectively, reduce

$$\varepsilon_{ij}^{0*}=\frac{1}{h}\int_{x_3}\varepsilon_{ij}^{*}\,dx_3 \tag{14a}$$

$$\kappa_{ij}^{*}=\frac{12}{h^3}\int_{x_3}\varepsilon_{ij}^{*}x_3\,dx_3 \tag{14b}$$

Employing Eshelby's inclusion method to the plate problem, integral type solutions for in-plane and out-of-plane displacements can be, respectively, expressed as [14]

$$u_i^0(x_1,x_2)=-\int_{\Omega}A_{jlmn}G_{ij,l}\varepsilon_{mn}^{0*}(x_1',x_2')dA \tag{15a}$$

$$w(x_1,x_2)=-\int_{\Omega}D_{jlmn}G_{33,jl}\kappa_{mn}^{*}(x_1',x_2')dA \tag{15b}$$

where the Green's functions $G_{ij}(\mathbf{x}-\mathbf{x}')$ and $G_{33}(\mathbf{x}-\mathbf{x}')$ are the displacement components u_i^0 and u_3 at point $\mathbf{x}$ when unit concentrated forces in the x_i-direction and x_3-direction, respectively, are applied at point $\mathbf{x}'$ in the infinite orthotropic Kirchhoff plate, of which the components are given in Appendix A.

3 Homogeneous Inclusion

3.1 Deformation of the Plate. For an orthotropic Kirchhoff plate containing an inclusion with mid-plane eigenstrain and eigencurvature, the in-plane and out-of-plane displacements due to a homogeneous inclusion in Eqs. (15*a*) and (15*b*) are, respectively, reduced to

$$u_i^0=\int_{\Omega}f_{ijk}\varepsilon_{jk}^{0*}\,d\Omega(x_1'\ x_2') \tag{16a}$$

$$w=\int_{\Omega}h_{mn}\kappa_{mn}^{*}\,d\Omega(x_1'\ x_2') \tag{16b}$$

where the functions f_{ijk} and h_{mn} are linearly proportional to the first and the second derivatives of the Green's functions for the extension and bending problems, respectively, as

$$f_{ijk}=-A_{mnjk}G_{im,n} \tag{17a}$$

$$h_{mn}=-D_{jlmn}G_{33,jl} \tag{17b}$$

Substituting Eqs. (A1*a*)–(A1*c*) into Eq. (17*a*), the components of the tensor f_{ijk} for the extension problem can be expressed as

$$f_{111}=-\sum_{n=1}^{2}(A_{12}+A_{66})(p_n^A)^2\eta_n\alpha_n\phi_{n,1}^A \tag{18a}$$

$$f_{112}=-\sum_{n=1}^{2}(A_{12}+A_{66})\eta_n\alpha_n\phi_{n,2}^A \tag{18b}$$

$$f_{122}=\sum_{n=1}^{2}\eta_n\alpha_n\phi_{n,1}^A \tag{18c}$$

$$f_{211}=-\sum_{n=1}^{2}\eta_n\alpha_n\beta_n\phi_{n,2}^A \tag{18d}$$

$$f_{222}=\sum_{n=1}^{2}\left(\frac{1}{p_n^A}\right)^2\eta_n\alpha_n\beta_n\phi_{n,2}^A \tag{18e}$$

$$f_{212}=\sum_{n=1}^{2}\eta_n\alpha_n\beta_n\phi_{n,1}^A \tag{18f}$$

where the constants p_n^A, α_n, β_n and the function ϕ_n^A are given in Eqs. (A2), (A3*a*), (A3*b*), and (A4*a*) in Appendix A, respectively, and the constant η_n is

$$\eta_n=A_{66}(A_{11}+(p_n^A)^2A_{12}) \tag{19}$$

The tensor f_{ijk} is symmetric in jk, that is $f_{ijk}=f_{ikj}$. Here the relations $\psi_{n,1}^A=-(1/p_n^A)\phi_{n,2}^A$ and $\psi_{n,2}^A=p_n^A\phi_{n,1}^A$, where the function ψ_n^A is defined in (A4*b*), are used. Substituting Eqs. (A7*a*) and (A7*b*) into Eq. (17*b*) for a bending problem, the following is derived

$$h_{11}=-\sum_{n=1}^{2}\nu_n\gamma_n\phi_n^D \tag{20a}$$

$$h_{22}=\sum_{n=1}^{2}\left(4D_{66}-\frac{\nu_n}{(p_n^D)^2}\right)\gamma_n\phi_n^D \tag{20b}$$

$$h_{12}=\sum_{n=1}^{2}2D_{66}\gamma_np_n^D\psi_n^D \tag{20c}$$

where the constants ϕ_n^D, γ_n are given in Eqs. (A8) and (A9), respectively, and the constant ν_n is

$$\nu_n=(D_{11}-D_{12}(p_n^D)^2) \tag{21}$$

The tensor h_{ij} is symmetric in ij. By introducing integrals Φ_n^* and Ψ_n^* defined by

$$\Phi_n^*=\int_{\Omega}\phi_n^*\,d\Omega \tag{22a}$$

$$\Psi_n^*=\int_{\Omega}\psi_n^*\,d\Omega \tag{22b}$$

the in-plane and out-of-plane displacements of Eqs. (16*a*) and (16*b*) could be explicitly expressed by these integrals and their derivatives. In this paper, the superscript * stands for the superscript A or D, except for an eigenstrain and an inhomogeneity, designating the constants or functions of the extension and bending problems, respectively. Here, it should be noted that the partial derivatives $\Psi_{n,1}^*$ and $\Psi_{n,2}^*$ are equal to the derivatives $-(1/p_n^*)\Phi_{n,2}^*$ and $p_n^*\Phi_{n,1}^*$, respectively.

3.2 Mid-Plane Strain, Curvature, Resultant Stress and Bending Moment. Substituting Eqs. (16*a*) and (16*b*) into the relations, $\varepsilon_{kl}^0=(u_{k,l}^0+u_{l,k}^0)/2$ and $\kappa_{kl}=-w_{,kl}$, respectively, the mid-plane strain and the curvature outside an inclusion with a uniform mid-plane eigenstrain and eigencurvature can be, respectively, expressed as

$$\varepsilon_{ij}^0=X_{ijkl}\varepsilon_{kl}^{0*}\quad(x_1\ x_2)\notin\Omega \tag{23a}$$

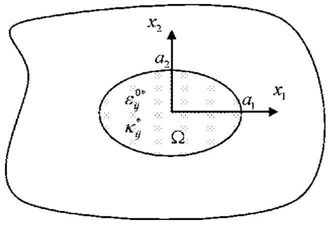


Fig. 2 Infinite plate containing an elliptic cylindrical inclusion with principal half axes a_1 and a_2, in which uniform mid-plane eigenstrain and eigencurvature are prescribed

$$\kappa_{ij} = Y_{ijkl}\kappa_{kl}^{*} \quad (x_1\ x_2) \notin \Omega \tag{23b}$$

where the tensors X_{ijkl} and Y_{ijkl} are

$$X_{ijkl} = \frac{1}{2}(F_{ikl,j} + F_{jkl,i}) \tag{24a}$$

$$Y_{ijkl} = -H_{kl,ij} \tag{24b}$$

where the functions F_{ijk} and H_{ij} are

$$F_{ijk} = \int_{\Omega} f_{ijk}\, d\Omega \quad (x_1\ x_2) \notin \Omega \tag{25a}$$

$$H_{ij} = \int_{\Omega} h_{ij}\, d\Omega \quad (x_1\ x_2) \notin \Omega \tag{25b}$$

The mid-plane strain and the curvature inside the inclusion are

$$\varepsilon_{ij}^{0} = S_{ijkl}\varepsilon_{kl}^{0*} \tag{26a}$$

$$\kappa_{ij} = T_{ijkl}\kappa_{kl}^{*} \tag{26b}$$

where the tensors S_{ijkl} and T_{ijkl} have the same expression as X_{ijkl} and Y_{ijkl}, respectively, except that the integrals in Eqs. (25a) and (25b) are evaluated for points inside the inclusion Ω. The tensors X_{ijkl}, Y_{ijkl}, S_{ijkl} and T_{ijkl} are symmetric with respect to an interchange of i and j or k and l, i.e., $X_{ijkl}=X_{jikl}=X_{ijlk}$, $Y_{ijkl}=Y_{jikl}=Y_{ijlk}$, etc. The solutions of Eqs. (23a), (23b), (26a), and (26b) are functions of the quasi-Newtonian potentials Φ_n^{*} for the inclusion Ω

$$\Phi_n^{*} = \int_{\Omega} \phi_n^{*}\, d\Omega \tag{27}$$

For an elliptic inclusion, the quasi-Newtonian potentials Φ_n^{*} can be obtained in a closed form as will be shown in the following section. Substituting the above mid-plane eigenstrain ε_{ij}^{0} and eigencurvature κ_{ij} into Eqs. (7a) and (7b) with a prescribed mid-plane eigenstrain and eigencurvature, respectively, the resultant stress and moment could be obtained. The expressions for the elastic energy in the matrix and in the inclusion and the interaction energy of the elastic field with another field are the same as those given by Eshelby [1] for an isotropic inclusion.

3.3 Elliptic Inclusion With Uniform Mid-Plane Eigenstrain and Eigencurvature. Considered here is an elliptic cylindrical inclusion Ω described by

$$\frac{x_1'^2}{a_1^2} + \frac{x_2'^2}{a_2^2} = 1 \tag{28}$$

where the point (0, 0) is its center. The mid-plane eigenstrain and eigencurvature are prescribed inside the inclusion, as shown in Fig. 2. The following relationship:

$$\Phi_n = \int_{\Omega} \phi_n\, d\Omega(x_1', x_2') = \int_{\Omega^*} \frac{1}{p_n^*}\phi_n\, d\Omega^*(x_1', p_n^* x_2') \tag{29}$$

where the constants, p_n^* are p_n^A or p_n^D, and Φ_n indicates the Newtonian potential at the point $(x_1, p_n^* x_2)$ due to an elliptic cylinder with a mass density $1/p_n^*$, semi-axes a_1 and $p_n^* a_2$, of which the explicit expressions are shown in Appendix B. The logarithmic Newtonian potential Φ_n and those derivatives are used for evaluating the mid-plane strain, curvature, corresponding resultant stress, and the bending moment. For a mathematical convenience, mechanical fields inside the elliptic inclusion with a uniform mid-plane eigenstrain and eigencurvature are considered here. Substituting Eqs. (18a)–(18f) into Eq. (25a), and then substituting these results into Eq. (24a), the nonzero components of the tensor S_{ijkl} can be given as

$$S_{1111} = -\sum_{n=1}^{2} (A_{12} + A_{66})(p_n^A)^2 \eta_n \alpha_n \Phi_{n,11}^A \tag{30a}$$

$$S_{1122} = \sum_{n=1}^{2} \eta_n \alpha_n \Phi_{n,11}^A \tag{30b}$$

$$S_{2211} = -\sum_{n=1}^{2} \eta_n \alpha_n \beta_n \Phi_{n,22}^A \tag{30c}$$

$$S_{2222} = \sum_{n=1}^{2} \left(\frac{1}{p_n^A}\right)^2 \eta_n \alpha_n \beta_n \Phi_{n,22}^A \tag{30d}$$

$$S_{1212} = -\sum_{n=1}^{2} \frac{(A_{12} + A_{66})}{2} \eta_n \alpha_n \Phi_{n,22}^A + \sum_{n=1}^{2} \frac{1}{2} \eta_n \alpha_n \beta_n \Phi_{n,11}^A \tag{30e}$$

Here it should be noted that the integrals in Eqs. (25a) and (25b) are evaluated for points inside the inclusion Ω, as mentioned above, and $\Phi_{n,12}^{*}=0$ for points inside the elliptic inclusion. Similarly, substituting Eqs. (20a)–(20c) into (25b), and then substituting these results into Eq. (24b), the nonzero components of the tensor T_{ijkl} can be given as

$$T_{1111} = \sum_{n=1}^{2} \nu_n \gamma_n \Phi_{n,11}^D \tag{31a}$$

$$T_{1122} = -\sum_{n=1}^{2} \left(4D_{66} - \frac{\nu_n}{(p_n^D)^2}\right) \gamma_n \Phi_{n,11}^D \tag{31b}$$

$$T_{2211} = \sum_{n=1}^{2} \nu_n \gamma_n \Phi_{n,22}^D \tag{31c}$$

$$T_{2222} = -\sum_{n=1}^{2} \left(4D_{66} - \frac{\nu_n}{(p_n^D)^2}\right) \gamma_n \Phi_{n,22}^D \tag{31d}$$

$$T_{1212} = \sum_{n=1}^{2} 2D_{66} \gamma_n \Phi_{n,22}^D \tag{31e}$$

Since the derivatives of the potential, $\Phi_{n,11}^{*}$ and $\Phi_{n,22}^{*}$, are constants inside the inclusion and, therefore, the Eshelby tensors S_{ijkl} and T_{ijkl} are also constants inside that region, the strain and curvature inside the elliptic inclusion are uniform, corresponding to those for an isotropic solid, first investigated by Eshelby [1].

Substituting Eq. (26a) into Eq. (7a), the components of the resultant stress inside the inclusion can be expressed as

$$N_{11} = [A_{11}(S_{1111} - 1) + A_{12}S_{2211}]\varepsilon_{11}^{0*} + [A_{11}S_{1122} + A_{12}(S_{2222} - 1)]\varepsilon_{22}^{0*} \quad (32a)$$

$$N_{22} = [A_{12}(S_{1111} - 1) + A_{22}S_{2211}]\varepsilon_{11}^{0*} + [A_{12}S_{1122} + A_{22}(S_{2222} - 1)]\varepsilon_{22}^{0*} \quad (32b)$$

$$N_{12} = A_{66}(S_{1212} - 1)\varepsilon_{12}^{0*} \quad (32c)$$

Similarly, substituting Eq. (26b) into Eq. (7b), the components of the bending moment inside the inclusion can be expressed as

$$M_{11} = [D_{11}(T_{1111} - 1) + D_{12}T_{2211}]\kappa_{11}^{*} + [D_{11}T_{1122} + D_{12}(T_{2222} - 1)]\kappa_{22}^{*} \quad (33a)$$

$$M_{22} = [D_{12}(T_{1111} - 1) + D_{22}T_{2211}]\kappa_{11}^{*} + [D_{12}T_{1122} + D_{22}(T_{2222} - 1)]\kappa_{22}^{*} \quad (33b)$$

$$M_{12} = D_{66}(T_{1212} - 1)\kappa_{12}^{*} \quad (33c)$$

For the region outside the inclusion, those fields might be evaluated without any difficulty via the quasi-Newtonian potentials and their derivatives for that region, as given in Appendix B.

4 Inhomogeneous Inclusion

Consider an infinite orthotropic Kirchhoff plate with stiffness A_{ijkl} and D_{ijkl} containing an elliptic inhomogeneity with stiffness A_{ijkl}^{*} and D_{ijkl}^{*}. A uniform mid-plane eigenstrain and eigencurvature are prescribed in the elliptic inhomogeneity. The major and minor semi-axes of the ellipse are a_1 and a_2, respectively. The elastic fields due to the inhomogeneity can be determined by the equivalent inclusion method proposed for an ellipsoidal inhomogeneity problem by Eshelby [1]. The resultant stress and bending moment due to the presence of an inhomogeneity with a uniform mid-plane eigenstrain ε_{kl}^{0*} and eigencurvature κ_{kl}^{*} can be simulated by the resultant stress and bending moment caused by a homogeneous inclusion when the eigenstrain and the eigencurvature are chosen properly, as illustrated in Fig. 3. The constitutive relation for the inhomogeneity problem is written as

$$N_{ij} = A_{ijkl}^{*}(\varepsilon_{kl}^{0} - \varepsilon_{kl}^{0*}) \quad (34a)$$

$$M_{ij} = D_{ijkl}^{*}(\kappa_{kl} - \kappa_{kl}^{*}) \quad \text{for } \mathbf{x} \text{ in } \Omega \quad (34b)$$

$$N_{ij} = A_{ijkl}\varepsilon_{kl}^{0} \quad (34c)$$

$$M_{ij} = D_{ijkl}\kappa_{kl} \quad \text{for } \mathbf{x} \text{ out of } \Omega \quad (34d)$$

By introducing the eigenstrain ε_{kl}^{0**} and the eigencurvature κ_{kl}^{**} in order to simulate the inhomogeneity problem in a manner analogous to the equivalent inclusion method in a three-dimensional elasticity, the constitutive relation for the homogeneous problem is written as

$$N_{ij} = A_{ijkl}(\varepsilon_{kl}^{0} - \varepsilon_{kl}^{0**}) \quad (35a)$$

$$M_{ij} = D_{ijkl}(\kappa_{kl} - \kappa_{kl}^{**}) \quad \text{for } \mathbf{x} \text{ in } \Omega \quad (35b)$$

$$N_{ij} = A_{ijkl}\varepsilon_{kl}^{0} \quad (35c)$$

$$M_{ij} = D_{ijkl}\kappa_{kl} \quad \text{for } \mathbf{x} \text{ out of } \Omega \quad (35d)$$

The conditions for the equivalency of the resultant stress and the bending moment in the two problems of an inhomogeneous and a homogeneous inclusion above are

$$A_{ijkl}^{*}(\varepsilon_{kl}^{0} - \varepsilon_{kl}^{0*}) = A_{ijkl}(\varepsilon_{kl}^{0} - \varepsilon_{kl}^{0**}) \quad (36a)$$

$$D_{ijkl}^{*}(\kappa_{kl} - \kappa_{kl}^{*}) = D_{ijkl}(\kappa_{kl} - \kappa_{kl}^{**}) \quad \text{for } \mathbf{x} \text{ in } \Omega \quad (36b)$$

Equations (36a) and (36b) determine ε_{kl}^{0**} and κ_{kl}^{**} for a given ε_{kl}^{0*} and κ_{kl}^{*}, respectively, in such a manner that the equivalency holds.

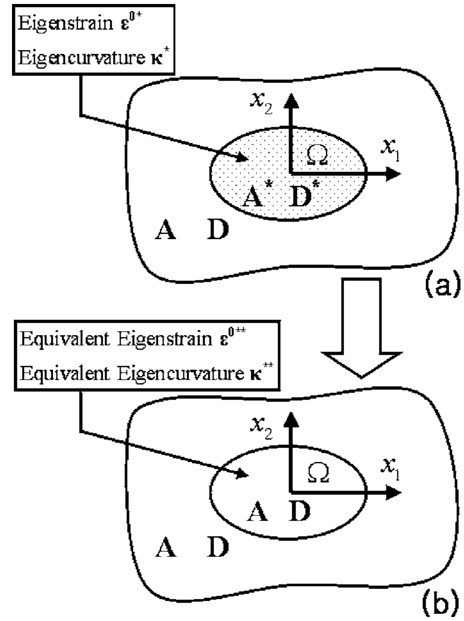


Fig. 3 Schematic illustration for equivalent inclusion method; (*a*) infinite plate containing an inhomogeneous inclusion with mid-plane eigenstrain and eigencurvature, (*b*) infinite plate containing an homogeneous inclusion with equivalent eigenstrain and eigencurvature, which are properly chosen to give the same resultant stress and bending moment inside the inclusion.

As mentioned earlier, ε_{kl}^{0} and κ_{kl} in the equations above can be obtained with known functions of ε_{kl}^{0**} and κ_{kl}^{**} when the eigenstrain problem in the homogeneous material is solved. For a uniform mid-plane eigenstrain and eigencurvature, it is convenient to define $\boldsymbol{\varepsilon}^{0**}=\{\varepsilon_{11}^{0**},\varepsilon_{22}^{0**},2\varepsilon_{12}^{0**}\}$ and $\boldsymbol{\kappa}^{**}=\{\kappa_{11}^{**},\kappa_{22}^{**},2\kappa_{12}^{**}\}$, while $\mathbf{S}$ and $\mathbf{T}$ are simplified by 3×3 matrices, of which the components are given as

$$\mathbf{S} = \begin{bmatrix} S_{1111} & S_{1122} & 0 \\ S_{2211} & S_{2222} & 0 \\ 0 & 0 & 2S_{1212} \end{bmatrix} \quad (37a)$$

$$\mathbf{T} = \begin{bmatrix} T_{1111} & T_{1122} & 0 \\ T_{2211} & T_{2222} & 0 \\ 0 & 0 & 2T_{1212} \end{bmatrix} \quad (37b)$$

Then, the equivalent eigenstrain and the eigencurvature in Eqs. (36a) and (36b) could be written in a compact matrix form, respectively, as

$$\boldsymbol{\varepsilon}^{0**} = (\mathbf{A}^{*}\mathbf{S} - \mathbf{AS} + \mathbf{A})^{-1}\mathbf{A}^{*}\boldsymbol{\varepsilon}^{0*} \quad (38a)$$

$$\boldsymbol{\kappa}^{**} = (\mathbf{D}^{*}\mathbf{T} - \mathbf{DT} + \mathbf{D})^{-1}\mathbf{D}^{*}\boldsymbol{\kappa}^{*} \quad (38b)$$

where $\mathbf{A}$ and $\mathbf{D}$ are defined by 3×3 matrices, of which the components are given in reduced notation as

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & A_{66} \end{bmatrix} \quad (39a)$$

$$\mathbf{D} = \begin{bmatrix} D_{11} & D_{12} & 0 \\ D_{21} & D_{22} & 0 \\ 0 & 0 & D_{66} \end{bmatrix} \quad (39b)$$

Similarly, $\mathbf{A}^{*}$ and $\mathbf{D}^{*}$ are defined as $\mathbf{A}$ and $\mathbf{D}$ when replacing A_{ij} and D_{ij} by A_{ij}^{*} and D_{ij}^{*}, respectively. Notably, the equivalent eigenstrain and eigencurvature are related to the mid-plane eigenstrain and eigencurvature in a similar function form, and they are different in the stiffness matrix and Eshelby's tensors.

In general, Eqs. (36*a*) and (36*b*) are a system of simultaneous equations with unknown ε_{ij}^{0**} and κ_{ij}^{**}, where the tensile and shear components are coupled. However, if the inclusion is also an orthotropic solid with elastic constants A_{ijkl}^{*} and D_{ijkl}^{*}, then the extension and shear components will be decoupled. As shown in Eqs. (36*a*) and (36*b*) the algebraic equations for calculating the equivalent eigenstrain and eigencurvature for the extension and bending problems, respectively, have the same forms. Hence, it might be sufficient to simply consider one of two problems; selected here is the bending problem. In order to find the tensile components of the equivalent eigencurvature κ_{ij}^{**}, it is necessary to solve the following simultaneous equations:

$$[D_{11}^{*}T_{1111}+D_{12}^{*}T_{2211}-D_{11}(T_{1111}-1)-D_{12}T_{2211}]\kappa_{11}^{**}+[D_{11}^{*}T_{1122} + D_{12}^{*}T_{2222}-D_{11}T_{1122}-D_{12}(T_{2222}-1)]\kappa_{22}^{**} = D_{11}^{*}\kappa_{11}^{*}+D_{12}^{*}\kappa_{22}^{*} \tag{40a}$$

$$[D_{12}^{*}T_{1111}+D_{22}^{*}T_{2211}-D_{21}(T_{1111}-1)-D_{22}T_{2211}]\kappa_{11}^{**}+[D_{12}^{*}T_{1122} + D_{22}^{*}T_{2222}-D_{12}T_{1122}-D_{22}(T_{2222}-1)]\kappa_{22}^{**} = D_{12}^{*}\kappa_{11}^{*}+D_{22}^{*}\kappa_{22}^{*} \tag{40b}$$

And the shear component of the equivalent eigencurvature can be easily obtained from Eq. (36*b*) as

$$\kappa_{12}^{**}=\frac{D_{66}^{*}\kappa_{12}^{*}}{2(D_{66}^{*}-D_{66})T_{1212}+D_{66}} \tag{41}$$

Then, the problem of the inhomogeneous inclusion can be analyzed with an equivalently simulated problem of a homogeneous inclusion with the equivalent eigencurvature obtained by Eqs. (40*a*), (40*b*), and (41). That is, the curvature can be obtained by Eq. (23*b*) with κ_{ij}^{**} replacing κ_{ij}^{*}. That result is then substituted into Eq. (7*b*) to obtain the bending moment, in which the curvatures κ_{ij}^{*} are also replaced by the equivalent curvature κ_{ij}^{**}.

5 Some Examples and Discussion

The solutions of this work could be applied to a thin orthotropic laminated plate with a symmetric plane normal to the direction of the x_3-axis, which is, according to the Kirchhoff plate theory, equivalent to an orthotropic plate. As a result of disregarding the effect of a transverse shear deformation, the Kirchhoff plate theory is actually restrictive in terms of providing a precise analysis of thin laminated structures. However, it is still the most celebrated theory among the numerous two-dimensional models of laminated plates from the point of view of a usefulness and instructiveness for an application to structural mechanics. As an example of an application of these kinds of works, Duong and Yu [12,13] analyzed the thermal stresses due to a low uniform operating temperature in a one-sided composite bonded repair of an aircraft, which greatly improves the reliability of cracked aerospace metallic structures. They showed that the effect of a out-of-plane bending became significant, and a good agreement was found between analytical predictions and the finite element results in spite of disregarding the transverse shear deformation. The analysis could also be basically applied to an isotropic plate. Due to the anisotropic characteristics of the composite materials extensively used for the structural components of an aircraft, it is also expected that the work presented here will constitute an improved approach for analyzing these kinds of structures.

5.1 Inhomogeneous Inclusion With a Thermal Mismatch. As an example of an application of this work to analyzing the stress fields due to a thermal mismatch such as that in a composite bonded repair of an aircraft, an inhomogeneous elliptic cylindrical inclusion thermally expanding in an orthotropic plate is considered. For a mathematical convenience, the bending problem of the plate is considered (The extension problem would not entail more difficulty than the bending problem in terms of an analysis.). The nonzero components of the eigencurvature κ_{ij}^{*} can be obtained by using Eq. (10*b*) as

$$\kappa_{11}^{*}=\frac{1}{D_{11}^{*}D_{22}^{*}-D_{12}^{*2}}\int_{x_3}[(D_{22}^{*}\tilde{C}_{11}-D_{12}^{*}\tilde{C}_{12})\alpha_1\Delta T+(D_{22}^{*}\tilde{C}_{12} - D_{12}^{*}\tilde{C}_{22})\alpha_2\Delta T]x_3\,dx_3 \tag{42a}$$

$$\kappa_{22}^{*}=\frac{1}{D_{11}^{*}D_{22}^{*}-D_{12}^{*2}}\int_{x_3}[(D_{11}^{*}\tilde{C}_{12}-D_{12}^{*}\tilde{C}_{11})\alpha_1\Delta T+(D_{11}^{*}\tilde{C}_{22} - D_{12}^{*}\tilde{C}_{12})\alpha_2\Delta T]x_3\,dx_3 \tag{42b}$$

where α_i, $i=1,2$, and ΔT are the thermal expansion coefficients of the inclusion along the x_i—directions and the temperature excursion during a high-altitude cruising of the aircraft, respectively. With the equivalent eigencurvature κ_{ij}^{**} obtained from Eqs. (40*a*) and (40*b*) for κ_{ij}^{*} given in Eqs. (42*a*) and (42*b*), the curvature and the bending moment due to the localized thermal loading can be obtained from Eqs. (23*b*) and (7*b*), respectively, with κ_{ij}^{**} replacing κ_{ij}^{*}.

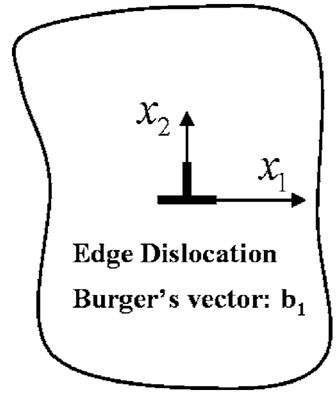


Fig. 4 Schematic diagram for edge dislocation on the $x_2=0$ plane ($x_1>0$) with Burger's vector b_1.

5.2 Edge Dislocation. The extension problem of the Kirchhoff plate is mathematically the same as the plane stress problem. Moreover, the results could be applied to the problem of a plane strain state by appropriately replacing the elastic constants. As an example of a plane strain problem, an edge dislocation at origin with the Burgers vector b_1 parallel to the x_1-direction, as shown in Fig. 4, is considered. Using the relationship between inclusions and dislocations proposed by Eshelby [3], a displacement due to an edge dislocation is given as

$$u_i^0=-b_1\int_0^{\infty}f_{i21}\,dx_1' \tag{43}$$

where the slip plane is the positive half plane x_1-x_2. Substituting Eqs. (18*b*) and (18*f*) into Eq. (43), the following is obtained:

$$u_1^0=b_1\sum_{n=1}^{2}(C_{12}+C_{66})\alpha_n\eta_nH_{n,2} \tag{44a}$$

$$u_2^0=-b_1\sum_{n=1}^{2}\alpha_n\beta_n\eta_nH_{n,1} \tag{44b}$$

where

$$H_n = x_1 \ln r_n^A - p_n^A x_2 \tan^{-1}\left(\frac{p_n^A x_2}{x_1}\right) \quad (45)$$

Here, the extensional elastic constants A_{ij} included in Eqs. (18*a*)–(18*f*), (44*a*), (44*b*), and (45) and the eigenvalue p_n^A should be replaced by the elastic constant C_{ij}. This is a new form of the solution for the problem of an edge dislocation.

5.3 Inclusion With Polynomial Eigencurvatures. Next, a bending problem of an infinite orthotropic plate containing an inclusion with a polynomial eigencurvature, prescribed as [14] is discussed;

$$\kappa_{ij}^* = \frac{12}{h^3}\int_{x_3} \varepsilon_{ij}^* x_3\, dx_3 = K_{ij} + K_{ijk}x_k + K_{ijkl}x_k x_l + \cdots, \quad (i,j,k,l,\ldots = 1,2) \quad (46)$$

where $K_{ij}, K_{ijk}, \ldots$, etc. are constants defined, respectively, by

$$K_{ij} = \frac{12}{h^3}\int_{-h/2}^{h/2} E_{ij}x_3\, dx_3, \quad K_{ijk} = \frac{12}{h^3}\int_{-h/2}^{h/2} E_{ijk}x_3\, dx_3, \quad \text{etc.} \quad (47)$$

where $E_{ij}, E_{ijk}, \ldots$ are constants of eigenstrains given in the form of polynomials of in-plane coordinates such as

$$\varepsilon_{ij}^* = E_{ij} + E_{ijk}x_k + E_{ijkl}x_k x_l + \cdots, \quad (i,j,k,l,\ldots = 1,2) \quad (48)$$

where $E_{ij}, E_{ijk}, \ldots$ are dependent only on x_3 and symmetric with respect to the free indices i, j and $E_{ijkl}=E_{ijlk}$, $E_{ijklm}=E_{ijkml}$, etc.

Substituting Eq. (46) into Eq. (16*b*), and then using the relation, $\kappa_{ij}=-w_{,ij}$, the curvature inside the inclusion is given by

$$\kappa_{ij} = T_{ijkl}(\mathbf{x})K_{kl} + T_{ijklm}(\mathbf{x})K_{klm} + T_{ijklmn}(\mathbf{x})K_{klmn} + \cdots \quad (49)$$

where, introducing an integral $H_{ijk\cdots l}$ defined by

$$H_{ijk\cdots l} = \int_{\Omega} x_k \cdots x_l h_{ij}\, d\Omega \quad (50)$$

the tensors $T_{ijk\cdots l}(\mathbf{x})$ are expressed in the integral $H_{ijk\cdots l}$ as

$$T_{ijk\cdots l}(\mathbf{x}) = -H_{k\cdots l,ij} \quad (51)$$

For a mathematical convenience, an inclusion with a first order eigencurvature, $\kappa_{ij}^*=K_{ij}+K_{ijk}x_k$ is considered. The explicit expressions for T_{ijklm} are given in Appendix C. When $\mathbf{x}$ is an interior point of the inclusion, then the *I*-integrals, I_{ij} in (C3*a*)–C3c are constants, and all their derivatives vanish. Thus, the tensor $T_{ijklm}(\mathbf{x})$ becomes a homogeneous linear function of $\mathbf{x}$. The mechanical fields outside the inclusion could also be obtained, since the *I*-integrals could be applied at any point of the matrix as well as inside the inclusion.

6 Summary

Analyzed in the present work is an infinite orthotropic Kirchhoff plate containing an elliptic homogeneous and inhomogeneous inclusion with an eigenstrain, where the displacement fields in the plate are expressed in terms of the in-plane and out-of-plane displacements on the mid-plane. By using Green's functions expressed in compact real form in a Cartesian coordinates system, which were obtained by using a Stroh-type formalism, integral type solutions of the displacements of the in-plane and out-of-plane direction were expressed in terms of a mid-plane eigenstrain and eigencurvature, respectively. For an elliptic cylindrical inclusion, the integrals could be explicitly expressed by calculating the quasi-Newtonian potentials. An edge dislocation and an inclusion with polynomial eigenstrains were also considered.

Acknowledgment

The author gratefully thanks Professor Earmme, Dept. of Mech. Eng., KAIST for his comments and valuable suggestion on the original manuscript. This work was supported by the Long-Term Nuclear R&D Program of the Korean Ministry of Science and Technology in 2006, and the Brain Korea 21 Project in 2005.

Appendix A: Green's Functions for Unit Concentrated Forces in the In-Plane and the Out-of-Plane Directions

Recently, a Stroh-like formalism for an anisotropic Kirchhoff plate has been reported [18–20]. From that formalism, solutions for various anisotropic plate problems, which had obtained in series form, are now available in explicit compact forms [21–24]. Green's functions derived from those works are relatively simple in a dual or cylindrical coordinates system; however, they are much more complicated in a Cartesian coordinates system. Moreover, the Green's functions are not favorable with respect to obtaining the Newtonian potentials associated with defect problems.

Using a slightly different method from the Stroh formalism for an anisotropic elasticity, where the linear composite variable of $z=x_1+ipx_2$ instead of $z=x_1+px_2$ is used, Green's functions for the unit concentrated forces in the in-plane and out-of-plane directions in an infinite orthotropic Kirchhoff plate were recently obtained in Cartesian coordinates [15]. For an extension problem, the explicit forms of the Green's functions in the Cartesian coordinates are

$$G_{11} = \sum_{n=1}^{2} (A_{12}+A_{66})^2 (p_n^A)^2 \alpha_n \phi_n^A \quad (A1a)$$

$$G_{12} = -\sum_{n=1}^{2} (A_{12}+A_{66}) p_n^A \alpha_n \beta_n \psi_n^A \quad (A1b)$$

$$G_{22} = -\sum_{n=1}^{2} \alpha_n \beta_n^2 \phi_n^A \quad (A1c)$$

where the constant p_n^A is the eigenvalue of a characteristic equation derived in a Stroh-like formalism and given as

$$p_n^A = \left[\frac{(\bar{A}_{12}-A_{12})(\bar{A}_{12}+A_{12}+2A_{66})}{4A_{22}A_{66}}\right]^{1/2} + (-1)^{n+1}\left[\frac{(\bar{A}_{12}+A_{12})(\bar{A}_{12}-A_{12}-2A_{66})}{4A_{22}A_{66}}\right]^{1/2} \quad (A2)$$

where $\bar{A}_{12}=\sqrt{A_{11}A_{22}}$. The constants α_n and β_n are, respectively,

$$\alpha_n = \frac{p_n^A}{2\pi A_{66}(A_{11}+(p_n^A)^2 A_{12})(A_{11}-(p_n^A)^2(A_{12}+2A_{66}))} \quad (A3a)$$

$$\beta_n = (A_{11}-(p_n^A)^2 A_{66}) \quad (A3b)$$

The functions ϕ_n^A and ψ_n^A are, respectively,

$$\phi_n^A = \ln r_n^A \quad (A4a)$$

$$\psi_n^A = \tan^{-1}\left(\frac{p_n^A(x_2-x_2')}{x_1-x_1'}\right) \quad (A4b)$$

where r_n^A is defined as

$$r_n^A = [(x_1-x_1')^2 + (p_n^A)^2(x_2-x_2')^2]^{1/2} \quad (A5)$$

The functions ϕ_n^A and ψ_n^A are called quasi-harmonic functions, because they are not harmonic functions in the coordinates (x_1,x_2) but harmonic in the transformed coordinates $(x_1,p_n^A x_2)$; thus

$$\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial(p_n^A x_2)^2}\right)\phi_n=0 \tag{A6}$$

with no summation over the subscript n. The Green's function G_{ij} is symmetric in ij, that is, $G_{ij}=G_{ji}$.

For a bending problem, the slope of the mid-plane θ_i is used to evolve the inclusion problem. The slope θ_i corresponds to the components of the derivatives of the Green's functions as $\theta_i=-G_{33,i}$. The explicit form for the derivatives of the Green's functions in the Cartesian coordinates are given as

$$G_{33,1}=\sum_{n=1}^{2}\gamma_n(x_1\phi_n^D-p_n^D x_2\psi_n^D) \tag{A7a}$$

$$G_{33,2}=-\sum_{n=1}^{2}p_n^D\gamma_n(p_n^D x_2\phi_n^D+x_1\psi_n^D) \tag{A7b}$$

where the constant p_n^D is the eigenvalue of the characteristic equation for a bending problem of an orthotropic Kirchhoff plate, derived in a Stroh-like formalism and given as

$$p_n^D=\left[\frac{1}{2}\left(\frac{D_{12}+2D_{66}}{D_{22}}+\sqrt{\frac{D_{11}}{D_{22}}}\right)\right]^{1/2}+(-1)^{n+1}\left[\frac{1}{2}\left(\frac{D_{12}+2D_{66}}{D_{22}}-\sqrt{\frac{D_{11}}{D_{22}}}\right)\right]^{1/2} \tag{A8}$$

The constants γ_n are

$$\gamma_n=\frac{p_n^D}{2\pi(D_{11}-(p_n^D)^4D_{22})} \tag{A9}$$

The functions ϕ_n^D and ψ_n^D correspond to the functions ϕ_n^A and ψ_n^A, respectively, when p_n^A is replaced by p_n^D.

Appendix B: Logarithmic Quasi-Newtonian Potential

The two-dimensional integral $\int_{\Omega^*}(1/p_n^*)\ln r_n\, d\Omega^*(x'_1,p_n^* x'_2)$ in Eq. (29) could be obtained after some manipulation of the integration for a three-dimensional ellipsoid, where the longest axis in the x_3-direction approaches infinity. More details of the evolution of this are given in the paper of Yang et al. [14], in which the integration result of the potential function is

$$\Phi_n^*=-\frac{1}{4p_n^*}(I^N(\lambda_N)-\bar{x}_r\bar{x}_r I_R^N(\lambda_N)) \tag{B1}$$

where the repeated lower case indices are summed from 1 to 2, and the upper case indices take on the same numbers as the corresponding lower case ones but they are not summed; for example,

$$\bar{x}_r\bar{x}_r I_R^N(\lambda_N)=\bar{x}_1^2 I_1^N(\lambda_N)+\bar{x}_2^2 I_2^N(\lambda_N)$$

$$\frac{I^N(\lambda_N)}{p_n^*}=\frac{I^1(\lambda_1)}{p_1^*}\quad\text{where } n=1=\frac{I^2(\lambda_2)}{p_2^*}\quad\text{where } n=2 \tag{B2}$$

The transformed coordinates $\bar{x}_1$ and $\bar{x}_2$ are x_1 and $p_n^* x_2$, respectively. The closed form of the functions $I^N(\lambda_N)$ and $I_R^N(\lambda_N)$ are presented as follows:

$$I^N(\lambda_N)=-4\pi a_1p_n^*a_2[\ln(\sqrt{a_1^2+\lambda_N}+\sqrt{(p_n^*a_2)^2+\lambda_N})-\ln(a_1+p_n^*a_2)] \tag{B3a}$$

$$I_1^N(\lambda_N)=\frac{4\pi a_1p_n^*a_2}{(p_n^*a_2)^2-a_1^2}\left[\sqrt{\frac{(p_n^*a_2)^2+\lambda_N}{a_1^2+\lambda_N}}-1\right] \tag{B3b}$$

$$I_2^N(\lambda_N)=\frac{4\pi a_1p_n^*a_2}{a_1^2-(p_n^*a_2)^2}\left[\sqrt{\frac{a_1^2+\lambda_N}{(p_n^*a_2)^2+\lambda_N}}-1\right] \tag{B3c}$$

where λ_N is zero for the interior point $\mathbf{x}$ of Ω and for the exterior point is the largest positive root of the following equation:

$$\frac{x_1^2}{a_1^2+\lambda_N}+\frac{(p_n^*x_2)^2}{(p_n^*a_2)^2+\lambda_N}=1 \tag{B4}$$

which is given by

$$\lambda_N=\tfrac{1}{2}[\bar{x}_m\bar{x}_m-\bar{a}_m\bar{a}_m+\sqrt{(\bar{x}_m\bar{x}_m-\bar{a}_m\bar{a}_m)^2+4(\bar{x}_1^2\bar{a}_2^2+\bar{x}_2^2\bar{a}_1^2-\bar{a}_m\bar{a}_m)}] \tag{B5}$$

where the transformed semi-axes of the ellipse, $\bar{a}_1$ and $\bar{a}_2$, are a_1 and $p_n^*a_2$, respectively.

Appendix C: Explicit Expression of the Tensor $\mathbf{T}_{ijklm}$

The two-dimensional integral $\int_\Omega x'_k\cdots x'_l \ln r_n d\Omega$ could be obtained after some manipulation of the integration of a three-dimensional harmonic potential for an ellipsoid. More details are given in a paper by Yang et al. [14]. Introducing integrals defined by

$$\Phi_n^{(i)}=\int_\Omega \bar{x}'_i \ln r_n^D\, d\Omega,\quad i=1,2 \tag{C1}$$

the components of the tensors T_{ijklm}, which is symmetric with respect to the interchange of i and j or k and l, could be explicitly written as

$$T_{11111}=\sum_{n=1}^{2}\nu_n\gamma_n\Phi_{n,11}^{(1)}\quad T_{11112}=\sum_{n=1}^{2}\nu_n\gamma_n\Phi_{n,11}^{(2)}$$

$$T_{11121}=\sum_{n=1}^{2}2D_{66}\gamma_n\Phi_{n,21}^{(1)}\quad T_{11122}=\sum_{n=1}^{2}2D_{66}\gamma_n\Phi_{n,21}^{(2)}$$

$$T_{11221}=-\sum_{n=1}^{2}\left(4D_{66}-\frac{\nu_n}{(p_n^D)^2}\right)\gamma_n\Phi_{n,11}^{(1)}\quad T_{11222}=-\sum_{n=1}^{2}\left(4D_{66}-\frac{\nu_n}{(p_n^D)^2}\right)\gamma_n\Phi_{n,11}^{(2)}$$

$$T_{12111}=\sum_{n=1}^{2}\nu_n\gamma_n\Phi_{n,12}^{(1)}\quad T_{12112}=\sum_{n=1}^{2}\nu_n\gamma_n\Phi_{n,12}^{(2)}$$

$$T_{12121}=\sum_{n=1}^{2}2D_{66}\gamma_n\Phi_{n,22}^{(1)}\quad T_{12122}=\sum_{n=1}^{2}2D_{66}\gamma_n\Phi_{n,22}^{(2)}$$

$$T_{12221}=-\sum_{n=1}^{2}\left(4D_{66}-\frac{\nu_n}{(p_n^D)^2}\right)\gamma_n\Phi_{n,12}^{(1)}$$

$$T_{12222}=-\sum_{n=1}^{2}\left(4D_{66}-\frac{\nu_n}{(p_n^D)^2}\right)\gamma_n\Phi_{n,12}^{(2)}$$

$$T_{22111}=\sum_{n=1}^{2}\nu_n\gamma_n\Phi_{n,22}^{(1)}\quad T_{12112}=\sum_{n=1}^{2}\nu_n\gamma_n\Phi_{n,22}^{(2)}$$

$$T_{22121}=-\sum_{n=1}^{2}2D_{66}\gamma_n(p_n^D)^2\Phi_{n,12}^{(1)}\quad T_{22122}=-\sum_{n=1}^{2}2D_{66}\gamma_n(p_n^D)^2\Phi_{n,12}^{(2)}$$

$$T_{22221} = -\sum_{n=1}^{2} \left(4D_{66} - \frac{\nu_n}{(p_n^D)^2}\right)\gamma_n \Phi_{n,22}^{(1)} \tag{C2}$$

$$T_{22222} = -\sum_{n=1}^{2} \left(4D_{66} - \frac{\nu_n}{(p_n^D)^2}\right)\gamma_n \Phi_{n,22}^{(2)}$$

where the logarithmic potential $\Phi_n^{(i)}$ is expressed in I-integrals as

$$\Phi_n^{(1)} = -\frac{1}{4p_n^D} a_1^2 x_1 [I_1^N(\lambda_N) - \bar{x}_r \bar{x}_r I_{R1}^N(\lambda_N)] \tag{C3a}$$

$$\Phi_n^{(2)} = -\tfrac{1}{4} a_2^2 p_n^D x_2 [I_2^N(\lambda_N) - \bar{x}_r \bar{x}_r I_{R2}^N(\lambda_N)] \tag{C3b}$$

where the higher order I-integrals $I_{ij}^N(\lambda_N)$ could be expressed by the lower ones $I_i^N(\lambda_N)$ given in Appendix B as [14]

$$I_{12}^N(\lambda_N) = I_{21}^N(\lambda_N) = \frac{I_2^N(\lambda_N) - I_1^N(\lambda_N)}{a_1^2 - (p_n^D a_2)^2} \tag{C4a}$$

$$3I_{11}^N(\lambda_N) = \frac{4\pi a_1 p_n^D a_2}{(a_1^2 + \lambda_N)\Delta(\lambda_N)} - I_{12}^N(\lambda_N) \tag{C4b}$$

$$3I_{22}^N = \frac{4\pi a_1 p_n^D a_2}{((p_n^D a_2)^2 + \lambda_N)\Delta(\lambda_N)} - I_{12}^N(\lambda_N) \tag{C4c}$$

References

[1] Eshelby, J. D., 1957, "The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems," Proc. R. Soc. London, Ser. A, **241**, pp. 376–396.
[2] Eshelby, J. D., 1959, "The Elastic Field Outside on Ellipsoidal Inclusion," Proc. R. Soc. London, Ser. A, **252**, pp. 561–569.
[3] Eshelby, J. D., 1961, "Elastic Inclusions and Inhomogeneities," *Progress in Solid Mechanics 2*, I. N. Sneddon and R. Hill, eds., North-Holland, Amsterdam, pp. 89–140.
[4] Mura, T., 1982, *Micromechanics of Defects in Solids*, Martinus Nijhoff Publishers, Hague, Netherlands.
[5] Jaswon, M. A., and Bhargava, R. D., 1961, "Two-dimensional Elastic Inclusion Problems," Proc. Cambridge Philos. Soc., **57**, pp. 669–680.
[6] Willis, J. R., 1964, "Anisotropic Elastic Inclusion Problems," Q. J. Mech. Appl. Math., **17**, pp. 157–174.
[7] Bhargava, R. D., and Radhakrishna, H. C., 1964, "Elliptic Inclusion in Orthotropic Medium," J. Phys. Soc. Jpn., **19**, pp. 396–405.
[8] Hwu, C., and Ting, T. C. T., 1989, "Two-dimensional Problems of the Anisotropic Elastic Solid With an Elliptic Inclusion," Q. J. Mech. Appl. Math., **42**, pp. 553–572.
[9] Beom, H. G., 1998, "Analysis of a Plate Containing an Elliptic Inclusion With Eigencurvatures," Arch. Appl. Mech., **68**, pp. 422–432.
[10] Beom, H. G., and Earmme, Y. Y., 1999, "The Elastic Field of an Elliptic Cylindrical Inclusion in a Laminate With Multiple Isotropic Layers," ASME J. Appl. Mech., **66**, pp. 165–171.
[11] Beom, H. G., and Kim, I. B., 1999, "Analysis of a Multilayered Plate Containing a Cuboidal Inclusion With Eigenstrains," Mech. Mater., **31**, pp. 729–741.
[12] Duong, C. N., and Yu, J., 2003, "Analysis of a Plate Containing a Polygon-Shape Inclusion With a Uniform Eigencurvature," ASME J. Appl. Mech., **70**, pp. 404–407.
[13] Duong, C. N., and Yu, J., 2003, "Thermal Stresses in a One-Sided Bonded Repair by a Plate Inclusion Model," J. Therm. Stresses, **26**, pp. 457–466.
[14] Yang, K. J., Beom, H. G., and Kang, K. J., 2005, "Thermal Stress Analysis for an Inclusion With Nonuniform Temperature Distribution in an Infinite Kirchhoff Plate," J. Therm. Stresses **28**, pp. 1123–1144.
[15] Yang, K. J., and Kang, K. J., 2005, "Green's Functions for a Unit Concentrated Force in an Infinite Orthotropic Kirchhoff Plate" (submitted to Mechanics Research Communications).
[16] Sendeckyj, G. P., 1967, "Ellipsoidal Inhomogeneity Problem," Ph.D. thesis, Northwestern University, Evanston.
[17] Jones, R. M., 1975, *Mechanics of Composite Materials*, McGraw-Hill, New York.
[18] 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.
[19] Cheng, Z. Q., and Reddy, J. N., 2002, "Octet Formalism for Kirchhoff Anisotropic Plates," Proc. R. Soc. London, Ser. A, **458**, pp. 1499–1517.
[20] Hwu, C., 2003, "Stroh-Like Formalism for the Coupled Stretching-Bending Analysis of Composite Laminates," Int. J. Solids Struct., **40**, pp. 3681–3705.
[21] Cheng, Z. Q., and Reddy, J. N., 2003, "Green's Functions for Infinite and Semi-Infinite Anisotropic Thin Plates," ASME J. Appl. Mech., **70**, pp. 260–267.
[22] Cheng, Z. Q., and Reddy, J. N., 2004, "Green's Functions for an Anisotropic Thin Plate With a Crack or an Anticrack," Int. J. Eng. Sci., **42**, pp. 271–289.
[23] Cheng, Z. Q., and Reddy, J. N., 2004, "Laminated Anisotropic Thin Plate With an Elliptic Inhomogeneity," Mech. Mater., **36**, pp. 647–657.
[24] Hsieh, M. C., and Hwu, C., 2002, "Anisotropic Elastic Plates With Holes/Cracks/Inclusions Subjected to Out-of-Plane Bending Moments," Int. J. Solids Struct., **39**, pp. 4905–4925.

P. Sharma
e-mail: psharma@uh.edu

L. T. Wheeler

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

Size-Dependent Elastic State of Ellipsoidal Nano-Inclusions Incorporating Surface/Interface Tension

Using a tensor virial method of moments, an approximate solution to the relaxed elastic state of embedded ellipsoidal inclusions is presented that incorporates surface/interface energies. The latter effects come into prominence at inclusion sizes in the nanometer range. Unlike the classical elastic case, the new results for ellipsoidal inclusions incorporating surface/interface tension are size-dependent and thus, at least partially, account for the size-effects in the elastic state of nano-inclusions. For the pure dilatation case, exceptionally simple expressions are derived. The present work is a generalization of a previous research that addresses simplified spherical inclusions. As an example, the present work allows us, in a straightforward closed-form manner, the study of effect of shape on the size-dependent strain state of an embedded quantum dot.
[DOI: 10.1115/1.2338052]

1 Introduction

The determination of elastic states of an embedded inclusion is of considerable importance in a wide variety of physical problems. In the classical elasticity context this problem was first solved rigorously by [1]. The latter work, both with and without modifications, has been employed to tackle a diverse set of problems: Localized thermal heating, residual strains, dislocation-induced plastic strains, phase transformations, overall or effective elastic, plastic and viscoplastic properties of composites, damage in heterogeneous materials, quantum dots, interconnect reliability, microstructural evolution, to name a few. The classical solution of an embedded inclusion neglects the presence of surface or interface energies and indeed, the effects of those are negligible except in the size range of tens of nanometers, where one contends with a significant surface-to-volume ratio. Clearly, the influence of surface/interface energies only extends to the nanoscale regime, as illustrated by various mechanical and optoelectronic applications such as nanostructures, nanocomposites, thin films, nanoelectronics, and quantum dots [2–10]. With this in mind, in a recent work, Sharma and Ganti [8] extended Eshelby's formalism to cover nano-inclusions by incorporating coupled bulk-surface elasticity. Explicit size-dependent expressions were presented for the elastic state of simplified spherical and circular shapes under radially symmetric loadings along with several illustrative applications. Sharma and Ganti [8] conclude that for shapes that admit a constant curvature (e.g., sphere, circular shape) and subjected to radial loadings, the elastic states are uniform. In the present work, we revisit the inclusion problem to address the more generalized ellipsoidal shape. We note here also several other works that have recently appeared addressing surface energy effects in inclusion problems: Duan et al. [11] generalized the work of Sharma and Ganti [8] to incorporate eigenstrains or loadings of arbitrary symmetry in spherical inhomogeneities. They find that interior stresses and strains in spherical inhomogeneities are *nonuniform* if the eigenstrain or applied stresses are not radially symmetric. This conclusion is also affirmed independently by the work of Lim et al. [12]. Effective size-dependent properties of composites containing spherical inclusions have been addressed by Duan et al. [11] while Dingreville et al. [13] investigate the "effective" properties of nanoparticles, nanowires, and thin films.

Apart from the aforementioned works related to surface energy, in the classical elasticity context, extensive work has been done on the embedded inclusion problem and related issues. For the sake of conciseness, and given the existence of several readily available reviews on this topic, a detailed literature survey is avoided in this work. Wherever appropriate, relevant papers are cited contextually. The reader is referred to the following monographs, review articles, books and references therein: Mura [14], Nemat-Nasser and Hori [15], Markov and Preziosi [16], Weng et al. [17], Bilby et al. [18], Mura et al. [19], and Mura [20].

In Sec. 2, we discuss some preliminary concepts and issues related to surface/interface energies and inclusions. In Secs. 3 and 4, we employ the tensor virial method of moments to establish approximate solutions of various orders to the ellipsoidal nano-inclusion problem. Some simple expressions are deduced for the purely dilatational problem in Sec. 5. Numerical results and the application to quantum dots are presented in Sec. 6 followed with a summary and the major conclusions.

2 Preliminaries and Background

In this section, the mathematical preliminaries draw on the formulations of Gurtin and Murdoch [21], Murdoch [22], and Gurtin et al. [23]. In the context of inclusions, a reference to the words "surface" or "interface" is meant for the internal free surface of cavities or adjoining region of the solid inclusion and the surrounding matrix. In the present work, we shall use these words and their variants interchangeably.

Consider an arbitrarily shaped smooth interface between an embedded inclusion and surrounding host matrix, characterized by a unit normal **n**. Let this interface be "attached" to the bulk (i.e., both inclusion and matrix) without slipping or any other discontinuity of displacements across it. This implies that we consider only a coherent interface. In contrast to the classical case where surface energies are neglected, we now require that the interface of the inclusion and the matrix be endowed with a deformation

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received October 1, 2005; final manuscript received April 12, 2006. Review conducted by M. R. Begley. 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.

dependent interfacial energy density, Γ. The interfacial or surface energy density is positive definite. This quantity is distinct from the bulk deformation dependent energy density due to the different coordination number of the surface/interface atoms, different bond lengths, angles and a different charge distribution [24].

Within the assumptions of infinitesimal deformation and a continuum field theory, the concept of surface stress and surface tension can be clarified by the (assumed *linearized*) relation between the interface/surface stress tensor, $\sigma^{\mathbf{s}}$, and the deformation dependent surface energy, $\Gamma(\varepsilon^{\mathbf{s}})$

$$\sigma^{\mathbf{s}} = \tau_o \mathbf{P}^{\mathbf{s}} + \frac{\partial \Gamma}{\partial \varepsilon^{\mathbf{s}}} \tag{1}$$

Where applicable, superscripts B and S indicate bulk and surface, respectively. Here, $\boldsymbol{\varepsilon}^{\mathbf{s}}$ is the strain tensor for surfaces that will result from the projection of the conventional bulk strain tensor on to the tangent plane of the surface or interface while τ_o is the deformation independent surface/interfacial tension. The surface projection tensor, $\mathbf{P}^{\mathbf{s}}$ which maps tensor fields from bulk to surface and vice versa is defined as

$$\mathbf{P}^{\mathbf{s}} = \mathbf{I} - \mathbf{n} \otimes \mathbf{n} \tag{2}$$

Here, "$\mathbf{I}$" is the identity tensor. Consider an arbitrary vector $\mathbf{v}$. The tensor or dyadic product $\otimes$ extends two vectors into a second order tensor, i.e., in indical notation, $A_{ij}=a_i b_j$. The surface gradient and surface divergence, then, take the following form [23]:

$$\begin{aligned} \nabla_s \mathbf{v} &= \nabla \mathbf{v} \mathbf{P}^s \\ \mathrm{div}_s(\mathbf{v}) &= \mathrm{Tr}(\nabla_s \mathbf{v}) \end{aligned} \tag{3}$$

Here we have also defined the surface gradient operator (∇_s) and the surface divergence (div_s). "Tr" indicates the trace operation. We shall employ both index and boldface notation as convenient. Unless otherwise stated, all tensors are referred to a Cartesian basis and isotropy is assumed throughout.

The equilibrium and isotropic constitutive equations of bulk elasticity are

$$\mathrm{div}\, \boldsymbol{\sigma}^{\mathbf{B}} = 0 \tag{4a}$$

$$\boldsymbol{\sigma}^{\mathbf{B}} = \lambda \mathbf{I}\mathrm{Tr}(\boldsymbol{\varepsilon}) + 2\mu\varepsilon \tag{4b}$$

At the interface, the concept of surface or interface elasticity [21,23], ordinarily ignored in the classical formulation is introduced

$$[\boldsymbol{\sigma}^B \cdot \mathbf{n}] + \mathrm{div}_s\, \sigma^S = 0 \tag{5a}$$

$$\sigma^s = \tau_o \mathbf{P}^s + 2(\mu^s - \tau_o)\varepsilon^s + (\lambda^s + \tau_o)\mathrm{Tr}(\varepsilon^s)\mathbf{P}^{\mathbf{s}} \tag{5b}$$

Here, λ and μ are the Lamé constants for the isotropic bulk material. Isotropic interfaces or surfaces can be characterized by surface Lamé constants λ^s, μ^s, and surface tension, τ_o. The square brackets indicate a jump of the field quantities across the interface. It is to be noted that only certain strain components appear within the constitutive law for surfaces due to the 2×2 nature of the surface stress tensor (i.e., only the tangential projection of the strains on the interface are included consequently, $\mathbf{P}^s . \mathbf{n}=\mathbf{0}$). In the absence of surface terms, Eq. (5*a*) reduce to the familiar normal traction continuity equations.

Thus, while the infinitesimal strain tensor in the bulk (both inclusion and matrix) is defined as usual in Eq. (6), the surface strains involve the use of projection tensor (Eq. (7))

$$\varepsilon = \tfrac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T) \tag{6}$$

$$\boldsymbol{\varepsilon}^{\mathbf{s}} = \tfrac{1}{2}(\mathbf{P}^s \nabla_s \mathbf{u} + \nabla_s \mathbf{u}^T \mathbf{P}^S) \tag{7}$$

Implicit in Eq. (7) is our assumption of a coherent interface. An incoherent interface requires additional measures of strain. The most generalized treatment of deformation measures that allows projections of jump in the displacement gradient (i.e., relative strain, twist, normal shear, tilt) and jumps in the normal and tangential displacement (stretch and slip) in addition to the coherent deformation considered in Eq. (7) has been addressed in detail by Gurtin et al. [23].

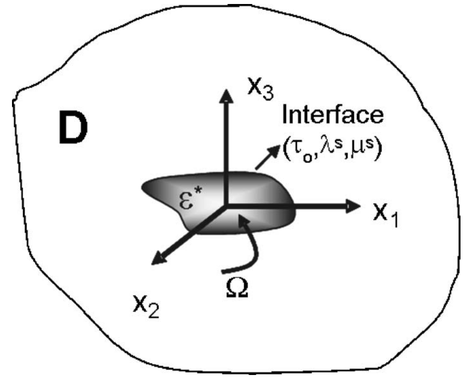


Fig. 1 Schematic of the problem

Consider a stress-free uniform transformation strain prescribed within the domain of the inclusion (Fig. 1). As per Mura's definition of an inclusion [14] we assume (*for the moment*) identical material properties for the inclusion and the matrix. The scenario where material properties differ is referred to as the inhomogeneity problem [14] and will be discussed in Sec. 6.

Sharma and Ganti [8] have derived the following general integral equation for arbitrary shaped inclusion linking the actual strain in the inclusion to the uniform transformation strain

$$\boldsymbol{\varepsilon} = \mathbf{S}{:}\boldsymbol{\varepsilon}^* + \mathrm{sym}\left\{ \nabla_x \otimes \underbrace{\int_S \mathbf{G}^{\mathbf{T}}(\mathbf{y}-\mathbf{x}) \,.\, \mathrm{div}_s \boldsymbol{\sigma}^{\mathbf{s}}(\mathbf{y}) dS_y} \right\} \tag{8}$$

Here $\mathbf{G}$ is the Green's tensor for isotropic classical elasticity [1]. ε^* is the so-called "eigenstrain" or a stress-free transformation strain such as due to for example, phase transformation, thermal expansion mismatch, lattice mismatch among others. The underlined term in Eq. (8) indicates the extra contributions due to surface/interface energy. $\mathbf{S}$ is the classical size-independent Eshelby tensor (see [1], Appendix A and [14]). The notation, sym{.}, represents the symmetric part of a second order tensor, $\mathbf{A}$, e.g., $\mathrm{sym}\{\mathbf{A}\}=\frac{1}{2}\{\mathbf{A}+\mathbf{A}^T\}$. Some further details on Eshelby's tensor formalism are included in Appendix A for ready reference.

Further simplification of Eq. (8) is difficult without additional assumptions regarding inclusion shape. Equation (8) implicitly gives the modified Eshelby's tensor for inclusions incorporating surface energies. This relation is implicit since the surface stress depends on the surface strain, which in turn is the projection of the conventional strain (ε) on the tangent plane of the inclusion-matrix interface.

For the spherical shape, Eq. (8) can be explicitly re-written as [8]

$$\boldsymbol{\varepsilon} = \mathbf{S}{:}\boldsymbol{\varepsilon}^* - \frac{2(\lambda^s + \mu^s)}{3KR_o}(\mathbf{S}{:}\mathbf{I})\mathrm{Tr}(\mathbf{P}^{\mathbf{s}}\boldsymbol{\varepsilon}\mathbf{P}^{\mathbf{s}}) - \frac{2\tau_o}{3KR_o}(\mathbf{S}{:}\mathbf{I}) \tag{9}$$

Here we note that for a sphere, curvature is the reciprocal of its radius, $\kappa = 1/R_o$ and K is the bulk modulus. To allow easy manipulation and in order to get analytical form of Eshelby tensor, it is necessary to resolve the fourth order tensors along the following basis: $\frac{1}{3}\delta_{ij}\delta_{kl}$ and $\frac{1}{2}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk})-\frac{1}{3}\delta_{ij}\delta_{kl}$ (see for example, [25]). And using the spherical Eshelby tensor [14], Eq. (9) can be rewritten (in spherical polar basis)

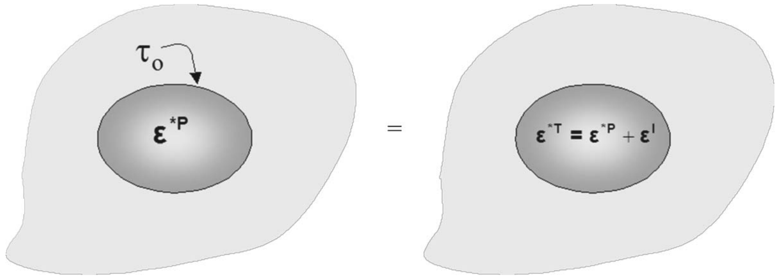


Fig. 2 Schematic of the solution

$$\varepsilon_{rr} = \varepsilon_{\theta\theta} = \varepsilon_{\phi\phi} = \frac{3K\varepsilon^* - 2\tau_o/R_o}{4\mu + 3K + 4(\lambda^s + \mu^s)/R_o} \tag{10}$$

A general feature of the integral Eq. (8) can be realized by noting that

$$\mathrm{div}_s\, \boldsymbol{\sigma}^{\mathbf{s}} = \mathrm{div}_s\{\mathbf{C}^{\mathbf{s}}\mathbf{P}^{\mathbf{s}}\boldsymbol{\varepsilon}\mathbf{P}^{\mathbf{s}} + \tau_o\mathbf{P}^{\mathbf{s}}\} \tag{11}$$

The surface divergence of surface stress tensor can only be uniform if the classical "bulk" strain *as well as* the projection tensor is uniform over the inclusion surface. Consider the identity

$$\mathrm{div}_s\mathbf{P}^{\mathbf{s}} = 2\kappa\mathbf{n} \tag{12}$$

Here κ is the mean curvature of the inclusion. For a general ellipsoid the curvature is nonuniform and varies depending upon the location at the surface. Thus, unlike for the spherical or circular shape, even for purely symmetric transformation strains such as pure dilation, we can expect a nonuniform strain state rendering perhaps impossible an exact solution of Eq. (8) for the general ellipsoidal shape.

3 Tensor Virial Method of Moments for Approximations to the Ellipsoidal Problem

The nonuniformity of the ellipsoidal curvature makes it difficult to solve for the elastic state via the implicit system of integral equations listed in Eq. (8). The direct use of the integral equation (8) is not very convenient for our purposes and thus in the remainder of this paper, based on some simplifying assumptions, we explore the so-called tensor virial method of moments [26] to deduce an approximate solution for an ellipsoidal inclusion undergoing a uniform transformation strain.

First, we only consider the effect of surface tension (i.e., τ_o) and ignore deformation dependent surface elasticity, for example, in the result of Eq. (10), the term $4(\lambda^s+\mu^s)/R_0$ would be discarded. This assumption is reasonable for small strains and indeed, as has been found in some technological applications, the deformation dependent surface elasticity effects can often be small compared to surface tension effect. Of course, in certain classes of problems, essential physics is lost by abandoning the deformation dependent surface elasticity (e.g., effective properties of nanocomposites, [8,11]; dislocation nucleation in *flat* nanosized thin films, [3]). However, for several problems of technological interest, consideration of τ_o is sufficient. Note that the recent work of Yang [27] may require further clarification as he proposes that the effective size-dependent properties of nanocomposites depend on surface tension of inhomogeneities and fails to include the effect of surface elasticity which (according to us) is the sole contributor to the effect Yang purports to investigate. Since the effect of surface tension manifests itself as a residual type effect (i.e., independent of external loading), we can immediately employ Eshelby's classic *gendanken* of cutting and welding operations [1] to put a physical perspective on the problem. Take the inclusion (containing a prescribed physical eigenstrain, say, a thermal expansion mismatch strain or that due to lattice mismatch) out of the matrix but with a surface tension equivalent to the interfacial tension of inclusion-matrix. From a classical perspective the inclusion should relax to a strain equal to the physical eigenstrain. However, in the context of coupled surface-bulk elasticity, an additional strain ensues due to the presence of interfacial tension. Thus the total effective eigenstrain is equal to the superposition of the initial prescribed eigenstrain (due to a physical mechanism) and the strain state of an ***isolated un-embedded*** inclusion under the action of a surface tension. This is shown schematically in Fig. 2. Mathematically

$$\boldsymbol{\varepsilon}^{*\mathbf{T}}(\mathbf{x}) = \boldsymbol{\varepsilon}^{*\mathbf{P}}(\mathbf{x}, \text{physical cause}) + \boldsymbol{\varepsilon}^{\mathbf{I}}(\mathbf{x}, \tau_o, \kappa) \tag{13}$$

Here we have indicated the major functional dependence of each type of eigenstrain. The superscripts "T," "P," and "I" stand for "total," "physical," and "isolated" respectively. In summary, if we are able to evaluate $\boldsymbol{\varepsilon}^I$, Eshelby's classical tensor type concept can be employed to determine the elastic state of the inclusion incorporating surface energy, i.e.,

$$\boldsymbol{\varepsilon}(\mathbf{x}) = \hat{\mathbf{S}}(\mathbf{x}) : \boldsymbol{\varepsilon}^{*\mathbf{T}}(\mathbf{x}) \tag{14}$$

Here $\hat{\mathbf{S}}$ is an Eshelby tensor type integral operator which defaults to Eshelby's classical tensor ($\mathbf{S}$) for *uniform* eigenstrains. For an inhomogeneous eigenstrain, the actual strain is then determined by the action of the integral operator, $\hat{\mathbf{S}}$. This distinction is necessary since even though (in this work) the physical eigenstrain is assumed to be (as often is) uniform, the contribution of surface tension is nonuniform, except in the case of a sphere and infinitely long circular cylindrical shape. Further details on $\hat{\mathbf{S}}$ are found in Appendix A. Note that the size dependence enters via the simulated eigenstrain, $\boldsymbol{\varepsilon}^I$. For spherical and cylindrical shapes, the calculation of this strain is trivial. For example, in the spherical inclusion case, the isolated strain (in spherical polar basis) is merely $\varepsilon^I_{rr}=\varepsilon^I_{\theta\theta}=\varepsilon^I_{\phi\phi}=-2\tau_o/3KR_o$. Substituting this into (13) and using Eshelby's tensor (Appendix A) reproduces Eqs. (9) and (10) and the result of Sharma and Ganti [8] provided that the surface elasticity constants, $\{\mu^s, \lambda^s\} \to 0$ in their formula, are set to zero which then also coincides with the result of Cahn and Larche [28].

For an ellipsoidal inclusion, even the isolated solution is not so trivial. We construct an approximate solution using the tensor virial method developed by Chandrasekhar [26,29].

Consider an isolated (i.e., un-embedded) triaxial ellipsoid made from the same material as the inclusion (Fig. 3). Further, let it

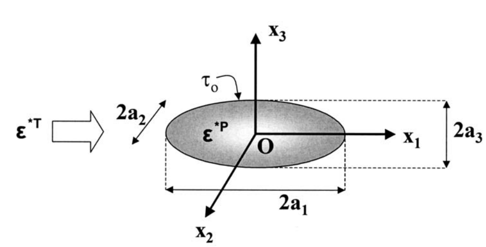


Fig. 3 Schematic of the problem for the isolated ellipsoidal particle under a surface tension

contain an eigenstrain identical to that of the inclusion, $\boldsymbol{\varepsilon}^{*P}$. Now, to incorporate surface energy (as per the discussion of the preceding paragraph), let the isolated ellipsoidal particle also be under the influence of a surface tension numerically equivalent to the interfacial tension of inclusion-matrix. A coordinate system with an origin at the ellipsoid centroid coincident with the principal axes of the ellipsoid (a_1, a_2, a_3) is adopted.

In the absence of any external body forces and in static equilibrium, the equation of motion for the isolated ellipsoidal particle can be written as

$$\operatorname{div} \boldsymbol{\sigma}^{\mathbf{I}} = 0 \tag{15a}$$

$$\boldsymbol{\sigma}^{\mathbf{I}} = \mathbf{C}:\boldsymbol{\varepsilon}^{\mathbf{I}} \tag{15b}$$

Where we note that $\boldsymbol{\varepsilon}^{I}$ can be found by simply inverting Eq. (15b). We take the first moment of Eq. (15a) to obtain

$$\int_V \mathbf{x}_i \frac{\partial \boldsymbol{\sigma}_{ik}^I}{\partial \mathbf{x}_k} dV = \int_V \frac{\partial(\mathbf{x}_j \boldsymbol{\sigma}_{ik}^I)}{\partial \mathbf{x}_k} dV - \int_V \sigma_{ij}^I dV = 0 \tag{16a}$$

$$\Rightarrow \int_S \mathbf{x}_j \sigma_{ik}^{I(-)} dS_k - \int_V \sigma_{ij}^I dV = 0 \tag{16b}$$

Gauss's theorem is used to obtain the surface integral in Eq. (16b). A superscript $(-)$ sign indicates that the referred quantity is applicable at distances infinitesimally close to the bounding surface but located in the interior. Since the inclusion is isolated, the $(+)$ quantities are identically zero. Recall that for uniform surface tension, the jump in the normal tractions is proportional to the surface tension (Eq. (5a)), i.e.,

$$[\sigma^I \,.\, \mathbf{n}] = -\,\sigma^{I(-)} \,.\, \mathbf{n} = -\,\tau_o \mathbf{n} \operatorname{div} \mathbf{n} \tag{17}$$

From which we obtain the relation

$$\int_V \sigma_{ij}^I dV = -\,\tau_o \int_S \mathbf{x}_j \operatorname{div} \mathbf{n} dS_i \tag{18}$$

Consistent with the notion of taking a first order moment, we assume that $\boldsymbol{\sigma}^I$ is uniform within the ellipsoidal particle. Then, as clearly apparent, Eq. (18) furnishes us a method to find the "average" strain of an isolated inclusion under the action of surface tension and eigenstrain via a (relatively) simpler surface integral on the right hand side. Note that, effectively, Eq. (18) approximates the nonuniform boundary condition at the inclusion-matrix interface in Eq. (5a) in an "average" sense. Implicit in this assumption is the notion that the average inclusion strain in the ellipsoidal inclusion is representative of the actual nonuniform strain that one would obtain from a rigorous solution of the integral equations in Eq. (8). Equation (18), which is the heart of the first moment approximation, is exact in three cases: (i) For inclusions with constant, curvature, i.e., spherical and circular shape (ii) the trivial case when surface tension is absent, (iii) for large "inclusion" size where curvature is effectively negligible. Clearly, the first moment approximation is expected to be inaccurate when ellipsoidal aspect ratios become extreme, e.g., flat crack like inclusion, and as such should be avoided. Since we are constructing an approximation for an ellipsoid as the associated fields depart marginally from the uniform case (sphere), our solution is not expected to satisfy the two-dimensional asymptotic limit when the ellipsoid degenerates to a circular cylinder (for which also the exact analytical solutions are known). To approximate an elliptical cylinder, our analysis must be repeated in a two-dimensional framework (although in that case, complex analysis tools may be more efficient possibly leading to exact solutions).

Using a Lemma by Rosenkilde [30,31]—see Appendix B, the surface integral in Eq. (18) can be further simplified to

$$\tau_o \int_S \mathbf{x}_j \mathbf{n}_{k,k} dS_i = \tau_o \int_S (\delta_{ij} - \mathbf{n}_i \mathbf{n}_j) dS = 2\mathbf{M}_{ij} \tag{19}$$

Note that the rightmost integrand in Eq. (19) is just the surface projection tensor. The derivation of Eq. (19) and other ellipsoidal surface integrals involving higher order moments of div $\mathbf{n}$ are given in Appendix B. We can then write the surface contributed eigenstrain of the ellipsoidal inclusion in the following simple manner:

$$\boldsymbol{\varepsilon}^I = -\frac{2}{V}(\mathbf{C})^{-1}\mathbf{M} \tag{20}$$

The tensor $\mathbf{M}$ tensor can be reduced to

$$M_{ii} = \pi(a_1 a_2 a_3)^2 \tau_o (A_j + A_k) \quad (i \neq j \neq k) \tag{21a}$$

$$A_i = \int_0^\infty \frac{dt}{\Delta(a_i^2 + t^2)}, \quad \Delta^2 = (a_1^2 + t^2)(a_2^2 + t^2)(a_3^2 + t^2) \tag{21b}$$

Summation convention has been suspended in Eq. (21). Our choice of coordinate system coincident with the principal axes of the ellipsoid and the ellipsoidal symmetry in general restrict nontrivial terms to only the diagonal ones. The A_i integrals are similar to the ones that appear in Eshelby's [1] classical work (but ***not*** the same). They are easily cast in terms of elliptical integrals

$$A_1 = \frac{\sin^2\phi\cos^2\theta F(\theta,\phi) + \cos^2\phi E(\theta,\phi) - (a_3/a_2)\sin\phi\cos\phi}{a_1^3 a_2 \sin^3\phi\cos^2\theta}$$

$$A_2 = \frac{\cos^2\theta F(\theta,\phi) - (a_3/a_2)E(\theta,\phi) + (a_3/a_2)\sin^2\theta\sin\phi\cos\phi}{a_1 a_2^3 \sin^3\phi\cos^2\theta\sin^2\theta} \tag{22}$$

$$A_3 = \frac{E(\theta,\phi)(a_3/a_2)^2 F(\theta,\phi)}{a_1 a_2 a_3^2 \sin^3\phi\sin^2\theta}$$

Here, the following ordering of semi-axes has been assumed, $a_1 > a_2 > a_3$. E and F are the incomplete elliptical integrals of the first and second kind, respectively

$$E(\theta,\phi) = \int_0^\phi \frac{d\varpi}{\sqrt{1 - \sin^2\varpi\sin^2\theta}}$$

$$F(\theta,\phi) = \int_0^\phi \sqrt{1 - \sin^2\varpi\sin^2\theta}\, d\varpi \tag{23}$$

$$\theta = \sin^{-1}\frac{a_1}{a_2}\sqrt{\frac{(a_2^2 - a_3^2)}{(a_1^2 - a_3^2)}}, \quad \phi = \sin^{-1}\frac{a_3}{a_1}$$

Equations (22) are obtained by substituting $t = a_3 \sin\varpi / \sqrt{\sin^2\phi - \sin^2\varpi}$ in Eq. (21).

Obviously, for simpler shapes such as spheres, circular cylinders and spheroids, these elliptical integrals (E and F) degenerate to well-known elementary expressions (e.g., [32]). The final (interior) strains and stresses of the embedded ellipsoidal inclusion can be conveniently expressed as

$$\begin{aligned} \varepsilon &= \mathbf{S}:\varepsilon^{*\mathbf{T}} = \mathbf{S}:\left\{\varepsilon^{*\mathbf{P}} - \frac{2}{V}\mathbf{C}^{-1}\mathbf{M}\right\} \\ \sigma &= \mathbf{C}:(\mathbf{S} - \mathbf{I}):\left\{\varepsilon^{*\mathbf{P}} - \frac{2}{V}\mathbf{C}^{-1}:\mathbf{M}\right\} \end{aligned} \tag{24}$$

Note that we have used the Eshelby tensor ($\mathbf{S}$), not the operator ($\hat{\mathbf{S}}$), since in the first moment approximation, the total eigenstrain is uniform.

4 Higher Order Virial Moments

In the previous section a first order virial moment of the equations of motion led to the approximation of uniform strain in the ellipsoidal particle under the influence of a surface tension and consequently uniform strain in the embedded inclusion. Successively higher order moments of the equations of motion can be taken to obtain more accurate approximations. To obtain a nontrivial set of self-sufficient equations, the desired strain must be a polynomial of degree one $m-1$ where m is the order of the moment, e.g., in the previous section, a first order moment was taken consistent with a uniform strain (zero order polynomial).

The equations of motion (Eq. (15a)) are now multiplied by $x_j x_m$ and integrated over the volume. Analogous to Eq. (16), we arrive at

$$\int_S x_j x_m \sigma_{ik}^{I(-)} dS_k - \int_V (x_m \sigma_{ij}^I + x_j \sigma_{im}^I) dV = 0 \tag{25}$$

Consistent with the second order virial moment, we shall assume the isolated elastic state to be a polynomial of degree $(m-1)$, which in the present case is a linear function of position, i.e.,

$$\sigma_{ij}^I = \sigma_{ij}^2 = \sigma_{ij}^1 + a_{ijk} x_k \tag{26}$$

Here, the coefficients a_{jik} are constants to be determined with the aid of Eq. (25). The first term in Eq. (26), σ_{ij}^1, is determined (as in the preceding section) by means of the first order viral moment equation.

Straightforward manipulation yields

$$\int_V (\sigma_{ij}^I x_m + \sigma_{im}^I x_j) dV = -2(M_{ijm} + M_{imj}) \tag{27}$$

Where we have used (Appendix B)

$$\tau_o \int_S x_j x_k \operatorname{div} \mathbf{n} dS_i = 2M_{ijk} + 2M_{ikj} \tag{28a}$$

$$M_{ijk} = \frac{\tau_o}{2} \int_S (\delta_{ij} - n_i n_j) x_k dS \tag{28b}$$

The reduction to Eq. (28b) is important since it immediately allows us to note that the integrand is manifestly an odd function. The ellipsoidal shape possesses three planes of symmetry and consequently an odd function must integrate out to zero on its surface. Thus, the second order virial moment approximation degenerates to the first moment approximation of the previous section, i.e., $a_{ijk}=0$. This of course implies that the "average" strain computed in the previous section is correct up to (first) linear order. Thus, at least a third order moment (or a quadratic approximation in the strain) is required to introduce a nonuniform strain in the elastic solution of an isolated ellipsoidal particle (and hence embedded inclusion).

In closure of this section, we point out the evident fact that due to the lengthy and tedious expressions involved, implementation is somewhat inconvenient beyond the first order approximation.

5 A Purely Dilating Ellipsoidal Inclusion

For the pure dilatational case ($\varepsilon^{*\mathbf{P}} = \varepsilon^p \mathbf{I}$), exceptionally simple expressions can be derived which we now proceed to outline. We only address the uniform strain approximation (i.e., first order virial moment). Consider Eq. (24) which provides the strain tensor for an embedded inclusion incorporating surface/interface energies via the second order tensor, $\mathbf{M}$. Taking trace on both sides we obtain

$$\operatorname{Tr}(\varepsilon) = \operatorname{Tr}(\mathbf{S}:\varepsilon^{*\mathbf{T}}) = \operatorname{Tr}(\mathbf{S}:\varepsilon^{*\mathbf{P}}) - \operatorname{Tr}\left\{\mathbf{S}:\frac{2}{V}\mathbf{C}^{-1}\mathbf{M}\right\} \tag{29}$$

At this point we appeal to a general result derived by Milgrom-Shtrikman [33] which specifies that the trace of classical Eshelby's tensor is a constant for all shapes, i.e., in other words, the dilatation within an inclusion is independent of shape. Equation (29) then transforms to

$$\operatorname{Tr}(\varepsilon) = \left(\frac{1+\nu}{1-\nu}\right)\varepsilon^p - \operatorname{Tr}\left\{\frac{2}{V}\left[\frac{1}{2\mu}S_{ijpq}M_{pq} + \frac{2\mu - 3K}{18\mu K}S_{ijmm}M_{qq}\right]\right\} \tag{30}$$

This can be further simplified to

$$\operatorname{Tr}(\varepsilon) = \underbrace{\left(\frac{1+\nu}{1-\nu}\right)\varepsilon^p}_{\substack{\text{Classical result:}\\ \text{Milgrom Shtrikman Trace}}} \underbrace{- \frac{2}{V}\left[\frac{1}{2\mu}S_{iipq}M_{pq} + \frac{2\mu - 3K}{18\mu K}\left(\frac{1+\nu}{1-\nu}\right)M_{qq}\right]}_{\text{Shape and size-dependent results}} \tag{31}$$

The first term in Eq. (31) is simply the Milgrom-Shtrikman [33] trace valid for all shapes and as apparent solely a function of Poisson ratio of the matrix. The second term which involves $\mathbf{M}$ can be made more explicit by using Eq. (21a)

$$\operatorname{Tr}(\mathbf{M}) = 2\pi (a_1 a_2 a_3)^2 \tau_o (A_i + A_j + A_k) = \mathcal{A}\tau_o \tag{32}$$

where we have used the definition of surface area ($\mathcal{A}$). Further, decomposing, $\mathbf{M}$ into a trace and trace-free part, i.e., $\mathbf{M} = \frac{1}{3}\operatorname{Tr}(\mathbf{M})\mathbf{I} + \mathbf{M}'$ final result then takes a simpler form

$$\operatorname{Tr}(\varepsilon) = \underbrace{\left(\frac{1+\nu}{1-\nu}\right)\varepsilon^p}_{\text{Classical}} \underbrace{- \frac{2\mathcal{A}\tau_o}{V}\frac{1}{9K}\left(\frac{1+\nu}{1-\nu}\right) - \frac{1}{2\mu}\frac{2}{V}S_{iipq}M'_{pq}}_{\text{Shape and size-dependent results}} \tag{33}$$

$\mathcal{A}$ for a triaxial ellipsoid is usually expressed as

$$\mathcal{A} = 2\pi c^2 + \frac{2\pi b}{\sqrt{a^2 - c^2}}[(a^2 - c^2)E(\theta) + c^2\theta]$$
$$\operatorname{sn}(\theta,k) = \sqrt{1 - \frac{c^2}{a^2}}; \quad k = \frac{a}{b}\sqrt{\frac{b^2 - c^2}{a^2 - c^2}} \tag{34}$$

Here E is the complete elliptical integral of the second kind and inversion of the Jacobi elliptic function sn is required. Volume is simply $(4\pi a_1 a_2 a_3/3)$. The form of our results presented in Eq. (33) is especially convenient in that one can judge the departure form "shape isotropy" (i.e., spherical shape) by the magnitude of $\mathbf{M}'$ (which is identically zero for a sphere). One can verify (once again) that upon substituting $\mathcal{A}/V = 3/R_0$ and $\mathbf{M}'=0$ the spherical inclusion result of Sharma and Ganti [8] is recovered provided one ignores the deformation dependent terms in their expression. The new results for dilatation depart from the classical Milgrom-Shtrikman trace since $\mathcal{A}/V$ and $\mathbf{M}'$ are both shape-dependent. Size-dependency, obviously, also enters through $\mathbf{M}'/V$ and $\mathcal{A}/V$.

6 Numerical Results and Applications to Shape and Size Effects in Band Gap of Quantum Dots

The incorporation of surface size-effects in the inclusion problem extends all previous application areas of Eshelby's inclusion problem to the nanoscale. The present solution, now also allows study of shape effects beyond trivial shapes such as spherical or circular.

In the previous sections, identical material properties were assumed for the matrix and the inclusion. The differing elastic constants can be taken care of easily through Eshelby's equivalent

Table 1 Properties used in numerical calculations

Property	Value
E_g^∞ (eV)	1.94 [36]
a_c+a_v (eV)	8.3 [36]
μ^M (GPa)	67
K^M (GPa)	102
K^H (GPa)	168
μ^H (GPa)	95
τ_o (J/m^2)	1.33

inclusion principle [1,14] when the eigenstrain is uniform (which is the case in our first order moment approximation). The equivalent inclusion principle was extended to the polynomial case by Sendeckyj [34] and thus can in principle be used for higher order moment approximations if desired although we hasten to point out that one must contend with yet more tedious expressions. First some general results are presented for *inclusions* followed by discussion of our results for quantum dots (*inhomogeneities*). The material parameters used for the quantum dot are summarized for convenience in Table 1 (Appendix C) and the physical eigenstrain is identified with the lattice mismatch between the quantum dot and the surrounding matrix. When discussing the applications to quantum dots, the differing properties of the inclusion and the matrix are taken into account using Eshelby's equivalent condition [1,14]

$$\mathbf{C}^{\mathbf{I}}\left[\mathbf{S}:\left(\boldsymbol{\varepsilon}^{\mathbf{f}}+\boldsymbol{\varepsilon}^{\mathbf{p}}-\frac{2}{V}(\mathbf{C}^{\mathbf{I}})^{-1}\mathbf{M}\right)-\left(\boldsymbol{\varepsilon}^{\mathbf{p}}-\frac{2}{V}(\mathbf{C}^{\mathbf{I}})^{-1}\mathbf{M}\right)\right]$$
$$=\mathbf{C}^{\mathbf{M}}\left[\mathbf{S}:\left(\boldsymbol{\varepsilon}^{\mathbf{f}}+\boldsymbol{\varepsilon}^{\mathbf{p}}-\frac{2}{V}(\mathbf{C}^{\mathbf{I}})^{-1}\mathbf{M}\right)-\left(\boldsymbol{\varepsilon}^{\mathbf{f}}+\boldsymbol{\varepsilon}^{\mathbf{p}}-\frac{2}{V}(\mathbf{C}^{\mathbf{I}})^{-1}\mathbf{M}\right)\right] \quad (35)$$

Here superscripts "I" and "M" indicate inhomogeneity and matrix, respectively. Equation (35) allows the determination of the fictitious eigenstrain, $\boldsymbol{\epsilon}^f$, necessary to simulate the perturbation due to differing elastic constants of the matrix and the inhomogeneity.

6.1 General Results. In this section, we assume an inclusion undergoing a dilatational transformation strain. The effect of aspect ratio is studied by for both prolate ($a_1=a_2<a_3$) and oblate ($a_1=a_2>a_3$) geometry. Obviously our results are generic enough to tackle arbitrary ellipsoidal shapes; prolate and oblate cases are illustrated for a clearer insight into the shape effects.

Percentage deviation of the dilation from the spherical shape is shown for both prolate and oblate spheroids in Figs. 4(*a*)–4(*d*).

The results are plotted with respect to spheroid size (a_1) and aspect ratio ($r=a_3/a_1$). As can be appreciated the presence of surface tension renders the solution difficult to normalize hence for a better perspective, results are plotted for three different parametric values of the eigenstrain ε^{P*} (0.02, 0.04, and 0.06). In Figs. 4(*a*) and 4(*b*), three aspect ratios ($a_3/a_1=r=1.5, 2.5, 5$) are shown. Clearly, the effect of shape depends on both size and the magnitude of the physical eigenstrain. Keeping in mind the rather typical value of surface tension we chose (1.33 J/m^2), consider the case of 2% mismatch strain which is about the norm for quantum dots/substrate systems grown using the Stranski-Krastanov method. For this case, even up to 2 nm inclusion size (at an aspect ratio of 5), the departure from the spherical shape (i.e., Milgrom-Shtrikman trace) is no more than 10%. Shape effects, however, get rapidly appreciable below 2 nm reaching almost 30% for an inclusion size of 1 nm. For larger mismatch eigenstrains (4% and 6%, respectively) the shape effects are indeed small barely reaching 6% deviation from the spherical shape at a mismatch of 6% and aspect ratio of 5. Figures 4(*c*) and 4(*d*) yields similar conclusions albeit different magnitudes.

There is a weak dependence of dilation on shape. One though does expect a quite a stronger effect on the principal strains which is illustrated for prolate inclusions in Fig. 5. It can be seen the strains inside the inclusion exhibit a strong nonhydrostatic effect. The equivalence of ε_{11} and ε_{22} is due to the obvious symmetry of the geometry.

6.2 Shape Effects in Quantum Dots. We next employ the present results to study the effect of quantum dot shape on the size-dependent strain and consequently its band gap. Quantum dots (QD) are of immense technological importance and are typi-

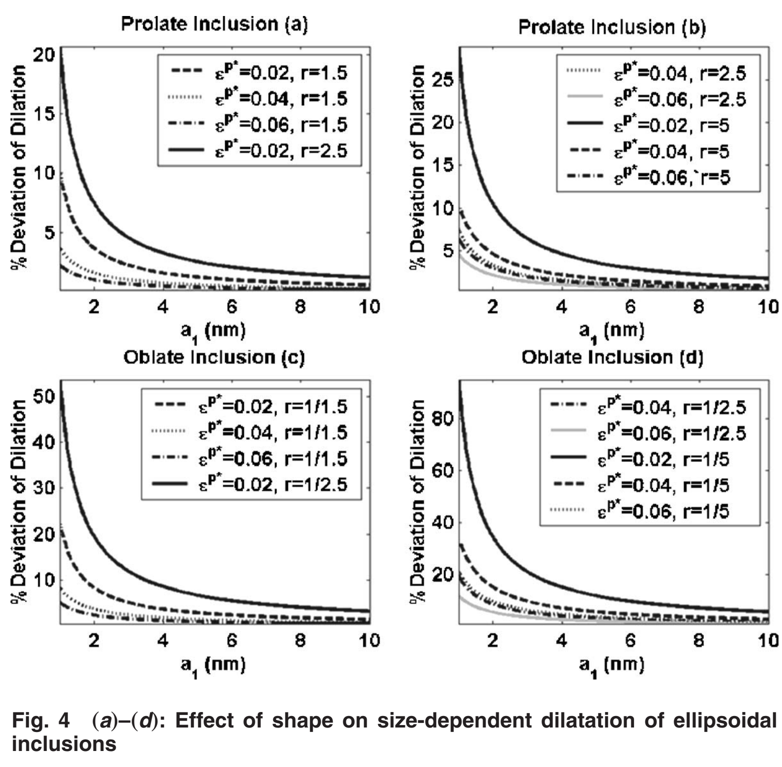


Fig. 4 (*a*)–(*d*): Effect of shape on size-dependent dilatation of ellipsoidal inclusions

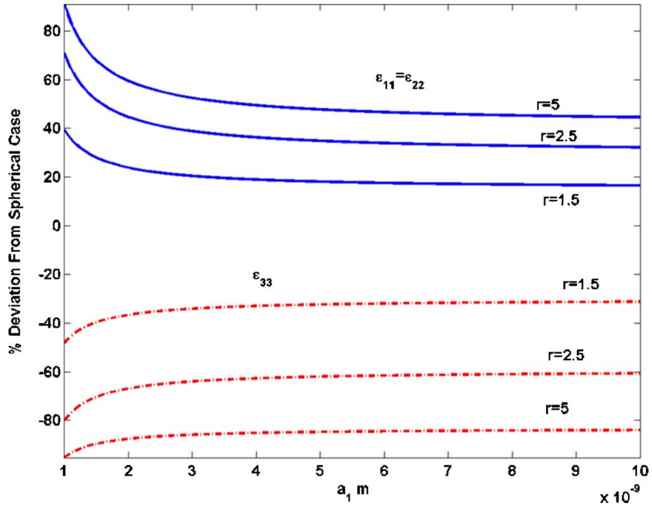


Fig. 5 Effect of shape on size-dependent principal strains of ellipsoidal inclusion, $a_1=a_2<a_3$, $r=a_3/a_1$. (*a*) $r=5$ (*b*) $r=2.5$ (*c*) $r=1.5$.

cally embedded in another semiconductor material with differing elastic constants and lattice parameter. The ensuing elastic relaxation within the QD is well known to impact their opto-electronic properties [35]. Acknowledging that quantum dots are often "fabricated" in the sub 10 nm regime, in our previous work [8], we reported significant shifts in band structure and optoelectronic properties for very small spherical quantum dots when one incorporates size-effects due to surface energies. While in most semiconductors the band structure is sensitive to all strain components, the major influence comes from the dilatation. From a classical point of view (based on classical elasticity theory), quantum dot optoelectronic properties have generally been considered relatively insensitive to shape (e.g., Andreev et al. [36])—although one can argue that this conclusion is a consequence of the several simplifying assumptions that typically made while proceeding to analytically calculate strain-band gap coupling. In light of Figs. 4 and 5 and the work of Sharma and Ganti [8] we can conclude that while size-effects due to surface effects may be significant at very small sizes, shapes effect is of secondary importance. A simple first order calculation can provide a numerical order to the shape effect[1]

$$E^{\infty}+E^{\text{surf}}(R_o,a_3/a_1)+E^{\text{cl}}=E^{\infty}+(a_c+a_v)(\varepsilon_{kk}-\varepsilon_{kk}^{*T}) + O(\textit{nondilatational terms}) \quad (36)$$

Where E^{surf} is the band gap shift due to the size and shape dependent contribution from surface energy induced strains while E^{cl} is the corresponding shift due to classical size-independent and (in the dilatational case) shape independent strain and finally E^{∞} is the unstrained bulk crystal band gap. Here (a_c+a_v) represents the dilatational deformation potential where, as indicated, we have ignored anisotropic terms both in elastic calculations as well as electronic calculations. Note that for the purposes of band structure calculations, the eigenstrain must be subtracted from the compatible strain. We take as an example, the technologically important $In_{32}GaN$ quantum dot system embedded in a GaN matrix. We fix the size of the quantum dot at $a_1=a_2=1$ nm to maximize the shape effect. We then obtain ΔE (net band gap shift due to shape at $a_1=a_2=1$ nm, $r=1$, and $r=5$) of about 52 meV which is well beyond the strict tolerances for many optoelectronic devices [35]. At a more realistic quantum dot size of 5 nm, we obtain an error of 7 meV which is small and most likely in the "noise" regime given the uncertainties in the determination of various other parameters of quantum dots (size, shape, material properties, etc.) that may have greater influence on the band structure than the variation in shape which we discuss.

7 Summary and Conclusions

In summary, we have extended a previous work on incorporation of surface/interface energies in the elastic state of inclusions to the ellipsoidal shape. In addition to the size dependency, the present formulation, also allows the possibility of a limited investigation of shape effects in various physical problems that typically employ the inclusion solution. Exact solution does not appear to be feasible for the problem addressed in this work and hence an approximate solution was constructed using the tensor virial method of moments. Although we employed only first order approximation in our numerical results, higher order approximations can be easily derived if such accuracy is required. In a sense, the present solution now incorporates shape effects into the scaling laws for inclusions that are valid at the nanoscale. In addition, exceptionally simple expressions were derived for the pure dilatational problem which is relevant for several physical applications. The discarding of deformation dependent surface elasticity prohibits use of the present results for calculations of effective properties of composites. The present work shares with its preceding companion article [8] much of the same limitations. For example, we have presented a completely isotropic formulation while interfacial/surface properties can be fairly anisotropic. In addition, we have assumed a perfectly coherent interface. In dealing with nano-inclusions it is important also to consider the degree of coherency, a complete investigation of which is left for future work.

Note added in proof. Our statement that residual stresses do not impact effective properties is only true under specific instances, however our statement regarding Ref. [27] stands. For further clarification on this issue, one may consult Huang and Sun [37]. The first author is grateful to Professor Z. P. Huang for educating him on this.

Acknowledgment

Help with plots from Xinyuan Zhang is gratefully acknowledged.

Appendix A: Eshelby's Classical Tensor

In classical isotropic elasticity, the strain field for the inclusion problem is given by [1,14]

$$\varepsilon_{ij}(x)=\frac{1}{8\pi(1-v)}[\Psi_{kl,klij}-2v\Phi_{kk,ij}-2(1-v)(\Phi_{ik,kj}+\Phi_{jk,ki})] \quad (A1)$$

Where, ψ and Φ are bi-harmonic and harmonic potentials of the inclusion shape (Ω). They are given as

$$\Psi_{ij}(\mathbf{x})=\varepsilon_{ij}^{*}\int_{\Omega}|\mathbf{x}-\mathbf{y}|d^3\mathbf{y} \quad (A2)$$

$$\Phi_{ij}(\mathbf{x})=\varepsilon_{ij}^{*}\int_{\Omega}\frac{1}{|\mathbf{x}-\mathbf{y}|}d^3\mathbf{y} \quad (A3)$$

Equation (A1) can then be cast into the more familiar expression

$$\boldsymbol{\varepsilon}(\mathbf{x})=\mathbf{S}(\mathbf{x}):\boldsymbol{\varepsilon}^{*} \quad \mathbf{x}\in\Omega$$
$$\boldsymbol{\varepsilon}(\mathbf{x})=\mathbf{D}(\mathbf{x}):\boldsymbol{\varepsilon}^{*} \quad \mathbf{x}\notin\Omega \quad (A4)$$

Mura's book [14] contains detailed listing of **S** and **D** tensor for various inclusion shapes (spheres, cylinders, ellipsoids, and cuboids). When the eigenstrain is nonuniform, it must be brought

[1]A more rigorous calculation say using the multi-band k.p. method or tight binding approach is required for accurate electronic band structure calculation. For our purposes, the approximation in Eq. (36) is sufficient to illustrate our point.

into the integral in which case, Eshelby's tensor is an integral operator.

Appendix B: Some Identities Related to Ellipsoidal Surface Integrals

In this Appendix, we derive some identities related to various order **M** tensor. We simply illustrate the method by giving a derivation of M_{ijkl} which is not available in the literature [30,31].

Consider the following lemma [30,31] for an arbitrary tensor field **A**:

$$\int_S \frac{\partial \mathbf{A}}{\partial x_l} dS_i = \int_S \frac{\partial \mathbf{A}}{\partial x_i} dS_l \tag{B1}$$

In Eq. (B1) we set $\mathbf{A}=n_l x_j x_k x_m$. The left side of Eq. (B1) can then be expanded to

$$\int_S (\mathrm{div}\ n) x_j x_k x_m + (n_j x_k x_m + n_k x_j x_m + n_m x_j x_k) dS_i \tag{B2}$$

The right side of Eq. (B1) is written as

$$\int_S (\delta_{ij} x_k x_m + \delta_{ik} x_j x_m + \delta_{im} x_j x_k) dS \tag{B3}$$

Equation (B3) can be used to derive the third order approximation if needed.

Appendix C: Material Properties for Numerical Calculations

The numerical values used are listed in Table 1. Further information on approximations of the interface elasticity constants is available from Sharma and Ganti [8]. For the quantum dot problem, all the properties in Table 1 and a dilatational physical eigenstrain corresponding to the lattice mismatch of 2% was used while for the inclusion problem, only μ^M K^M τ_o were used.

References

[1] Eshelby, J. D., 1957, "The Determination of the Elastic Field of an Ellipsoidal Inclusion and Related Problems," Proc. R. Soc. London, Ser. A, **241**, pp. 376–396.

[2] Miller, R. E., and Shenoy, V. B., 2000, "Size-Dependent Elastic Properties of Nanosized Structural Elements," Nanotechnology, **11**(3), pp. 139–147.

[3] Cammarata, R. C., Sieradzki, K., and Spaepen, F., 2000, "Simple Model for Interface Stresses With Application to Misfit Dislocation Generation in Epitaxial Thin Films," J. Appl. Phys., **87**(3), pp. 1227–1234.

[4] Kukta, R., Kouris, D., and Sieradzki, K., 2003, "Adatoms and Their Relation to Surface Stress," J. Mech. Phys. Solids, **51**(7), pp. 1243–1266.

[5] Kouris, D., Peralta, A., and Sieradzki, K., 2000, "Surface Islands and Their Elastic Interaction With Adatoms," Surf. Sci., **445**, pp. 420–429.

[6] Sharma, P., and Ganti, S., 2002, "Interfacial Elasticity Corrections to the Elastic States of Quantum Dots," Phys. Status Solidi B, **234**, pp. R10–12.

[7] Sharma, P., Ganti, S., and Bhate, N., 2003, "The Effect of Surfaces on the Size-Dependent Elastic State of (Nano) Inhomogeneities," Appl. Phys. Lett., **82**, pp. 535–537.

[8] Sharma, P., and Ganti, S., 2004, "Size-Dependent Eshelby's Tensor for Embedded Nano-Inclusions Incorporating Surface/Interface Energies," J. Appl. Mech., **71**, p. 663.

[9] He, L. H., Lim, C. W., and Wu, B. S., 2004, "A Continuum Model for Size-Dependent Deformation of Elastic Films of Nano-Scale Thickness," Int. J. Solids Struct., **41**(3–4), pp. 847–857.

[10] Lim, C. W., and He, L. H., 2004, "Size-Dependent Nonlinear Response of Thin Elastic Films With Nano-Scale Thickness," Int. J. Mech. Sci., **46**(11), pp. 1715–1726.

[11] Duan, H. L., Wang, J., Huang, Z. P., and Karihaloo, B. L., 2005, "Eshelby Formalism for Nano-Inhomogeneities," Proc. R. Soc. London, Ser. A, **461**(2062), pp. 3335–3353.

[12] Lim, C. W., Li, Z. R., and He, L. H., 2005, "Size Dependent, Non-Uniform Elastic Field Inside a Nano-Scale Spherical Inclusion Due to Interface Stress," Int. J. Solids Struct. (to be published).

[13] Dingreville, R., Qu, J., and Cherkaoui, M., 2005, "Surface Free Energy and its Effect on the Elastic Behavior of Nano-Sized Particles, Wires and Films," J. Mech. Phys. Solids, **53**(8), pp. 1827–1854.

[14] Mura, T., 1987, *Micromechanics of Defects in Solids*, Martinus Nijhoff, Hague, Netherlands.

[15] Nemat-Nasser, S., and Hori, M., 1999, *Micromechanics: Overall Properties of Heterogeneous Solids*, Elsevier Science, New York.

[16] Markov, K., and Preziosi, L., 2000, *Heterogeneous Media: Micromechanics Modeling Methods and Simulations*, Birkhauser Verlag, Switzerland.

[17] Weng, G. J., Taya, M., and Abe, H., eds, 1990, *Micromechanics and Inhomogeneity: The Toshio Mura Anniversary Volume*, Springer-Verlag, Berlin.

[18] Bilby, B. A., Miller, K. J., and Willis, J. R., 1984, IUTAM/IFC/ICM Symposium on Fundamentals of Deformation and Fracture (Sheffield, England, 2–5 April 1984, Eshelby Memorial Symposium), Cambridge University Press, Cambridge, UK.

[19] Mura, T., Shodja, H. M., and Hirose, Y., 1996, "Inclusion Problems," Appl. Mech. Rev., **49**(10), pp. S118–S127.

[20] Mura, T., 2000, "Some New Problems in the Micromechanics," Mater. Sci. Eng., A, **285**(1), pp. 224–228(5).

[21] Gurtin, M. E., and Murdoch, A. I., 1975, "A Continuum Theory of Elastic Material Surfaces," Arch. Ration. Mech. Anal., **59**, pp. 291–323.

[22] Murdoch, A. I., 1976, "The Propagation of Surface Waves in Bodies With Material Boundaries," J. Mech. Phys. Solids, **24**, pp. 137–146.

[23] Gurtin, M. E., Weissmuller, J., and Larche, F., 1998, "The General Theory of Curved Deformable Interfaces in Solids at Equilibrium," Philos. Mag. A, **78**, pp. 1093–1109.

[24] Ibach, H., 1997, "The Role of Surface Stress in Reconstruction, Epitaxial Growth and Stabilization of Mesoscopic Structures," Surf. Sci. Rep., **29**(5–6), pp. 193–263.

[25] Kleinert, H., 1989, *Gauge Field in Condensed Matter*, World Scientific, Singapore, Vol. 2.

[26] Chandrasekhar, S., 1969, *Ellipsoidal Figures of Equilibrium*, Yale University Press, New Haven.

[27] Yang, F., 2004, "Size-Dependent Effective Modulus of Elastic Composite Materials: Spherical Nanocavities at Dilute Concentrations," J. Appl. Phys., **95**(7), pp. 3516–3520.

[28] Cahn, J. W., and Larche, F., 1982, "Surface Stress and the Chemical Equilibrium of Small Crystals. II. Solid Particles Embedded in a Solid Matrix," Acta Metall., **30**(1) pp. 51–56.

[29] Chandrasekhar, S., 1965, "The Stability," Proc. R. Soc. London, **A286**, p. 1.

[30] Rosenkilde, C. E., 1967a, "Surface-Energy Tensors," J. Math. Phys., **8**(1), pp. 84–88.

[31] Rosenkilde, C. E., 1967b, "Surface-Energy Tensors," J. Math. Phys., **8**(1), pp. 88–97.

[32] MacMillan, W. D., 1958, *The Theory of the Potential*, Dover, New York.

[33] Milgrom, M., and Shtrikman, S., 1992, "The Energy of Inclusions in Linear Media Exact Shape-Independent Relations," J. Mech. Phys. Solids, **40**, pp. 927–937.

[34] Sendeckyj, G. P., 1967, "Ellipsoidal Inhomogeneity Problem," Ph.D. Dissertation, Northwestern University, Evanston, IL.

[35] Singh, J., 1992, *Physics of Semiconductors and Their Heterostructures*, McGraw-Hill, New York.

[36] Andreev, A. D., Downes, J. R., Faux, D. A., and O'Reilly, E. P., 1999, "Strain Distributions in Quantum Dots of Arbitrary Shape," J. Appl. Phys., **86**(1), pp. 297–305.

[37] Huang, Z. P., and Sun, L., 2006, "Size-Dependent Effective Properties of a Heterogeneous Material With Interface Energy Effect: From Finite Deformation Theory to Infinitesimal Strain Analysis," Acta Mech., **190**, pp. 151–163.

Ramon Quintanilla
Matemàtica Aplicada 2,
Universitat Politècnica de Catalunya,
Colon 11, 09222 Terrassa,
Barcelona, Spain
e-mail: ramon.quintanilla@upc.edu

Giuseppe Saccomandi
Dipartimento di Ingegneria dell'Innovazione,
Sezione di Ingegneria Industriale,
Università di Lecce,
73100 Lecce,
Italy
e-mail: giuseppe.saccomandi@unile.it

The Importance of the Compatibility of Nonlinear Constitutive Theories With Their Linear Counterparts

We show, by considering a special class of nonlinear viscoelastic materials, that consistency of a mechanical model with classical linear viscoelasticity, may be a fundamental condition to ensure a mathematical and physical well-posedness behavior. To illustrate our arguments we use a rectilinear class of shear motions that we investigate in the static and quasistatic case in the framework of a simple boundary value problem and the classical recovery phenomenon. [DOI: 10.1115/1.2338053]

1 Introduction and Basic Equations

The aim of this paper is to point out the importance of the compatibility of nonlinear theories with the classical corresponding linear theories. For the sake of simplicity, we shall restrict our attention to nonlinear incompressible elasticity and nonlinear viscoelasticity of differential type, but it is not hard to extend our arguments to other (including more complex) theories.

As is conventional in continuum mechanics, the motion of a body is described by a relation $\mathbf{x}=\mathbf{x}(\mathbf{X},t)$, where $\mathbf{x}$ denotes the current coordinates of a point occupied by the particle of coordinates $\mathbf{X}$ in the reference configuration at the time t. The deformation gradient $\mathbf{F}(\mathbf{X},t)$ and the spatial velocity gradient $\mathbf{L}(\mathbf{X},t)$ associated with the motion are defined as $\mathbf{F}:=\partial\mathbf{x}/\partial\mathbf{X}$ and $\mathbf{L}:=\text{grad}\,\mathbf{v}$, respectively ($\mathbf{v}=\partial\mathbf{x}/\partial t$). The other geometrical and kinematical quantities of interest are the left Cauchy-Green strain tensor $\mathbf{B}:=\mathbf{F}\mathbf{F}^{\mathrm{T}}$, and the stretching tensor $\mathbf{D}=\frac{1}{2}(\mathbf{L}+\mathbf{L}^{\mathrm{T}})$. For incompressible materials we have that $\det\mathbf{F}=1$ and $\operatorname{tr}\mathbf{D}=0$, therefore the Cauchy stress, $\mathbf{T}$, can be determined only within an arbitrary spherical part.

In the case of a viscoelastic material of grade 1, by definition it must be [1] $\mathbf{T}=\mathcal{G}(\mathbf{D},\mathbf{B})$. The more general isotropic and incompressible material of grade 1 *linear* in $\mathbf{D}$ is characterized by the constitutive equation

$$\mathbf{T}=-p\mathbf{I}+\alpha_1\mathbf{B}+\alpha_2\mathbf{B}^{-1}+\alpha_3\mathbf{D}+\alpha_4(\mathbf{D}\mathbf{B}+\mathbf{B}\mathbf{D})+\alpha_5(\mathbf{D}\mathbf{B}^2+\mathbf{B}^2\mathbf{D}) \tag{1}$$

where the response functions α_1, α_2, α_3, α_4, α_5 depend are functions of the principal invariants

$$I_1=\operatorname{tr}\mathbf{B},\quad I_2=\operatorname{tr}\mathbf{B}^{-1}$$

and p is the indeterminate pressure introduced by the isochoric constraint $I_3\equiv\det\mathbf{B}=1$.

In the purely hyperelastic case we have $\alpha_3=\alpha_4=\alpha_5=0$ and there exists a scalar function (the strain-energy) $W=W(I_1,I_2)$ such that $\alpha_1=2W_1$, $\alpha_2=-2W_2$ where $W_i=\partial W/\partial I_i$. The constitutive equation (1) contains the natural and direct generalization of the classical linear Kelvin-Voigt viscoelasticity to a nonlinear setting as it has been proposed previously by several authors [2].

At the very least, to model *real* material behavior, the response functions α_i should be compatible with fairly general empirical mechanical response results obtained from carefully controlled large deformation tests on isotropic materials of special kinds. This raises the related question of what reasonable restrictions on the form of the response functions must be imposed, a delicate and central matter sometimes refereed to as Truesdell's *Hauptproblem* in the framework of nonlinear elasticity. This problem is intimately related with the interplay between stability, thermodynamics, and the mathematical conditions that guarantee existence and uniqueness of solutions in nonlinear elasticity.

In nonlinear elasticity experimental data appear to support the following *empirical inequalities*:

$$\alpha_1>0,\quad \alpha_2\leqslant 0 \tag{2}$$

In fact, a variety of tests by Rivlin and Saunders, Treloar, and others on incompressible rubber-like materials support these inequalities [2].

The inequalities (2) are not a definitive solution to the Truesdell's Hauptproblem. It is only known that empirical inequalities are useful in some physical situations, because they reflect the intuitive and expected behavior of materials. Other requirements on the response coefficients may come from the questions of existence and uniqueness of the solution. It should be recognized, however, that the classical Hadamard notion [3], commonly used in linear theories, of well-posedness is too restrictive and in the nonlinear theories of materials it is very hard to justify fundamental restrictions on the constitutive functions besides those arising from material symmetry and frame indifference. On the other hand numerical methods (as for example the one used in commercial FEM softwares) require some regularity and stability results to ensure stability, robustness, and reliability.

It is well known that for a linear, incompressible, isotropic material there is only one independent elastic constant: the *infinitesimal shear modulus* denoted μ. The connection between the derivatives of the strain energy density $W(I_1,I_2)$ and this constant is given by the relation

$$W_1+W_2=\mu/2 \tag{3}$$

where the partial derivatives with respect to the invariants are evaluated for $I_1=I_2=3$ (i.e., in the unstressed configuration).

In this paper we say that a given strain energy W^* is *compatible* with the linear theory if it satisfies the condition

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received October 13, 2005; final manuscript received May 24, 2006. Review conducted by E. M. Arruda. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California–Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME Journal of Applied Mechanics.

$$0 < (W_1^* + W_2^*)\big|_{I_1=I_2=3} < \infty \tag{4}$$

so that the corresponding infinitesimal shear modulus is finite, nonzero, and positive.

Compatibility of nonlinear elasticity is (in part) contained in the first of the empirical inequalities (2) that requires $W_1 > 0$ for all deformations, but it is not common to find this requirement clearly stated in the literature. For example, this is not included in the list of desiderata in the book of Antman [4] (Chap. XIII, Sec. 2). The reason for this situation is that in many of the usual requirements the compatibility with the linear theory is implicit. To the best of our knowledge, one of the few authors in the framework of the nonlinear elasticity who is usually careful to ensure compatibility with the linear limit of the models he introduces is Ogden.

The aim of the present note is to show by an explicit and simple example the consequences of using models that contradict (4) and to extend in a suitable way the definition of compatibility with the linear theory to the framework of a nonlinear viscoelastic theory. It is possible that our discussion will be considered pleonastic by readers with a sophisticated background, but recently in the biomechanical literature several models not compatible with the linear theory have been proposed. This is the case of the constitutive model for arterial wall mechanics given by Takamizawa and Hayashi [5], where instead of $\mathbf{B}$ as a measure of strain, the authors use the Green-Lagrange strain tensor $\mathbf{E}=(\mathbf{C}-\mathbf{I})/2$ where $\mathbf{C} = \mathbf{F}^{\mathrm{T}}\mathbf{F}$ and the strain energy is given as

$$W = -c \ln(1-Q) \tag{5}$$

with

$$Q = \tfrac{1}{2}c_1E_{RR}^2 + \tfrac{1}{2}c_2E_{\Theta\Theta}^2 + \tfrac{1}{2}c_3E_{ZZ}^2 + c_4E_{RR}E_{\Theta\Theta} + c_5E_{\Theta\Theta}E_{ZZ} + c_6E_{ZZ}E_{RR} \tag{6}$$

Here c is a material parameter with the dimension of stress, c_i ($i=1,\ldots,6$) are nondimensional material parameters and E_{RR}, $E_{\Theta\Theta}$, E_{ZZ} are the components of $\mathbf{E}$ in the radial, circumferential, and axial directions respectively. In the isotropic limit Horgan and Saccomandi [6] have shown that this model is equivalent to

$$W^{\mathrm{TH}} = -c \ln\left[1 - \frac{(I_1-3)^2}{J_m^2}\right] \tag{7}$$

where J_m is the limiting chain parameter. It is very simple to check that for (7) we have

$$(W_1^{\mathrm{TH}} + W_2^{\mathrm{TH}})\big|_{I_1=I_2=3} = 0$$

i.e., the infinitesimal shear modulus associated with (7) is null.

It seems that although the check of the compatibility with the linear theory is not formalized in a explicit way by some authors, it is clearly done in several papers. In a recent study by Gasser, Ogden, and Holzapfel [7] the authors derive an anisotropic model, always to be applied in arterial wall mechanics, able to take into account the distributed collagen fiber orientations. In the case of an isotropic distribution it has been shown that the general anisotropic model reduces to

$$W^{\mathrm{GOH}} = \frac{k_1}{2k_2}\left\{\exp\left[\frac{k_2}{9}(I_1-3)^2\right] - 1\right\} \tag{8}$$

which is similar to the Fung-Demiray equation but it depends on $(I_1-3)^2$ and not on (I_1-3). Also in this case is very simple to check that for (8) we have

$$(W_1^{\mathrm{GOH}} + W_2^{\mathrm{GOH}})\big|_{I_1=I_2=3} = 0$$

and the infinitesimal shear modulus associated with this model is null. For this reason in Gasser, Ogden and Holzapfel [7] a classical neo-Hookean term that ensures the compatibility with linear elasticity is added to the strain energy (8).

Similar problems may be found also in the case of viscoelastic models. For example, to describe the constitutive law for ligaments and tendons the following nonlinear viscosity function has been proposed [8]

$$\alpha_3 = \nu(I_1-3)^n \tag{9}$$

where ν is a positive constant and $n>0$. It is clear that in the linear limit the viscosity function (9) is null for $n \neq 0$.

The plan of the paper is the following. In the next section we shall study (1) in the static and quasi-static case when a rectilinear shear motion or deformation is considered. We shall show that in considering such problems is possible to point out in a simple and rigorous way what kind of problems we encounter when the nonlinear theory is not compatible with the linear counterpart. Section 3 is then devoted to specific examples where exact solutions are considered. In the first example, we show that for an elastic material such that the infinitesimal shear modulus is zero a singularity develops in the second derivative of the deformation. In the second example, we show that for elastic materials such that the infinitesimal shear modulus is infinite we have nonexistence results. In the third example, we show that for viscoelastic materials with a zero viscosity parameter in the linear limit the classical recovery problem is not well posed. We make concluding remarks in Sec. 4; in particular, we recall past results and comment them in the light of our present finding.

2 Rectilinear Shear Motions

In the following, we shall consider the propagation of a single *shear wave*

$$x_1 = X_1, \quad x_2 = X_2, \quad x_3 = X_3 + f(X_1,t) \tag{10}$$

or the static version of (10), i.e., a *rectilinear shear* deformation

$$x_1 = X_1, \quad x_2 = X_2, \quad x_3 = X_3 + f(X_1) \tag{11}$$

The class of motions (10) and the corresponding class of static deformations (11) are not *universal*, in the sense that they are not the *same* solution of the balance equation in absence of body forces for all the materials in the constitutive class (1). On the other hand, it is remarkable to note that for every material in the constitutive class (1), under reasonable and general mathematical assumptions on the response functions, we expect that there is at least one choice of f such that the balance equations in absence of body forces are satisfied.

Only when we consider simple shear, i.e., $f=KX_1$ where K is an arbitrary constant, we obtain an universal deformation and when $f=K(t)X_1$ where $K(t)$ is an arbitrary function of time, we obtain an universal motion but in the quasistatic approximation of the equations of motion (i.e., neglecting the inertia terms).

2.1 Static Deformation. When we consider (11) and we introduce for the amount of shear the classical notation

$$k = \frac{df}{dX_1} \tag{12}$$

the balance equations, div $\mathbf{T}=0$, noting that in this case $\mathbf{D}\equiv\mathbf{0}$, reduce to the following system of three scalar equations:

$$\frac{\partial p}{\partial x_1} = \frac{\partial}{\partial x_1}[2W_1 + 2(1+k^2)W_2], \quad \frac{\partial p}{\partial x_2} = 0,$$
$$\frac{\partial p}{\partial x_3} = \frac{\partial}{\partial x_1}[2(W_1+W_2)k] \tag{13}$$

From the usual compatibility conditions for the existence of a pressure field, we obtain that it must be

$$p(x_1,x_3) = [2W_1 + 2(1+k^2)W_2] + \pi x_3 \tag{14}$$

and therefore from $(13)_3$ the determining equation for the rectilinear shear displacement $f(X_1)$ is given by

$$\frac{d[2(W_1+W_2)k]}{dX_1} = \pi \tag{15}$$

Here π it is the constant gradient of pressure in the x_3-direction.

Usually in (15), the quantity

$$M(k) = 2(W_1 + W_2)$$

denoting the secant modulus of shear or the generalized shear modulus, is assumed to be positive for all $k \geqslant 0$ in view of (2). Here, we are interested in replacing this last condition by the weaker requirement $M(k) > 0$ for all $k > 0$ and to examine some of the consequences of this assumption.

To illustrate our ideas, we shall consider a specific boundary value problem (BVP) with Dirichlet boundary conditions. To this end let us consider a slab bonded at both $X_1=0$ and $X_1=H$ to rigid surfaces and with infinite dimensions along the directions X_2 and X_3. We suppose that the deformation of the slab is due to the existence of a pressure gradient generated by applied traction at $X_3=\pm\infty$ and for this reason π must be different from zero. In this framework, the appropriate boundary conditions are

$$f(0) = 0, \quad f(H) = 0 \tag{16}$$

Let $X=X_1/H$, $\bar{f}=f/H$, $\bar{W}=W/\tilde{\mu}$, and $\bar{k}=(d\bar{f}/dX)$. Integration of (15) enables the BVP to be rewritten as

$$\bar{M}(\bar{k})\frac{d\bar{f}}{dX} = \frac{\pi H}{\tilde{\mu}}(X + C_1)$$

$$\bar{f}(0) = 0, \quad \bar{f}(1) = 0 \tag{17}$$

In the above equations $\tilde{\mu}$ is a constant with the dimensions of the shear modulus. When the infinitesimal shear modulus is defined is obvious to consider $\tilde{\mu}\equiv\mu$ but here we are interested in situations where this may not be the case (for the sake of simplicity of notation, we drop the bar henceforth).

The two point BVP (17) was introduced by Zhang and Rajagopal [9] and it is quite interesting because it enjoys a special invariance, as was noted by Pucci and Saccomandi [10]. Indeed, under our hypothesis we have that $df(X)/dX$ is zero in $(0,1)$ only for $X=-C_1$. If we are searching for a continuous solution of our BVP this means that $-C_1 \in (0,1)$. We recall that $M(k)$ is an even function around $X=-C_1$ and by squaring (17) we see that $(df(X)/dX)^2$ is also an even function. This means that $df(X)/dX$ is odd and $f(X)$ must be even. Therefore, $C_1=-1/2$.

The possibility to fix a priori C_1 allows the original BVP to be translated into a standard initial value problem (IVP). To this end, let us consider, the equation given by

$$[M(k)k]_{X=0} = -\frac{\pi H}{2\mu} \tag{18}$$

and let us cast this equation in terms of the unknown $(df/dX)|_{x=0}$. When at least a solution, say χ, of (18) exists then by solving the IVP

$$\left[M + 2\frac{dM}{dk}\left(\frac{df}{dX}\right)^2\right]\frac{d^2f}{dX^2} = \frac{\pi H}{\mu}$$

$$f(0) = 0, \qquad \left.\left(\frac{df}{dX}\right)\right|_{X=0} = \chi \tag{19}$$

for all points in the range $(0,1/2)$ we obtain the solution of the BVP (17). (The solution in the range $(1/2,1)$ is then obtained by symmetry.) This happenstance means not only that we have more powerful numerical methods at our disposal for solving (17) but also that when it is possible to rewrite Eq. (19) in *normal form* we may use the standard theorem of Cauchy to prove uniqueness and existence of our original BVP.

To summarize, to have existence and uniqueness of the IVP (19) and therefore of the BVP (17) when $M(k)$ is a continuous function of its argument k we must have that:

- the function $M(k)k$ is monotone (to ensure the uniqueness of χ),
- the $\lim_{k\to\pm\infty} M(k)k=\infty$ (to ensure the existence of χ), and
- the function $[M+2(dM/dk)k^2]^{-1}$ satisfy the usual requirements of the classical Cauchy theorem (continuity to ensure existence and Lipschitzianity to ensure uniqueness).

It is clear that if $M(0)=0$ when $k=0$ we have that the above last condition is contradicted and indeed Eq. $(19)_1$ cannot be recast in normal form. This is exactly what happens for $X=1/2$ if the model we are considering it is not well defined in the linear limit. On the other hand, if we have $\lim_{k\to 0} M(k)=\infty$ also in this case Eq. $(19)_1$ is problematic.

The deformation we are considering tell us that when the solution of a BVP contains a neighborhood in which shear vanishes, because the linearized limit applies, difficulties can arise when the linearized limit is not physically meaningful.

We emphasize that here we are considering a BVP which for the invariance property is easily handled, and in many situations we are able to find exact solutions to (17). When we are considering more complex BVP it is necessary to consider numerical methods such as the shooting method or approximate methods such as the Picard's iterations [11]. In this case we have continuous dependence on initial conditions and parameters are forthcoming. This kind of result, in the standard formulation, needs not only that the function $[M+2(dM/dk)k^2]^{-1}$ is continuous and satisfies the Lipschitz condition in the appropriate domain, but that this functions is bounded for all values of its argument.

2.2 Quasistatic Motion. For the case of nonlinear viscoelasticity we have to consider that the motion (10) and the amount of shear are therefore also a function of time, i.e., $k=k(X_1,t)$. In this case

$$[D_{ij}] = \frac{1}{2}\begin{bmatrix} 0 & 0 & k_t \\ 0 & 0 & 0 \\ k_t & 0 & 0 \end{bmatrix} \tag{20}$$

and the equation of motion, $\operatorname{div}\mathbf{T}=\rho(\partial^2\mathbf{x}/\partial t^2)$, reduces to the three scalar equations

$$p_{x_1} = \{\alpha_1 + \alpha_2 + \alpha_2 k^2 + [(\alpha_4 + 2\alpha_5)k + \alpha_5 k^3]k_t\}_{x_1}$$

$$p_{x_2} = 0 \tag{21}$$

and

$$\rho f_{tt} = -p_{x_3} + \frac{dT_{13}}{dx_1} \tag{22}$$

where the subscripts signify partial derivatives. Then, by a simple computation, we check that

$$T_{13} = (\alpha_1 - \alpha_2)k + \left[(\alpha_3 + \alpha_4 + \alpha_5) + \tfrac{1}{2}(\alpha_4 + 4\alpha_5)k^2 + \tfrac{1}{2}\alpha_5 k^4\right]k_t \tag{23}$$

From the Eqs. (21) and (22) we deduce that

$$p(x_1,x_3,t) = \alpha_1 + \alpha_2 + \alpha_2 k^2 + (\alpha_4 + 2\alpha_5)k_t k + \alpha_5 k_t k^3 + h_1(t)x_3 + h_0(t) \tag{24}$$

where $h_1(t)$ and $h_2(t)$ are arbitrary functions of time.

From the right hand side (rhs) of (22) it is clear that the requirement that a model must be compatible with the linear theory of viscoelasticity implies $(\alpha_1-\alpha_2)|_{k=0}\neq 0$, as in the previous section, *and* that the infinitesimal viscosity defined as $(\alpha_3+\alpha_4+\alpha_5)_{k=0}$ must be such that

$$0 < (\alpha_3 + \alpha_4 + \alpha_5)_{k=0} < \infty \tag{25}$$

To clarify the importance of the requirement (25) we shall investigate the usual *quasi-static* recovery phenomena associated with the class of motion (10) with the inertial terms neglected. In this case, we assume that

$$f(X_1,t) = K(t)X_1 \tag{26}$$

and we investigate the decay process characterized by the decreasing shear displacement $K(t)$ from an arbitrary initial shear K_0 at which load is suddenly removed. What is expected is that starting from $K(0)=K_0$ we obtain $K(t)\to 0$ as $t\to\infty$. Because, in this case the shear stress is (23) the shear displacement in the recovery problem is obtained by solving the following first-order ordinary differential equation

$$(\alpha_1 - \alpha_2)k + \left[(\alpha_3 + \alpha_4 + \alpha_5) + \tfrac{1}{2}(\alpha_4 + 4\alpha_5)k^2 + \tfrac{1}{2}\alpha_5 k^4\right]k_t = 0 \tag{27}$$

noting that from (26) we have $k=K(t)$. This is in separable form and may be reduced to quadrature as

$$\int \frac{(\alpha_3 + \alpha_4 + \alpha_5) + \frac{1}{2}(\alpha_4 + 4\alpha_5)k^2 + \frac{1}{2}\alpha_5 k^4}{(\alpha_1 - \alpha_2)k}dk = -t + \text{constant} \tag{28}$$

In (28) the arbitrary constant in the rhs is used to fix the initial condition and we therefore have to check directly that the asymptotic value for $t\to\infty$, corresponds to our physical expectations for the solution defined in (28). Since $(\alpha_1-\alpha_2)|_{k=0}\neq 0$ and the various constitutive functions are very regular (we may suppose analytical functions) a necessary and sufficient condition for $k(t)$ defined by (28) to go to zero in an infinite time is given by the condition that $\Phi(0)\neq 0$ where the function $\Phi=\Phi(k)$ is defined by

$$\Phi(k) = (\alpha_3 + \alpha_4 + \alpha_5) + \tfrac{1}{2}(\alpha_4 + 4\alpha_5)k^2 + \tfrac{1}{2}\alpha_5 k^4 \tag{29}$$

and clearly this condition requires $(\alpha_3+\alpha_4+\alpha_5)_{k=0}\neq 0$.

When $\Phi(0)\neq 0$ the integral on the left hand side (lhs) of (28) is a *generalized* integral which does not converge and therefore the value $k=0$ will be reached in an *infinite* time as it is desired. If $\Phi(0)=0$, then under the assumption of analyticity of $\Phi(k)$, a standard theorem of basic real analysis ensures that $\Phi(k)=k^n\Psi(k)$ where n is the multiplicity of the root $k=0$ and $\Psi(k)\neq 0$. This means that the integral on the lhs of (29) is no longer a generalized integral and there is a finite value $t=t^*$ such that $k(t^*)=0$, and for $t>t^*$ the behavior of the shear deformation will be in any case *unphysical*.

Another interesting result is

$$\lim_{k\to 0}\Phi(k) = \infty \tag{30}$$

then, from (28), it follows that a singularity for the amount of shear will develop in *finite* time.

3 Examples

3.1 Development of Singularities. The first example involves the strain energy

$$W^{(1)} = \frac{\mu}{4}(I_1 - 3)^2 \tag{31}$$

where μ is a parameter, not to be identified with the infinitesimal shear modulus, which in this case is null. The BVP (17) may be recast in the dimensionless form as

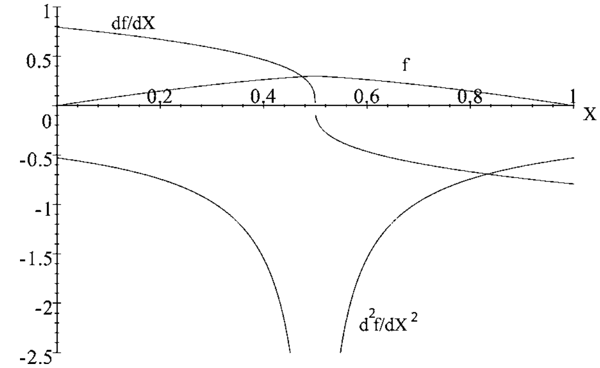


Fig. 1 Solution (35) for typical values of the parameters

$$\left(\frac{df}{dX}\right)^3 = \pm\varepsilon^2\left(X - \frac{1}{2}\right), \quad f(0)=0, \quad f(1)=0 \tag{32}$$

where ε is a number associated with the pressure gradient (we assume $\varepsilon>0$). Now, when we select the—sign on the rhs of (32) we obtain

$$\frac{df}{dX} = \begin{cases} \sqrt[3]{\varepsilon^2\left(\frac{1}{2} - x\right)} & \text{for } 0 \leqslant x \leqslant 1/2 \\ -\sqrt[3]{\varepsilon^2\left(x - \frac{1}{2}\right)} & \text{for } 1/2 < x \leqslant 1 \end{cases} \tag{33}$$

and when we select the + sign the result is

$$\frac{df}{dX} = \begin{cases} -\sqrt[3]{\varepsilon^2\left(\frac{1}{2} - x\right)} & \text{for } 0 \leqslant x \leqslant 1/2 \\ \sqrt[3]{\varepsilon^2\left(x - \frac{1}{2}\right)} & \text{for } 1/2 < x \leqslant 1 \end{cases} \tag{34}$$

In the following we shall consider only the case (34). The exact solution of (34) is given by a direct integration as

$$f(X) = -\frac{3}{4\varepsilon^2}\left\{\sqrt[3]{\left[\varepsilon^2\left(x - \frac{1}{2}\right)\right]^4} - \sqrt[3]{\left(\frac{\varepsilon^2}{2}\right)^4}\right\} \tag{35}$$

On the other hand we remark that

$$\left.\left(\frac{df}{dX}\right)\right|_{X=0} = -\sqrt[3]{\frac{\varepsilon^2}{2}} \tag{36}$$

and therefore the IVP corresponding to the BVP under scrutiny is obtained directly from (32) as

$$\frac{d^2 f}{dX^2} = -\frac{\varepsilon^2}{3(df/dX)^2}, \quad f(0)=0, \quad \left.\left(\frac{df}{dX}\right)\right|_{X=0} = -\sqrt[3]{\frac{\varepsilon^2}{2}} \tag{37}$$

In Fig. 1 we have plotted this exact solution for $\varepsilon=1$. The solution is symmetric with respect $X=1/2$, the first derivative of the solution develops in $X=1/2$ a vertical tangent and for this reason in this point, the second derivative blows up. This is a major problem when we try to solve the BVP numerically, both transforming it to an IVP or using numerical methods suitable for the two point BVPs as shooting methods.

Moreover, also from the physical point of view in $X=1/2$, we have to record a strong localization of strain that needs some mechanical interpretation that in materials as soft tissues and elastomers seems to be *artificial*. It is clear that the level of regularity of the solutions is in any case a matter that needs some mechanical interpretation, but the singularity here displayed may be very problematic when we cannot locate it a priori as in our special example. A similar problem where it is not possible to locate the singularity a priori (because no invariance results is obtainable) is discussed in Horgan and co-workers [12]. This example also clarifies the problems we have from a mathematical point of view in dealing with models (7) and (8).

3.2 Nonexistence. Let us now consider the strain energy

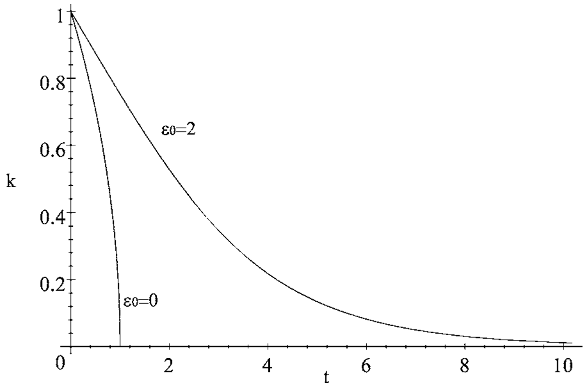


Fig. 2 Recovery shear as in (44) in the case $\epsilon_0=0$ and $\epsilon_0 \neq 0$

$$W^{(2)} = \mu n(I_1 - 3)^{1/n} \tag{38}$$

where $n>1$ and again μ is a parameter not to be confused with the infinitesimal shear modulus since

$$M(k) = \mu k^{2(1-n)/n} \tag{39}$$

and therefore the infinite shear modulus is infinite. The BVP (17) may be recast in dimensionless from as

$$\left(\frac{df}{dX}\right)^{(2-n)/n} = \tilde{\varepsilon}\left(X - \frac{1}{2}\right), \quad f(0)=0, \quad f(1)=0 \tag{40}$$

where $\tilde{\varepsilon}$ is a constant related to the pressure gradient. It is clear that in this case the solution of (40) is very problematic. For $n=2$ clearly, at first sight, no solution exists. For $n=4$ we have that also in this case we cannot have existence because form

$$\left(\frac{df}{dX}\right)^{-1/2} = \tilde{\varepsilon}\left(X - \frac{1}{2}\right) \tag{41}$$

it is clear that for $X>1/2$ we have this equation has no meaning in the real domain. When we consider the second order equation corresponding to (41)

$$-\frac{1}{2(df/dX)^{3/2}}\frac{d^2f}{dX^2} = \tilde{\varepsilon}$$

an equation that cannot be recast in normal form. Indeed, if we consider

$$\frac{d^2f}{dX^2} = -2\tilde{\varepsilon}(df/dX)^{3/2}$$

this admits the spurious solution $f(X)\equiv 0$.

3.3 A Recovery Problem. An example suitable for illustrating the quasistatic motion result is obtained considering a viscoelastic media with linear elastic generalized shear modulus, say μ, with the remaining response coefficients chosen so that $\alpha_4=\alpha_5=0$ and

$$\alpha_3 = \nu_0 + \nu_1 k^2 \tag{42}$$

In this case we have that (28) is given by (here $\varepsilon_i=\nu_i/\mu$, for $i=0,1$)

$$\int \frac{\varepsilon_0 + \varepsilon_1 k^2}{k} dk = -t + \text{constant} \tag{43}$$

and by considering $\bar{k}=k/K_0$ (and dropping the bar) we obtain

$$k^2 + \frac{\varepsilon_0}{K_0\varepsilon_1}\ln k = -\frac{t}{K_0\varepsilon_1} + 1 \tag{44}$$

In Fig. 2 we show the shear functions defined in (44) by considering the case $\varepsilon_0=0$ and the case $\varepsilon_0=2$ (here $K_0\varepsilon_1=1$). When there is no compatibility with the linear theory we see that we have recovery in a finite time. In our example this is not a truly dramatic behavior because the amount of shear does not become negative or does not start to increase again after the recovery time, but in any case it is not what we expect from a model of differential type. More exotic behaviors may be obtained considering when

$$\alpha_3 = \nu k^{2n} \tag{45}$$

with $0<n<1$. In this case

$$k(t) = (K_0^n - 2n\varepsilon t)^{1/2n} \tag{46}$$

with $\varepsilon=\mu/\nu$. It is clear from (46) that when, for example, $n=1/2$ the shear starts to increase again after the recovery time, and for $n=2/3$ it becomes negative after the recovery time.

Now let us consider

$$\alpha_3 = \frac{\nu_0}{k} \tag{47}$$

considering to the case (30). Then,

$$\int \frac{\nu_0}{\mu k^2} dk = t + \text{constant} \tag{48}$$

and by considering $\varepsilon=\mu t/(\nu_0 K_0)$, we obtain

$$\frac{1}{k} = 1 - \varepsilon t \tag{49}$$

and as it easy to check that this leads to an *unacceptable* singularity for finite time.

4 Concluding Remarks

We have shown for the model (1) the importance of compatibility with classical linear elasticity and viscoelasticity in the sense explained in the preceding sections. For the sake of simplicity we have chosen to consider an elementary framework to present our remarks, but it is clear that these considerations, may be extended to more general models and more complex deformations and motions.

For example if we consider a generalized neo-Hookean elastic material for which $W=W(I_1)$ and the antiplane shear deformations

$$x_1 = X_1, \quad x_2 = X_2, \quad x_3 = X_3 + u(X_1, X_2) \tag{50}$$

it is well known that in this case the determining equation for the shear displacement $u(X_1,X_2)$ is the quasi-linear partial differential equation

$$[M(k)u_{,\beta}]_{,\beta} = \pi \tag{51}$$

where π is a constant and

$$M(k) = 2W_1 = \left.\frac{\partial W}{\partial I_1}\right|_{I_1=3+k^2} \tag{52}$$

where

$$k^2 = |\nabla u|^2 \tag{53}$$

Such equations have been studied by Serrin [13]. Now if $M(k)>0$ for all $k>0$, we have that equation (51) is elliptic, but to ensure existence and uniqueness of the solution we need uniform ellipticity and this is possible only if $M(k)>0$ for all $k\geq 0$.

On the other hand, if we consider the dynamical equation (22) when the inertia terms are not neglected we have that the existence and uniqueness theorem has been provided by MacCamy [14]. In this paper it is essential to the result of global existence that the condition $(\alpha_3+\alpha_4+\alpha_5)_{k=0}\neq 0$ is satisfied otherwise it is not possible to ensure global existence of the solution. We note that the McCamy condition seems to be quite sharp because also in recent works on the argument you will find that the compatibility with linear theory cannot be avoided to find a well-posed mathematical problem [15].

As we have already declared, the problem of the level of regularity of the solutions depends on the mathematical setting. For example, when we are considering phase transitions we have to search for solutions in a weaker class than usual. On the other hand, existence of solution in classical nonlinear elasticity for Dirichlet boundary value problems (see for example Ciarlet [16] Sec. 6.4) are derived in the Sobolev space $W^{2,p}$. This means that in this setting the compatibility with the linear elasticity is certainly needed. Our example shows that when this compatibility is missed the second derivative of the solution is not continuous. On the other hand, when we consider existence theorems based on the powerful methods of the calculus of variations the existence of minimizers is established in relative coarse spaces and it is not always a simple matter to know when the minimizers are smooth enough to qualify as classical or weak solutions of the BVP of interest.

In any case compatibility with the corresponding linear theory for a nonlinear model is so important in its mathematical and mechanical implications that we think it is worth checking this property explicitly and directly. At least when we are considering classical solutions as is usual in the framework of many applications where, for example, practical methods of computational mechanics as nonlinear finite element methods based on the tangent modulus idea [17] are used. In this framework the noncompatibility may deliver us unpleasant surprises as becomes clear if we consider the definition of the initial tangent modulus.

Acknowledgment

This work is supported by: Aspectos de estabilidad en termomecanica (BFM2003-00309), GNFM of INDAM and PRIN 2004: Modelli Matematici per la Dinamica del DNA. G.S. acknowledge the hospitality and support of UPC. The authors are grateful for useful comments to Michel Destrade and Ray W. Ogden.

References

[1] Truesdell, C., and Noll, W., 2004, *The Non-Linear Field Theories of Mechanics*, 3rd ed., Springer-Verlag, Berlin.

[2] Beatty, M. F., 1996, "Introduction to Nonlinear Elasticity," in *Nonlinear Effects in Fluids and Solids*, M. M. Carroll and M. A. Hayes, eds., Plenum Press, New York, pp. 16–112.

[3] Weinberger, H. F., 1965, *A First Course in Partial Differential Equations*, Waltham, Blaisdell.

[4] Antman, S. S., 1995, *Nonlinear Problems of Elasticity*, Springer-Verlag, New York.

[5] Takamizawa, K., and Hayashi, K., 1987, "Strain Energy Density Function and Uniform Strain Hypothesis for Arterial Mechanics," J. Biomech., **20**, pp. 7–17.

[6] Horgan, C. O., and Saccomandi, G., 2003, "A Description of Arterial Wall Mechanics Using Limiting Chain Extensibility Constitutive Models," Biomech. Modeling in Mechanobiology, **1**, pp. 251–266.

[7] Gasser, T. C., Ogden, R. W., and Holzapfel, G. A., 2006, "Hyperelastic Modelling of Arterial Layers with Distributed Collagen Fibre Orientations," J. R. Soc., Interface, **3**, pp. 15–35.

[8] Pioletti, D. P., Rakotomanana, L. R., Benvenuti, J.-F., and Leyvraz, P.-F., 1998, "Viscoelastic Constitutive Law in Large Deformations: Application to Human Knee Ligaments and Tendons," J. Biomech., **31**, pp. 753–757.

[9] Zhang, J. P., and Rajagopal, K. R., 1992, "Some Inhomogeneous Motions and Deformations Within the Context of a Non Linear Elastic Solid," Int. J. Eng. Sci., **30**, pp. 919–938.

[10] Pucci, E., and Saccomandi, G., 1999, "Some Remarks on the Gent Model of Rubber Elasticity," *CanCNSM Proceedings*, E. M. Croitoro, ed., University of Victoria Press, Victoria, Canada, pp. 163–172.

[11] Bailey, P. B., Shampine, L. F., and Waltman, P. A., 1968, *Nonlinear Two Point Boundary Value Problems*, Academic Press, New York.

[12] Horgan, C. O., Saccomandi, G., and Sgura, I., 2003, "A Two-Point Boundary Value Problem for the Axial Shear of Hardening Isotropic Incompressible Nonlinearly Elastic Materials," SIAM J. Appl. Math., **65**, pp. 1712–1727.

[13] Serrin, J., 1969, "The Problem of Dirichlet for Quasilinear Elliptic Differential Equations with Many Independent Variables," Philos. Trans. R. Soc. London, Ser. A, **264**, pp. 413–496.

[14] MacCamy, R. C., 1970, "Existence, Uniqueness and Stability of Solutions of the Equation $u_{tt}=\partial/\partial x(\sigma(u_x)+\lambda(u_x)u_{xt})$," Indiana Univ. Math. J., **20**, pp. 231–238.

[15] Georgiev, V., Rubino, B., and Sampalmieri, R., 2005, "Global Existence for Elastic Waves with Memory," Arch. Ration. Mech. Anal., **176**, pp. 303–330.

[16] Ciarlet, P. G., 1988, *Mathematical Elasticity: Vol. 1*, North Holland, Amsterdam.

[17] Nicholson, D. W., 2003, *Finite Element Analysis: Thermomechanics of Solids*, CRC Press, Boca Raton, FL.

Single Degree of Freedom Model for Thermoelastic Damping

Jagannathan Rajagopalan
e-mail: jrajago2@uiuc.edu

M. Taher A. Saif[1]
e-mail: saif@uiuc.edu

Department of Mechanical Science and Engineering,
University of Illinois Urbana-Champaign,
Urbana, IL 61801

Finding the thermoelastic damping in a vibrating body, for the most general case, involves the simultaneous solving of the three equations for displacements and one equation for temperature (called the heat equation). Since these are a set of coupled nonlinear partial differential equations there is considerable difficulty in solving them, especially for finite geometries. This paper presents a single degree of freedom (SDOF) model that explores the possibility of estimating thermoelastic damping in a body, vibrating in a particular mode, using only its geometry and material properties, without solving the heat equation. In doing so, the model incorporates the notion of "modal temperatures," akin to modal displacements and modal frequencies. The procedure for deriving the equations that determine the thermoelastic damping for an arbitrary system, based on the model, is presented. The procedure is implemented for the specific case of a rectangular cantilever beam vibrating in its first mode and the resulting equations solved to obtain the damping behavior. The damping characteristics obtained for the rectangular cantilever beam, using the model, is compared with results previously published in the literature. The results show good qualitative agreement with Zener's well known approximation. The good qualitative agreement between the predictions of the model and Zener's approximation suggests that the model captures the essence of thermoelastic damping in vibrating bodies. The ability of this model to provide a good qualitative picture of thermoelastic damping suggests that other forms of dissipation might also be amenable for description using such simple models. [DOI: 10.1115/1.2338054]

1 Introduction

In most solids, the strain and temperature fields are coupled. When temperature is changed, volume changes, when volume is changed by elastic deformation, temperature changes. The constant that relates the change of length (strain) with the change in temperature of a material is its thermal expansion coefficient α. When a body is elastically deformed (with volume change), thus increasing its potential energy, and is allowed to oscillate freely, the body gradually loses its potential energy and returns to its stable equilibrium even if it does not exchange energy with the environment, for example, by air drag or friction. One fundamental mechanism responsible for this dissipation is known as thermoelastic damping, wherein potential energy is converted to heat. If the body is thermally isolated from its surroundings thermoelastic damping leads to an increase in its temperature.

Thermoelastic damping can be understood from two different, but equivalent standpoints. One method is to view it as a process of dissipation of mechanical energy. In general, the stress field in a vibrating body is nonuniform and hence some regions become hotter relative to others due to thermoelastic coupling. This results in heat flow within the body, if it has finite thermal conductivity, k. Due to this heat flow, the temperature field created by thermoelastic effect in a vibrating body becomes out of phase with the stress field. Thus the temperature induced strain field is out of phase with the stress field. This phase difference between stress and strain fields leads to dissipation of mechanical energy. If k is zero, the stress and temperature fields are always in phase and hence no dissipation takes place. If k is very large, the body remains isothermal and again there is no dissipation. The second way is to visualize thermoelastic damping is in terms of generation of entropy. Due to inhomogeneities in the stress field, local temperature gradients are created in the body. This leads to irreversible heat flow until the temperature, T, becomes uniform throughout the body, i.e., the attainment of thermal equilibrium. Since thermal equilibrium corresponds to the state of maximum entropy, there has to be a net increase in the entropy, S, of the body during this process. This increase in entropy has to come at the cost of the potential (strain)/kinetic energy of the system, since the total energy, U, of the system remains constant. The entropy increase can also be considered as an increase in heat content of the body.

The thermoelastic damping in a vibrating body can be obtained, for the most general case, by simultaneously solving the three equations for displacements and one equation for temperature (the heat equation) which comprise the equations of thermoelasticty [1]. Since these are a set of coupled nonlinear partial differential equations there is considerable difficulty in solving them, especially for finite geometries. Over the years, analytical solutions to thermoelastic damping have been obtained for certain simple geometries. Zener first studied thermoelastic relaxation as a source of damping in mechanical systems using the "standard model" of an anelastic solid [2] and developed a general theory of thermoelastic damping in a series of papers [3–5] in the 1930s. He showed that the damping behavior of transversely vibrating cantilever beam can be well approximated by a single relaxation peak with a characteristic relaxation time τ. He further showed that τ, which gives a measure of the time needed for temperature equalization through diffusion, is proportional to b^2/χ, where b is the thickness of the beam and χ is the thermal diffusivity.

Alblas [6] developed a generalised theory for thermoelastic dissipation in vibrating bodies using the three dimensional thermoelastic equations and derived the solution for the coupled thermoelastic equations in terms of normalized orthogonal eigen functions. In a later publication [7], Alblas generalized the results and obtained explicit expressions for thermoelastic damping in vibrating elastic beams, including the circular rod and the rectangular beam. Chadwick [8] derived the coupled equations govern-

[1]Author to whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received October 24, 2005; final manuscript received April 9, 2006. Review conducted by O. M. 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

ing the thermoelastic behavior of thin plates and beams, and demonstrated that these reduce to the classical equation of motion and of heat conduction in the limit of zero coupling. Lord and Shulman [9] derived a generalized theory of thermoelasticity by considering a modified form of the Fourier heat conduction equation that took into account the time lag needed to establish steady state conduction.

In recent times, there has been renewed interest in thermoelastic damping, especially due to its contribution to energy dissipation in micromechanical resonators. High frequency micromechanical resonators, with potential applications as RF filters [10], charge detectors [11], and sensors [12,13], need to have very little energy dissipation or very high quality factor, Q. Since thermoelastic damping is a fundamental damping mechanism, it imposes an upper limit on the quality factor that can be obtained in any oscillator. Lifshitz et al. [14] evaluated the importance of thermoelastic damping at micro and nano scales and concluded that it remains relevant at these scales. Further, they derived an exact expression for thermoelastic damping in thin rectangular beams which compared favorably with Zener's well known approximation [2]. Photiadis et al. [15] proposed a simple model of thermoelastic dissipation assuming that the energy loss occurred only due to dissipation of the flexural component of motion. Houston et al. [16,17] used this model to predict thermoelastic dissipation in a single-crystal double paddle oscillator and found that the predictions agreed well with experimental observations at high temperatures (above 150 K). Nayfeh et al. [18] derived analytical expressions for quality factors in microplates, using perturbation method, by decoupling the heat equation from the equation of motion. In a recent work, Norris et al. [19] presented a general method of calculating thermoelastic damping in vibrating elastic solids, by treating elasticity as an uncoupled forcing term in the heat equation. Using this method the authors obtained a new equation of motion for flexural vibration of thin plates incorporating a thermoelastic damping term.

As mentioned earlier, the difficulty in solving thermoelastic equations, and hence finding the damping, arises because of the need to solve the displacements and heat equation simultaneously. In this paper, we explore the possibility whether thermoelastic damping in a vibrating body can be estimated based on its material properties and geometry only and without solving the heat equation. Towards the end, we introduce the notion of a modal temperature, similar to modal displacements that are often used to study vibrations of continuous systems. The rationale for introducing a modal temperature is as follows. The local strain and its rate of change, in the presence of thermoelastic coupling, determine the local temperature gradients and the rate of temperature change. Since the strain field and its rate of change can be captured using a modal displacement and the corresponding modal frequency, it might be possible to find a modal temperature that describes the temperature field in the body. This model assumes no heat flow in or out of the body and that the thermal gradients created by the strain field in one direction are much larger compared to the other two. While these assumptions may seem quite restrictive, most practical structures used in micromechanical thin beams and plates obey these assumptions [14,18].

In Sec. 2, we outline a simple spring-mass model for thermoelastic damping, and show that it exhibits damping characteristics similar to those observed in real systems. We then describe a more generalized SDOF model for thermoelastic damping and outline a procedure to derive its governing equations in Sec. 3. In Sec. 4 we use this model to find the thermoelastic damping in a vibrating cantilever beam. In Sec. 5 we compare the results obtained from the model with those previously obtained in the literature and discuss the reasons for differences between them. In the final section we discuss as to how this SDOF model can be adapted to other geometries and potentially to model other forms of dissipation.

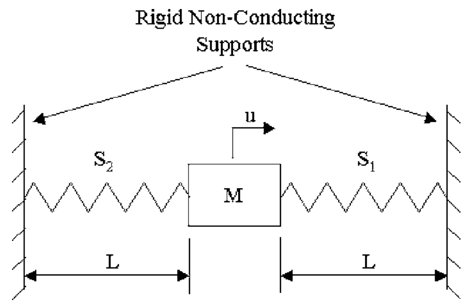


Fig. 1 Schematic of spring-mass model for thermoelasticity

2 Idealized Spring-Mass Model for Thermoelasticity

The spring-mass model (Fig. 1) comprises of a single mass M attached to two identical springs. The springs, labeled S_1 and S_2, are attached to rigid supports at their other ends. The springs are simply two elastic bars of length L, with elastic modulus E and cross-sectional area A and hence are of stiffness $K=AE/L$. The supports are assumed to be nonconducting and there is no heat transfer from the springs to the surroundings and vice versa. The springs are assumed to have very large thermal conductivity due to which the temperature within them is uniform at all times. In all the analyses below, the following initial conditions are imposed, i.e., at $t=0$:

(1) The displacement, u, of the mass satisfies, $u(0)=u_0$ and $du/dt|_{t=0}=0$;
(2) the entire system is at a uniform temperature T_0.

2.1 Spring-Mass Model with Nonconducting Mass. When the mass is displaced by a distance u ($u \ll L$), the strain in S_1, ϵ_1, is $-u/L$, and the strain in S_2, ϵ_2, is u/L. These strains induce a change in the temperature of the two springs because of thermoelastic coupling. It is assumed that the strains are uniform in each of the springs and that there is no thermal contact between the two springs, i.e., the mass M has zero thermal conductivity.

The generalized heat equation in the presence of thermoelastic coupling for an isotropic solid is given by [20]

$$\frac{\partial T}{\partial t} = \frac{k}{\rho c_v}\nabla^2 T - \frac{E\alpha T}{(1-2\nu)\rho c_v}\frac{\partial}{\partial t}\sum \epsilon_{kk} \tag{1}$$

where k is the thermal conductivity, ρ is the density, c_v is the specific heat capacity at constant volume, E is the Young's modulus, α is the linear thermal expansion coefficient, ν is the Poisson's ratio, and ϵ_{kk} are the normal components of strain.

When a body is subject to uniaxial stress (σ_{xx}), the equation reduces to

$$\frac{\partial T}{\partial t} = \frac{k}{\rho c_v}\nabla^2 T - \frac{E\alpha T}{\rho c_v}\frac{\partial \epsilon_{xx}}{\partial t} \tag{2}$$

If the thermal conductivity, k, is zero, the equation further reduces to

$$\frac{\partial T}{\partial t} = -\frac{E\alpha T}{\rho c_v}\frac{\partial \epsilon}{\partial t} \tag{3}$$

Here, ϵ_{xx} has been replaced by ϵ. Equation (3) can be viewed as a relation connecting heat generation (or absorption) rate with the strain rate for the case of uniaxial stress.

If the strain ϵ has no spatial variation, as is assumed for the springs in the spring-mass model, the partial time derivatives in Eq. (3), can be replaced by total derivatives. Hence Eq. (3) reduces to

$$\frac{dT}{dt} = -\frac{E\alpha T}{\rho c_v}\frac{d\epsilon}{dt} \tag{4}$$

Since the change in temperature due to thermoelastic effect is very small compared to the initial temperature T_0, T on the right hand side of Eq. (4) can be replaced by T_0. Integrating Eq. (4), after this modification, we get

$$\theta = -\frac{E\alpha T_0}{\rho c_v}\epsilon + N \tag{5}$$

where $\theta = T - T_0$, is the change in temperature from the initial temperature T_0 at the current strain ϵ, $N = (E\alpha T_0/\rho c_v)\epsilon_0$ is the constant of integration and ϵ_0 is the strain at $t=0$.

When S_1 is subjected to a strain $-u/L$, its temperature increases by $\theta_1 = E\alpha T_0 u/\rho c_v L + N_1$. Due to thermal expansion, the length of the spring increases by $\Delta L_1 = \alpha\theta_1 L = E\alpha^2 T_0 u/\rho c_v + \alpha N_1 L$. In effect, the net compression experienced by S_1 is $u(1+E\alpha^2 T_0/\rho c_v) + \alpha N_1 L$. Hence, the force exerted by S_1 on the mass M is given by $F_1 = -K(u(1+E\alpha^2 T_0/\rho c_v) + \alpha N_1 L)$. Similarly, the force exerted by S_2 is $F_2 = -K(u(1+E\alpha^2 T_0/\rho c_v) - \alpha N_2 L)$. The initial conditions, $\theta_1 = \theta_2 = 0$ and $u = u_0$ at $t=0$ can be used to find N_1 and N_2.

Hence, the equation of motion of the spring-mass system is given by

$$M\frac{d^2u}{dt^2} + 2K\left(1 + \frac{E\alpha^2 T_0}{\rho c_v}\right)u + K(\alpha N_1 L - \alpha N_2 L) = 0 \tag{6}$$

which is of the form

$$a'\frac{d^2u}{dt^2} + b'u + c' = 0 \tag{7}$$

and has the general solution $u = -c'/b' + P_1\cos(\omega t) + P_2\sin(\omega t)$ where $\omega = \sqrt{b'/a'}$. On imposing the initial conditions the solution reduces to $u = -c'/b' + P\cos(\omega t)$ where P is determined by the condition $u = u_0$ at $t=0$.

The solution to the equation of motion of the discrete spring-mass system with zero thermal conductivity leads to the following conclusions:

(1) The motion is harmonic with no decay in amplitude, i.e., with no damping, as is the case with any continuous system that has zero thermal conductivity;
(2) the natural frequency, $\sqrt{2K(1+E\alpha^2 T_0/\rho c_v)/M}$ of the thermoelastically coupled system is higher than $\sqrt{2K/M}$, the frequency of the uncoupled system ($\alpha=0$), again a characteristic of any continuous system;
(3) the point at which the mass experiences zero force is shifted from $u=0$, of the uncoupled system, to $u = c'/b' \neq 0$.

2.2 Spring-Mass Model with Conducting Mass. In this analysis, the mass M is assumed to have a finite thermal conductivity which leads to heat flow between the springs when there is a finite temperature difference between them. The rate of heat flow between the springs is assumed to be proportional to the difference in temperature between the springs, i.e., $\dot{q} = kL_c(T_1 - T_2) = kL_c(\theta_1 - \theta_2)$ (since $T_1 = T_0 + \theta_1$ and $T_2 = T_0 + \theta_2$). Here k is the thermal conductivity of the mass and L_c is a parameter that determines the heat transfer across it.

In a small time interval dt, let du be the displacement of the mass and $d\theta_1$ and $d\theta_2$ be the change in temperature in S_1 and S_2 (Fig. 1). The change in force exerted by S_1 and S_2 on the mass is given by

$$dF_1 = -K(du + L\alpha\, d\theta_1) \tag{8}$$

$$dF_2 = -K(du - L\alpha\, d\theta_2) \tag{9}$$

Dividing by dt throughout, and using $dF/dt = Md^3u/dt^3$ the equation of motion of the spring-mass system becomes

$$M\frac{d^3u}{dt^3} + 2K\frac{du}{dt} + KL\alpha\left(\frac{d\theta_1}{dt} - \frac{d\theta_2}{dt}\right) = 0 \tag{10}$$

In Eq. (10), the masses of the springs (M_s) have been neglected since they are very small compared to M. To obtain the expressions for θ_1 and θ_2, the heat balance equation needs to be used. In the absence of thermal conductivity the relation governing the rate of temperature change as a function of the strain rate is given by Eq. (4). Since, in the spring-mass system K is analogous to E, M_s/L is analogous to ρ and u is analogous to ϵ, Eq. (4) for the mass-spring system becomes

$$\frac{dT}{dt} = \frac{d\theta}{dt} = -\frac{K\alpha LT}{M_s c_v}\frac{du}{dt} \tag{11}$$

If there is heat transfer, the heat transfer rate given by $\dot{q} = kL_c(\theta_1 - \theta_2)$ has to be added to Eq. (11). Hence, the heat balance equations for S_1 and S_2 are given by

$$M_s c_v\frac{d\theta_1}{dt} = K\alpha LT_1\frac{du}{dt} - kL_c(\theta_1 - \theta_2) \tag{12}$$

$$M_s c_v\frac{d\theta_2}{dt} = -K\alpha LT_2\frac{du}{dt} + kL_c(\theta_1 - \theta_2) \tag{13}$$

The first term on the RHS of these two equations can be considered as the rate at which heat is generated/absorbed due to change in displacement with respect to time. Equation (10) along with Eqs. (12) and (13) describe the dynamics of the spring-mass system in the presence of thermoelastic coupling. These are a set of nonlinear differential equations and have to be solved numerically since there is no straightforward method for obtaining an analytical solution. But, since the change in temperatures, θ_1 and θ_2, induced by thermoelastic effect is very small compared to the initial temperature T_0, we can replace T_1 and T_2 in the first term on the RHS of Eqs. (12) and (13) by T_0, i.e., linearize Eqs. (12) and (13) about T_0 and consider only the zeroth order term. Therefore, we get

$$M_s c_v\frac{d\theta_1}{dt} = -M_s c_v\frac{d\theta_2}{dt} = K\alpha LT_0\frac{du}{dt} - kL_c(\theta_1 - \theta_2) \tag{14}$$

One consequence of linearization is that there will be no net increase in the temperature of the body, i.e., the net heat generated in the body will be zero, even though there is a reduction in potential energy. The linearized equations can easily be solved analytically as will be shown below.

Defining a new variable $\theta^* = \theta_1 - \theta_2$, Eqs. (10) and (14) can be reduced to

$$M\frac{d^3u}{dt^3} + 2K\frac{du}{dt} + KL\alpha\frac{d\theta^*}{dt} = 0 \tag{15}$$

$$M_s c_v\frac{d\theta^*}{dt} = 2K\alpha LT_0\frac{du}{dt} - 2kL_c\theta^* \tag{16}$$

Integrating Eq. (15) with respect to time and using the initial condition, $Md^2u/dt^2 = -2Ku$ and $\theta^*(0) = 0$, we get

$$M\frac{d^2u}{dt^2} + 2Ku + KL\alpha\theta^* = 0 \tag{17}$$

Substituting for θ^* in Eq. (16) from Eq. (17), we get

$$G_1\frac{d^3u}{dt^3} + G_2\frac{d^2u}{dt^2} + G_3\frac{du}{dt} + G_4 u = 0 \tag{18}$$

where $G_1 = M_s M c_v/KL\alpha$, $G_2 = 2kL_c M/KL\alpha$, $G_3 = 2(KL\alpha T_0 + M_s c_v/L\alpha)$ and $G_4 = 4kL_c/L\alpha$ are positive constants. Since Eq. (18) is a third order ordinary differential equation with constant coefficients, we can look for solutions of the form $u = Ae^{pt}$. Substituting this in Eq. (18), we get

$$G_1p^3 + G_2p^2 + G_3p + G_4 = 0 \tag{19}$$

If p_1, p_2, and p_3 are the roots of Eq. (19), the solution for u is given by

$$u = U_1e^{p_1t} + U_2e^{p_2t} + U_3e^{p_3t} \tag{20}$$

where U_1, U_2, and U_3 are determined by the initial conditions.

Since, all the coefficients in Eq. (19) are positive, p can either be real and negative or complex with negative real part. Unless the system is over damped, p_1 and p_2 will be complex with negative real parts and p_3 will be real and negative. Let $p_1=r+is$ and $p_2=r-is$, where r is negative and hence represents the decaying component of the solution in steady state while s represents the harmonic component. Since the damping due to thermoelastic effect is small, in general r will be much smaller in magnitude compared to s. If p_3 is comparable in magnitude to s, the transient part (the third term in the RHS of Eq. (20)) goes to zero in a small time and the system achieves steady state quickly. Using the initial conditions, $u(0)=0$ and $\dot{u}(0)=0$ it can be shown that $U_1=U_2=P$, where P is a constant, and hence the solution in steady state reduces to $u=2Pe^{rt}\cos(st)$.

The parameter L_c, as mentioned earlier, determines the heat flow in the system and hence the damping due to thermoelastic effect. $L_c=0$ corresponds to the adiabatic case while $L_c=\infty$ corresponds to the isothermal case, both of which lead to zero damping. For any other finite value of L_c there will be finite dissipation due to thermoelastic effect with the damping peaking at an unique value of L_c. In effect, one needs to determine L_c for the system of interest before the model can be used to predict its thermoelastic behavior. The equation of motion obtained for the case of nonconducting mass in the previous section can also be obtained from Eqs. (16) and (17) by making $L_c=0$.

In the spring-mass model, the amplitude exhibits an exponential decay in steady state as is expected in real systems. The model implicitly incorporates the notion of a modal temperature as it models the heat flow in terms of the difference of two lumped temperatures T_1 and T_2 which represent two parts of the system that have opposite states of strain. This indicates that it might be possible to find a modal temperature for real systems.

3 Generalized Model for Thermoelastic Damping

The generalized model provides a framework for estimating the thermoelastic damping in a thermoelastically coupled system vibrating in a particular mode. In estimating the damping using this model, we make the following assumptions:

(1) The dynamics of the vibrating system can be captured by a single modal displacement and frequency;
(2) the stress and strain fields in the system and their rate of change, in the absence of thermoelastic coupling, are completely known for the mode of vibration we are interested in;
(3) the stress field remains unaltered even in the presence of thermoelastic coupling;
(4) the temperature change induced is determined by the uncoupled strain;
(5) there is no heat flow from the surroundings to the system or vice versa.

In effect, we try to incorporate the effect of thermoelastic coupling by treating it as a perturbation from the uncoupled state.

To obtain the thermoelastic damping using this model we adopt the following procedure:

(1) We partition the vibrating system into two regions, 1 and 2, that are anti-phase with respect to strain (i.e., the strains of these two regions are of opposite nature at all times) and choose one point in each of these regions to be their respective "modal points;"
(2) we take the temperature *changes* at these modal points (from now on referred to as "modal temperatures") to represent the temperature fields of the two regions. This is the crucial approximation underlying this model as it reduces the temperature field of the distributed system to two lumped modal temperatures.
(3) we use the heat balance for the two regions and the energy conservation in the system to obtain the three governing equations necessary to solve for the three independent variables, namely the two modal temperatures and the modal displacement;
(4) finally, we use the time evolution of the modal displacement, obtained by solving the governing equations, to compute the damping in the system.

We will show that the damping characteristics obtained from the model is insensitive to the choice of the modal points. In other words, we get the same damping characteristics irrespective of the choice of modal points as long as we follow the definitions consistently.

3.1 Derivation of Governing Equations. We start by considering the the energy conservation equation, which for a general system vibrating in a particular mode, in the absence of thermoelastic coupling, is given by

$$\frac{1}{2}M_{\mathrm{eff}}\left(\frac{du}{dt}\right)^2 + \frac{1}{2}K_{\mathrm{eff}}u^2 = E_0 \tag{21}$$

where, M_{eff} is the effective mass and K_{eff} is the effective stiffness with respect to the modal displacement, u, while E_0 is the initial energy of the system. In the presence of coupling this equation modifies to

$$\frac{1}{2}M_{\mathrm{eff}}\left(\frac{du}{dt}\right)^2 + \frac{1}{2}K_{\mathrm{eff}}u^2 - \frac{1}{2}\int_V \sigma_{ii}\alpha\theta\, dV + \rho c_v\int_V \theta\, dV = E_0 \tag{22}$$

where θ is the change in temperature from the initial temperature T_0 and V is the volume of the system. The third term on the LHS of Eq. (22) accounts for the strain energy due to thermal strain while the last term accounts for the change in heat content of the system. Equation (22) can be expanded as

$$\frac{1}{2}M_{\mathrm{eff}}\left(\frac{du}{dt}\right)^2 + \frac{1}{2}K_{\mathrm{eff}}u^2 - \frac{1}{2}\left(\int_{V_1} \sigma_{ii}\alpha\theta\, dV + \int_{V_2} \sigma_{ii}\alpha\theta\, dV\right) + \rho c_v\left(\int_{V_1} \theta\, dV + \int_{V_2} \theta\, dV\right) = E_0 \tag{23}$$

where V_1 and V_2 are the volumes of the two regions, 1 and 2, and hence $V_1+V_2=V$.

To proceed further, we consider a mode where only one stress component, σ, contributes most of the strain energy. As we are interested in solving for the modal displacement, u, and the modal temperatures, θ_1 and θ_2, we need to formulate the governing equations in terms of u, θ_1, and θ_2. Towards this end, we replace the integrals in Eq. (23) in terms of θ_1, θ_2 and σ_1, σ_2, the stresses at the modal points, by defining

$$\rho c_v\int_{V_i} \theta\, dV = A_i\theta_i \quad i=1,2 \tag{24}$$

$$\frac{1}{2}\int_{V_i} \sigma\alpha\theta\, dV = B_i\sigma_i\alpha\theta_i \quad i=1,2 \tag{25}$$

where A_i and B_i are constants. Equation (23), hence, becomes

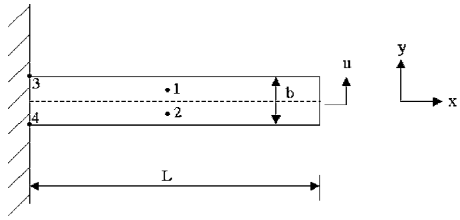


Fig. 2 Schematic of cantilever beam

$$\frac{1}{2}M_{\text{eff}}\left(\frac{du}{dt}\right)^2+\frac{1}{2}K_{\text{eff}}u^2-(B_1\sigma_1\alpha\theta_1+B_2\sigma_2\alpha\theta_2)+A_1\theta_1+A_2\theta_2=E_0 \tag{26}$$

Next, we consider the heat balance for the two regions represented by θ_1 and θ_2. The heat balance equations, assuming that at any time both θ_1 and θ_2 are much smaller than T_0, will be be similar to Eq. (14) and are given by

$$A_1\frac{d\theta_1}{dt}=-A_2\frac{d\theta_2}{dt}=H\frac{du}{dt}-kL_c(\theta_1-\theta_2) \tag{27}$$

where H is the constant relating the rate of heat generation/absorption in the two regions to the modal displacement (strain) rate. Equation (27) implies that the net heat generated in the system is zero which, as mentioned earlier, is a consequence of considering the linearized heat equation. Differentiating Eq. (26) with respect to time and noting that both σ_1, σ_2 are proportional to u, i.e., $\sigma_i=C_iu$, $i=1,2$ (C_i are constants), we get, using Eq. (27)

$$M_{\text{eff}}\left(\frac{d^2u}{dt^2}\right)\left(\frac{du}{dt}\right)+K_{\text{eff}}u\frac{du}{dt}-D_1\alpha\left(\frac{du}{dt}\theta_1+\frac{d\theta_1}{dt}u\right)-D_2\alpha\left(\frac{du}{dt}\theta_2+\frac{d\theta_2}{dt}u\right)=0 \tag{28}$$

where $D_i=B_iC_i$.

Equations (27) and (28) define the dynamics of the system in the presence of thermoelastic coupling. We can find the solution to these equations in the steady state by assuming, $u=u_0e^{i\omega t}$ and $\theta_1=\theta_{10}e^{i\omega t}$, $\theta_2=\theta_{20}e^{i\omega t}$. In general, ω will be complex with $\text{Re}(\omega)$ giving the new eigen frequency in the presence of thermoelastic damping and $|\text{Im}(\omega)|$ giving the attenuation of vibration. A measure of thermoelastic damping is then be given by [14]

$$Q^{-1}=2\left|\frac{\text{Im}(\omega)}{\text{Re}(\omega)}\right| \tag{29}$$

where Q^{-1} is the inverse of the quality factor.

Equation (27) involves a system parameter L_c which governs the heat flow within the system. L_c, which in general will depend on the geometry of the system, determines the lag between the temperature and stress fields and hence the dissipation. For certain simple cases we can get a good estimate of L_c by considering the heat flow in the system. But for more complicated geometries, it might have to be deduced from experiments. In Sec. 5, we provide an intuitive picture of L_c and outline a method for estimating it.

4 Application of Generalized Model to Vibrating Cantilever Beam

To estimate the thermoelastic damping in a thin cantilever beam vibrating in its first mode using the model, we first consider the small flexural displacements of the cantilever beam in the absence of thermoelastic coupling. The cantilever beam is taken to be of length L, thickness $b(\ll L)$ and width $w(\ll L)$ (Fig. 2). The x axis is defined to be parallel to the beam axis and y and z axes are parallel to surfaces with dimensions b and w, respectively. Since $b\ll L$ and $w\ll L$, it can be assumed that any plane cross section, initially perpendicular to the beam axis, remains perpendicular to the neutral plane during bending. If the surfaces of the beam are assumed to be stress free, only the σ_{xx} component of stress will be present. Since b is small (thin beam) this assumption holds good in the interior as well. Under these assumptions it can be shown that [20]

$$\epsilon_{xx}=-y\frac{\partial^2 v}{\partial x^2} \tag{30}$$

and the equation of motion is given by [20]

$$\rho A\frac{\partial^2 v}{\partial t^2}+EI\frac{\partial^4 v}{\partial x^4}=0. \tag{31}$$

where $A=bw$ is the area of its cross section, $I=b^3w/12$ is the moment of inertia about the z axis and $v(x,t)$ is the displacement of the beam along y direction at time t.

If the cantilever is given an initial displacement along y direction, such that its shape matches with its first mode shape and the tip displacement is u_0, and set into vibration, $v(x,t)$ will take the form $v(x,t)=v_0(x)\cos(\omega t)$, where $v_0(x)$ is the mode shape and ω is the natural frequency of the cantilever. The general solution for $v_0(x)$ is given by

$$v_0(x)=B(\sin(qx)-\sin h(qx)+R(\cos(qx)-\cos h(qx))) \tag{32}$$

where $R=(\cos(qL)+\cos h(qL))/(\sin(qL)-\sin h(qL))$ and B is determined by the initial displacement u_0 and is approximately equal to $u_0/2.724$. Since we are considering only the first mode, $q=q_1\approx 1.875/L$ and $R\approx 1.362$. Hence $v(x,t)$, for the first mode, is given by

$$v(x,t)=\frac{u_0\cos(\omega t)}{2.724}(\sin(q_1x)-\sin h(q_1x)+R(\cos(q_1x)-\cos h(q_1x))) \tag{33}$$

As $u_0\cos(\omega t)\equiv u(t)$, where $u(t)$ is the displacement of the cantilever tip, Eq. (33) can be written as

$$v(x,t)=\frac{u(t)}{2.724}(\sin(q_1x)-\sin h(q_1x)+R(\cos(q_1x)-\cos h(q_1x))) \tag{34}$$

We can find M_{eff} of the cantilever with respect to u by equating the kinetic energy of the cantilever to $M_{\text{eff}}(du/dt)^2/2$. K_{eff} can similarly be found by equating the potential energy to $K_{\text{eff}}u^2/2$. M_{eff} and K_{eff}, the modal mass and stiffness thus found, are given by $M_{\text{eff}}\approx 0.25M$ and $K_{\text{eff}}\approx 0.2575Eb^3w/L^3$. $M=\rho bLw$ is the total mass of the cantilever.

To proceed further, we choose the points 1 and 2, as shown in Fig. 2, to be the modal points and designate the temperature *changes* at these points, θ_1 and θ_2, to be the modal temperatures. The modal points have been chosen symmetrically merely for convenience. From Eqs. (24) and (25) we get

$$\rho c_v\int_0^L\int_0^{b/2}\int_{-w/2}^{w/2}\theta\,dz\,dy\,dx=A_1\theta_1 \tag{35}$$

$$\rho c_v\int_0^L\int_{-b/2}^{0}\int_{-w/2}^{w/2}\theta\,dz\,dy\,dx=A_2\theta_2 \tag{36}$$

$$\frac{1}{2}\int_0^L\int_0^{b/2}\int_{-w/2}^{w/2}\sigma\alpha\theta\,dz\,dy\,dx=B_1\sigma_1\alpha\theta_1 \tag{37}$$

$$\frac{1}{2}\int_0^L \int_{-b/2}^{0} \int_{-w/2}^{w/2} \sigma\alpha\theta\, dz\, dy\, dx = B_2\sigma_2\alpha\theta_2 \tag{38}$$

As H is the constant relating the rate of heat generation/absorption in the two regions to the modal displacement rate, we have

$$\rho c_v \int_0^L \int_0^{b/2} \int_{-w/2}^{w/2} \frac{\partial\theta}{\partial t} dz\, dy\, dx = H\frac{du}{dt} \tag{39}$$

To calculate the constants A_i, B_i, and H exactly we need to know the actual temperature profile, which can be obtained only by solving the heat equation for the system under consideration with appropriate boundary conditions. Since the very purpose of the model is to get an estimate of damping without solving the heat equation, we make the following approximation. We solve for the temperature profile for the case of no heat flow ($k \to 0$), so that the temperature profile is solely determined by the strain, and use it to determine the constants. As the stress field is also known, we can compute the constants A_1, A_2, B_1, B_2, and H. The values of the constants are $A_1 = A_2 = 0.5\ \rho c_v bLw$, $B_1 = B_2 = 0.5445\ wbL$, and $H = 0.172\ E\alpha T_0 wb^2/L$.

Using the stress field, we find the constants C_1 and C_2 that relate σ_1 and σ_2 with u to be $C_1 = -C_2 = -0.344Eb/L^2$. With these constants, together with M_{eff} and K_{eff} we can obtain the equations that govern the dynamics of the cantilever in the presence of thermoelastic coupling from Eqs. (27) and (28). The equations are given by

$$0.5Mc_v\frac{d\theta_1}{dt} = -0.5Mc_v\frac{d\theta_2}{dt} = 0.172\frac{E\alpha T_0 wb^2}{L}\frac{du}{dt} - kL_c(\theta_1 - \theta_2) \tag{40}$$

$$M_{\text{eff}}\frac{du}{dt}\left(\frac{d^2u}{dt^2}\right) + K_{\text{eff}}u\frac{du}{dt} + 0.1873\frac{E\alpha b^2 w}{L}\left(\frac{du}{dt}(\theta_1 - \theta_2) + u\left(\frac{d\theta_1}{dt} - \frac{d\theta_2}{dt}\right)\right) = 0 \tag{41}$$

Using $\theta^* = \theta_1 - \theta_2$, Eqs. (40) and (41) reduce to

$$0.5Mc_v\frac{d\theta^*}{dt} = 0.344\frac{E\alpha T_0 wb^2}{L}\frac{du}{dt} - 2kL_c\theta^* \tag{42}$$

$$M_{\text{eff}}\left(\frac{d^2u}{dt^2}\right)\left(\frac{du}{dt}\right) + K_{\text{eff}}u\frac{du}{dt} + 0.1873\frac{E\alpha b^2 w}{L}\left(\theta^*\frac{du}{dt} + u\frac{d\theta^*}{dt}\right) = 0 \tag{43}$$

5 Results and Discussion

To examine the validity of the approach outlined above in modeling thermoelastic dissipation in rectangular cantilever beams, we solve Eqs. (42) and (43) and compare the results with those previously published in literature. For solving the equations we take $\alpha = 2.616\times10^{-6}/{}^\circ\text{C}$, $c_v = 713\ \text{J/KgK}$, $k = 156\ \text{W/mK}$, $E = 1.68\times10^{11}\ \text{N/m}^2$, $\rho = 2330\ \text{Kg/m}^3$. Assuming a steady state solution of the form $u = u_0e^{i\omega t}$ and $\theta^* = \theta_0^* e^{i\omega t}$, we can solve Eqs. (42) and (43) for the complex valued frequency ω and hence the damping.

Figure 3 gives the plots of the normalized attenuation, $\zeta = |\text{Im}(\omega)|/(\omega_0\Delta_E)$ and normalized frequency shift, $\beta = (\text{Re}(\omega) - \omega_0)/(\omega_0\Delta_E)$, for a beam of dimensions $L = 4\times10^{-4}$ m, $b = 4\times10^{-5}$ m, and $w = 4\times10^{-5}$ m, as a function of L_c, where Δ_E is the relaxation strength [2] and ω_0 is the isothermal natural frequency. The relaxation strength can be understood as follows. When a periodic stress of frequency ω is applied to a body, the stress and strain amplitudes are related through the frequency dependent complex elastic modulus. The dissipation in the body, Q^{-1}, if small, is then equal to the ratio of the imaginary and real parts of the complex modulus giving $Q^{-1}/\Delta = \omega\tau/(1+\omega^2\tau^2)$, where Δ is the relaxation strength and τ is the relaxation time. For a thermoelastic solid the relevant relaxation strength is that of the Young's modulus giving $\Delta_E = (E_{ad} - E)/E = E\alpha^2T_0/\rho c_v$, where E_{ad} and E are the adiabatic and isothermal values of the Young's modulus.

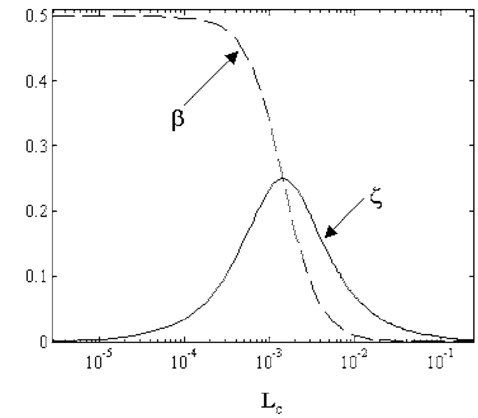


Fig. 3 Plot of normalized attenuation, ζ, and normalized frequency shift, β, as a function of L_c

As can be seen from Fig. 3, the frequency shift attains a maximum when $L_c \to 0$ and goes to zero as L_c becomes larger. The damping on the other hand attains a peak for an intermediary value of L_c and tends to zero at both the extremes. Since $L_c \to 0$ represents an adiabatic system and large values of L_c represent a nearly isothermal system, it can be seen that the damping behavior and frequency shift predicted by the model agrees well with results obtained by Zener [3] and Lifshitz et al. [14].

As mentioned in Sec. 3, we need to know the L_c of a system before we can find its thermoelastic damping. One method to get an estimate of L_c is as follows. Since, we take the heat flow between the two regions to be equal to $kL_c(\theta_1 - \theta_2)$, we have

$$k\int_S \frac{\partial\theta}{\partial n}dS = kL_c(\theta_1 - \theta_2) \tag{44}$$

where S is the surface through which heat flows between the regions represented by θ_1 and θ_2. To calculate L_c exactly from Eq. (44) we need to know the actual temperature profile. To circumvent this problem we make the same approximation that we made for evaluating the constants in Eqs. (35)–(39), i.e., solve for the temperature profile for the case of no heat flow ($k \to 0$) and use it to determine L_c. For the particular case of the vibrating cantilever that we are considering here, Eq. (44) becomes

$$k\int_0^L \int_{-w/2}^{w/2} \left.\frac{\partial\theta}{\partial y}\right|_{y=0} dz\, dx = kL_c(\theta_1 - \theta_2) \tag{45}$$

On solving Eq. (45) we get $L_c = 2Lw/b$. It is worth noting that for the simple case of cantilever beam considered here we can get an order of magnitude of L_c based on elementary physical considerations. The heat flow rate, $\dot{q}$, between two bodies at temperatures T_1 and T_2 is given by $\dot{q} = kA(T_1 - T_2)/d$ where k is the thermal conductivity, A is the cross sectional area, and d is the length of the heat conduction path. A comparison with $\dot{q} = kL_c(\theta_1 - \theta_2)$, used in the model, reveals that L_c is analogous to A/d. The appropriate area and length for the cantilever beam are $A = Lw$ and $d = b$, which gives $L_c \propto Lw/b$.

Figure 4 shows the plots of Q^{-1}/Δ_E as a function of b/L, with $L = 4\times10^{-4}$ m and $w = 4\times10^{-5}$ m, as obtained from the model (solving Eqs. (42) and (43) with $L_c = 2Lw/b$) and from Zener's approximation (Eq. (46)). Zener's approximation for Q^{-1} is given by

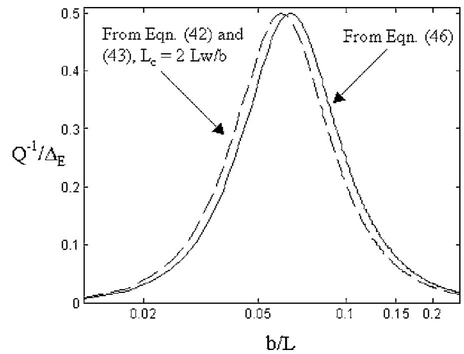


Fig. 4 Plot of Q^{-1}/Δ_E versus b/L obtained from the model and Zener's result

$$\frac{Q^{-1}}{\Delta_E} = \frac{\omega\tau}{1+\omega^2\tau^2} \tag{46}$$

where $\tau = b^2/\pi^2\chi$. Since $\omega \propto b/L^2$, Q^{-1}/Δ_E is just a function of b^3/L^2. In other words, if b^3/L^2 is kept constant, the damping remains constant. If we solve Eqs. (42) and (43) keeping b^3/L^2 constant, the damping remains constant, irrespective of the value of w. Further, if we choose $L_c = 2.466Lw/b$, the damping predicted by the model and Eq. (50) match exactly. These results imply that

(1) The damping predicted by the model is also just a function of b^3/L^2;
(2) the functional form of the relation between damping and b^3/L^2 predicted by the model is the same as that of Zener's approximation.

While the qualitative damping behavior predicted by the model agrees well with Zener's result, the quantitative agreement is not exact because of the approximation involved in evaluating the constants in Eqs. (35)–(39) as well as L_c. In estimating these constants and L_c, we implicitly assume that the temperature distribution within the cantilever remains similar irrespective of whether there is heat flow or not. In other words, if θ_A and θ_B represent the temperature changes at two arbitrary points in steady state, we assume θ_A/θ_B, with heat flow (k is finite) is equal to θ_A/θ_B when there is no heat flow ($k=0$). But, solving the heat equation explicitly gives a temperature distribution that depends on k, as shown in for example [14], leading to the discrepancy. This k dependence of the temperature distribution, at first sight, would suggest that at larger values of k, the damping characteristics obtained from the model will be substantially different from those obtained from Zener's result. But, the closeness between the damping characteristics obtained from the model and Zener's result remains identical for all finite values of k. This suggests that the degree of approximation is insensitive to the value of k, which we think is a consequence of averaging the temperature distribution in evaluating the constants.

As mentioned earlier, the damping characteristics obtained from the model is independent of the choice of modal points. For example, if we choose the two modal points as the corners of the upper and lower halves of the cantilever (points 3 and 4 in Fig. 2), we obtain

$$0.09785Mc_v\frac{d\theta^*}{dt} = 0.344\frac{E\alpha T_0 wb^2}{L}\frac{du}{dt} - 2kL_c\theta^* \tag{47}$$

$$M_{\mathrm{eff}}\left(\frac{d^2u}{dt^2}\right)\left(\frac{du}{dt}\right) + K_{\mathrm{eff}}u\frac{du}{dt} + 0.03665\frac{E\alpha b^2 w}{L}\left(\theta^*\frac{du}{dt} + u\frac{d\theta^*}{dt}\right) = 0 \tag{48}$$

where M_{eff} and K_{eff} are the same as in the previous case but $L_c = 0.3914Lw/b$ instead of $2Lw/b$. Solving Eqs. (47) and (48) leads to exactly the same damping behavior as shown in Fig. 4.

The functional dependence of L_c on the dimensions (like length, radius, width, etc.) of the vibrating system remains the same irrespective of the size. Once this functional relationship is known, the damping in any similar system vibrating in the same mode can be easily obtained. This will be especially useful for estimating thermoelastic damping in systems with complex geometries, as the model requires only the knowledge of the stress field. For example if we need to find the thermoelastic damping of beams with U-shaped cross section vibrating in a particular mode, we need to determine the functional relationship only once, and we can use the model to predict the damping at all size scales. But, when applying the model to more complex geometries two issues need to be taken into consideration:

(1) The functional dependence of L_c on the the dimensions of the body may not be obvious, i.e., while it was easy to see that $L_c \propto L$, w and $L_c \propto 1/b$ for the rectangular cantilever beam, such dependence may not be apparent for more complex geometries;
(2) Estimating L_c by solving Eq. (44) might be more difficult.

One way to overcome these problems would be to experimentally obtain the damping for the system of interest at various size scales, and determine the functional dependence of L_c on the dimensions using it. For this method to work, though, one would have to make sure that energy dissipation due to other causes are minimal. A second method would be to obtain the damping at different size scales by numerically solving the equations of thermoelasticity and determine the relationship between L_c and the dimensions of the system.

The simple model described in this work essentially depends on finding one macroscopic system parameter, L_c, to determine the thermoelastic damping in a vibrating solid. We have shown that this model provides a reasonably good estimate of thermoelastic damping at least for the case of a rectangular cantilever beam. This raises the possibility that similar macroscopic parameters and a corresponding damping model can be used to estimate dissipation due to other relaxation processes like dislocation dynamics or grain boundary relaxation. For example, the average grain size might turn out to be a macroscopic parameter on which grain boundary relaxation depends. Finding a good damping model, though, might prove to be more difficult, as many of these relaxation processes, unlike thermoelastic damping, do not have a well developed theoretical framework or governing equations.

6 Concluding Remarks

A SDOF model for estimating thermoelastic damping in structures vibrating in specific modes is proposed in this work. The model, incorporating the notion of modal temperatures, estimates the thermoelastic damping in a structure without explicitly solving the heat equation. The good qualitative agreement between the predictions of the model and Zener's well known approximation, for the rectangular cantilever beam, suggests that the model captures the essence of thermoelastic damping in vibrating bodies. The model can be used to estimate thermoelastic damping in structures with more complicated geometries if their corresponding characteristic length, L_c, can be obtained.

The simplicity of the model rests on several assumptions: (a) The dynamics of the vibrating body can be captured by a single modal displacement and frequency, (b) the stress and strain fields are completely known for the uncoupled state, (c) the dynamic response of the body in the presence of thermoelastic coupling differs only slightly from the uncoupled state, and (d) there is no thermal interaction between the body and the environment. Finally, the ability of this model to provide a good qualitative picture of thermoelastic damping suggests that other forms of dissipation might also be amenable for description using such simple models.

Acknowledgment

This work was supported by National Science Foundation under the Grant NSF ECS-0304243.

References

[1] Nowacki, W., 1986, *Thermoelasticity*, 2nd ed., Pergamon Press, New York.
[2] Nowick, A. S., and Berry, B. S., 1972, *Anelastic Relaxation in Crystalline Solids*, Academic, New York.
[3] Zener, C., 1937, "Internal Friction in Solids. I. Theory of Internal Friction in Reeds," Phys. Rev., **52**, pp. 230–235.
[4] Zener, C., 1938, "Internal Friction in Solids. II. General Theory of Thermoelastic Internal Friction," Phys. Rev., **53**, pp. 90–99.
[5] Zener, C., Otis, W., and Nuckolls, R., 1938, "Internal Friction in Solids. III. Experimental Demonstration of Thermoelastic Internal Friction," Phys. Rev., **53**, pp. 100–101.
[6] Alblas, J. B., 1961, "On the General Theory of Thermo-Elastic Friction," Appl. Sci. Res., Sect. A, **10**, pp. 349–362.
[7] Alblas, J. B., 1981, "A Note on the General Theory of Thermoelastic Damping," J. Therm. Stresses, **4**, pp. 333–355.
[8] Chadwick, P., 1962, "On the Propagation of Thermoelastic Disturbances in Thin Plates and Rods," J. Mech. Phys. Solids, **10**, pp. 99–109.
[9] Lord, H. W., and Shulman, Y., 1967, "A Generalized Dynamical Theory of Thermoelasticity," J. Mech. Phys. Solids, **15**, pp. 299–309.
[10] Nguyen, C. T. C., 2004, "Vibrating RF Mems for Next Generation Wireless Applications," Customs Integrated Circuits Conferencee, 2004, *Proceedings of the IEEE 2004*, 3–6, Oct. 2004, IEEE, Piscataway, NJ, pp. 257–274.
[11] Cleland, A. N., and Roukes, M. L., 1998, "A Nanometer-Scale Mechanical Electrometer," Nature (London), **392**, pp. 160–162.
[12] Tilman, H. A. C., Elwenspoek, M., and Fluitman, J. H. J., 1992, "Micro Resonant Force Gauges," Sens. Actuators, A, **30**, pp. 35–53.
[13] Zook, J. D., Burns, D. W., Guckel, H., Sniegowski, J. J., Engelstad, R. L., and Feng, Z., 1992, "Characteristics of Polysilicon Resonant Microbeams," Sens. Actuators, A, **35**, pp. 51–59.
[14] Lifshitz, R., and Roukes, M. L., 2000, "Thermoelastic Damping in Micro- and Nanomechanical Systems," Phys. Rev. B, **61**, pp. 5600–5609.
[15] D. M. Photiadis, A. B. H. H., Liu, X., Bucaro, J. A., and Marcus, M. H., 2002, "Thermoelastic Loss Observed in a High q Mechanical Oscillator," Physica B, **316**, pp. 408–410.
[16] Houston, B. H., Photiadis, D. M., Marcus, M. H., Bucaro, J. A., Liu, X., and Vignola, J. F., 2002, "Thermoelastic Loss in Microscale Oscillators," Appl. Phys. Lett., **80**, pp. 1300–1302.
[17] Houston, B. H., Photiadis, D. M., Vignola, J. F., Marcus, M. H., Liu, X., Czaplewski, D., Sekaric, L., Butler, J., Pehrsson, P., and Bucaro, J. A., 2004, "Loss Due to Transverse Thermoelastic Currents in Microscale Resonators," Mater. Sci. Eng., A, **370**, pp. 407–411.
[18] Nayfeh, A. H., and Younis, M. I., 2004, "Modeling and Simulations of Thermoelastic Damping in Microplates," J. Micromech. Microeng., **14**, pp. 1711–1717.
[19] Norris, A. N., and Photiadis, D. M., 2005, "Thermoelastic Relaxation in Elastic Structures with Applications to Thin Plates," Q. J. Mech. Appl. Math., **58**, pp. 143–163.
[20] Landau, L. D., and Lifshitz, E. M., 1959, *Theory of Elasticity*, Pergamon Press, London.

Null-Field Approach for the Multi-inclusion Problem Under Antiplane Shears

Jeng-Tzong Chen
e-mail: jtchen@mail.ntou.edu.tw

An-Chien Wu

Department of Harbor and River Engineering,
National Taiwan Ocean University,
Keelung 20224,
Taiwan

*In this paper, we derive the null-field integral equation for an infinite medium containing circular holes and/or inclusions with arbitrary radii and positions under the remote antiplane shear. To fully capture the circular geometries, separable expressions of fundamental solutions in the polar coordinate for field and source points and Fourier series for boundary densities are adopted to ensure the exponential convergence. By moving the null-field point to the boundary, singular and hypersingular integrals are transformed to series sums after introducing the concept of degenerate kernels. Not only the singularity but also the sense of principle values are novelly avoided. For the calculation of boundary stress, the Hadamard principal value for hypersingularity is not required and can be easily calculated by using series sums. Besides, the boundary-layer effect is eliminated owing to the introduction of degenerate kernels. The solution is formulated in a manner of semi-analytical form since error purely attributes to the truncation of Fourier series. The method is basically a numerical method, and because of its semi-analytical nature, it possesses certain advantages over the conventional boundary element method. The exact solution for a single inclusion is derived using the present formulation and matches well with the Honein et al.'s solution by using the complex-variable formulation (Honein, E., Honein, T., and Hermann, G., 1992, Appl. Math., **50**, pp. 479–499). Several problems of two holes, two inclusions, one cavity surrounded by two inclusions and three inclusions are revisited to demonstrate the validity of our method. The convergence test and boundary-layer effect are also addressed. The proposed formulation can be generalized to multiple circular inclusions and cavities in a straightforward way without any difficulty.* [DOI: 10.1115/1.2338056]

1 Introduction

The distribution of stress in an infinite medium containing circular holes and/or inclusions under the antiplane shear has been studied by many investigators. However, analytical solutions are rather limited except for simple cases. To the authors' best knowledge, an exact solution of a single inclusion was derived by Honein et al. [1] using the complex potential. Besides, analytical solutions for two identical holes and inclusions were obtained by Stief [2] and by Budiansky and Carrier [3], respectively. Zimmerman [4] employed the Schwartz alternative method for plane problems with two holes or inclusions to obtain a closed-form approximate solution. In addition, Sendeckyj [5] proposed an iterative scheme for solving problems of multiple inclusions. However, the approach is rather complicated and explicit solutions were not provided. Numerical solutions for problems with two unequal holes and/or inclusions were provided by Honein et al. [1] using the Möbius transformations involving the complex potential. Not only antiplane shears but also screw dislocations were considered. Numerical results were presented by Goree and Wilson [6] for an infinite medium containing two inclusions under the remote shear. Bird and Steele [7] used a Fourier series procedure to revisit the antiplane elasticity problems of Honein et al.'s paper [1]. To approximate the Honein et al.'s infinite problem, an equivalent bounded-domain approach with the stress applied on the outer boundary was utilized. A shear stress σ_{zr} on the outer boundary is used in place of a stress σ_{zy} at infinity to approach the Honein et al.'s results as the radius becomes large. Wu [8] solved the analytical solution for two inclusions under the remote shear in two directions by using the conformal mapping and the theorem of analytic continuation. Based on the technique of analytical continuity and the method of successive approximation, Chao and Young [9] studied the stress concentration on a hole surrounded by two inclusions. For a triangle pattern of three inclusions, Gong [10] employed the complex potential and Laurent series expansion to calculate the stress concentration. Complex variable boundary element method was utilized to deal with the problem of two circular holes by Chou [11] and Ang and Kang [12], independently. To provide a general solution to the antiplane interaction among multiple circular inclusions with arbitrary radii, shear moduli, and location is not trivial. Mathematically speaking, only circular boundaries in an infinite domain are concerned here. Mogilevskaya and Crouch [13] have also employed Fourier series expansion technique and used the Galerkin method instead of collocation technique to solve the problem of circular inclusions in 2D elasticity. The advantage of their method is that one can tackle a lot of inclusions even inclusions touching one another. However, they did not expand a fundamental solution into a degenerate kernel in the polar coordinate. Degenerate kernels play an important role not only for mathematical analysis [14] but also for numerical implementation. For example, the spurious eigenvalue [15], fictitious frequency [16], and degenerate scale [17] have been mathematically and numerically studied by using degenerate kernels for problems with circular boundaries. One gain is that exponential convergence instead of algebraic convergence in the boundary element method (BEM) can be achieved using the Fourier expansion [14]. Chen et al. [18] have successfully solved the antiplane problem with circular holes using the null-field integral equation in conjunction with the degenerate kernel and Fourier series. The

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 15, 2005; final manuscript received May 22, 2006. Review conducted by Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California–Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

extension to biharmonic problems was also implemented [19]. This paper extends the idea to solve problems with circular inclusions.

By introducing a multi-domain approach, an inclusion problem can be decomposed into two parts. One is the infinite medium with circular holes and the other is the problem with each circular inclusion. After considering the continuity and equilibrium conditions on the interface between the matrix and inclusion, a linear algebraic system is obtained and the unknown Fourier coefficients in the algebraic system can be determined. Then, the field potential and stress are easily obtained. Furthermore, an arbitrary number of circular inclusions can be treated by using the present method without any difficulty. One must take care the vector decomposition in using the adaptive observer system for the nonconfocal case. Also, the boundary stress is easily determined by using series sums instead of employing the sense of Hadamard principal value. A general purpose program for arbitrary number of circular inclusions with various radii and arbitrary positions was developed. The infinite medium with multiple circular holes [18] can be solved as a limiting case of zero shear modulus of inclusions by using the developed program. Several examples solved previously by other researchers [1–3,6,8–10] were revisited to see the accuracy and efficiency of the present formulation. In addition, the test of convergence is done and the boundary-layer effect for the calculation of stresses is also addressed.

2 Problem Statement

The displacement field of the antiplane deformation is defined as

$$u = v = 0, \quad w = w(x,y), \tag{1}$$

where w is the only nonvanishing component of displacement with respect to the Cartesian coordinate which is a function of x and y. For a linear elastic body, the stress components are

$$\sigma_{zx} = \mu \frac{\partial w}{\partial x}, \tag{2}$$

$$\sigma_{zy} = \mu \frac{\partial w}{\partial y}, \tag{3}$$

where μ is the shear modulus. The equilibrium equation can be simplified to

$$\frac{\partial \sigma_{zx}}{\partial x} + \frac{\partial \sigma_{zy}}{\partial y} = 0. \tag{4}$$

Thus, we have

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = \nabla^2 w = 0. \tag{5}$$

Equation (5) indicates that the governing equation of this problem is the Laplace equation. We consider an infinite medium subject to N circular inclusions bounded by the B_k contour $(k=1,2,\ldots,N)$ for either matrix or inclusions under the antiplane shear σ_{zx}^{∞} and σ_{zy}^{∞} at infinity or equivalently under the displacement $w^{\infty} = \sigma_{zx}^{\infty} x/\mu + \sigma_{zy}^{\infty} y/\mu$ as shown in Fig. 1(a). By taking the free body along the interface between the matrix and inclusions, the problem can be decomposed into two systems. One is an infinite medium with N circular holes under the remote shear and the other is N circular inclusions bounded by the B_k contour which satisfies the Laplace equation as shown in Figs. 1(b) and 1(c), respectively. From the numerical point of view, this is the so-called multi-domain approach. For the problem in Fig. 1(b), it can be superimposed by two parts. One is an infinite medium under the remote shear and the other is an infinite medium with N circular holes which satisfies the Laplace equation as shown in Figs. 1(d) and 1(e), respectively. This part was solved efficiently by Chen et al. [18] and the null-field equation approach is adapted here again. Therefore, one exterior problem for the matrix is shown in Fig. 1(e) and several interior problems for nonoverlapping inclusions are shown in Fig. 1(c). According to the null-field integral formulation in Ref. [18], the two problems in Figs. 1(e) and 1(c) can be solved in a unified manner since they both satisfy the Laplace equation.

3 A Unified Formulation for Exterior and Interior Problems

3.1 Dual Boundary Integral Equations and Dual Null-Field Integral Equations. The boundary integral equation for the domain point can be derived from the third Green's identity [20], we have

$$2\pi w(x) = \int_B T(s,x) w(s) dB(s) - \int_B U(s,x) t(s) dB(s), \quad x \in D, \tag{6}$$

$$2\pi \frac{\partial w(x)}{\partial n_x} = \int_B M(s,x) w(s) dB(s) - \int_B L(s,x) t(s) dB(s), \quad x \in D, \tag{7}$$

where $t(s) = \partial w(s)/\partial n_s$, s and x are the source and field points, respectively, B is the boundary, D is the domain of interest, n_s and n_x denote the outward normal vector at the source point s and field point x, respectively, and the kernel function $U(s,x) = \ln r$, ($r \equiv |x-s|$), is the fundamental solution which satisfies

$$\nabla^2 U(s,x) = 2\pi \delta(x-s), \tag{8}$$

in which $\delta(x-s)$ denotes the Dirac-delta function. The other kernel functions, $T(s,x)$, $L(s,x)$, and $M(s,x)$, are defined by

$$T(s,x) \equiv \frac{\partial U(s,x)}{\partial n_s}, \quad L(s,x) \equiv \frac{\partial U(s,x)}{\partial n_x}, \quad M(s,x) \equiv \frac{\partial^2 U(s,x)}{\partial n_s \partial n_x}. \tag{9}$$

By collocating x outside the domain $(x \in D^c)$, we obtain the dual null-field integral equations as shown below

$$0 = \int_B T(s,x) w(s) dB(s) - \int_B U(s,x) t(s) dB(s), \quad x \in D^c, \tag{10}$$

$$0 = \int_B M(s,x) w(s) dB(s) - \int_B L(s,x) t(s) dB(s), \quad x \in D^c, \tag{11}$$

where D^c is the complementary domain. Based on the separable property, the kernel function $U(s,x)$ is expanded into the degenerate form by separating the source point and field point in the polar coordinate [21]

$$U(s,x) = \begin{cases} U^i(R,\theta;\rho,\phi) = \ln R - \sum_{m=1}^{\infty} \frac{1}{m}\left(\frac{\rho}{R}\right)^m \cos m(\theta-\phi), & R \geqslant \rho \\ U^e(R,\theta;\rho,\phi) = \ln \rho - \sum_{m=1}^{\infty} \frac{1}{m}\left(\frac{R}{\rho}\right)^m \cos m(\theta-\phi), & \rho > R \end{cases}, \tag{12}$$

where the superscripts "i" and "e" denote the interior $(R>\rho)$ and exterior $(\rho>R)$ cases, respectively. The origin of the observer system for the degenerate kernel is (0,0). Figure 2 shows the graph of separate expressions of fundamental solutions where

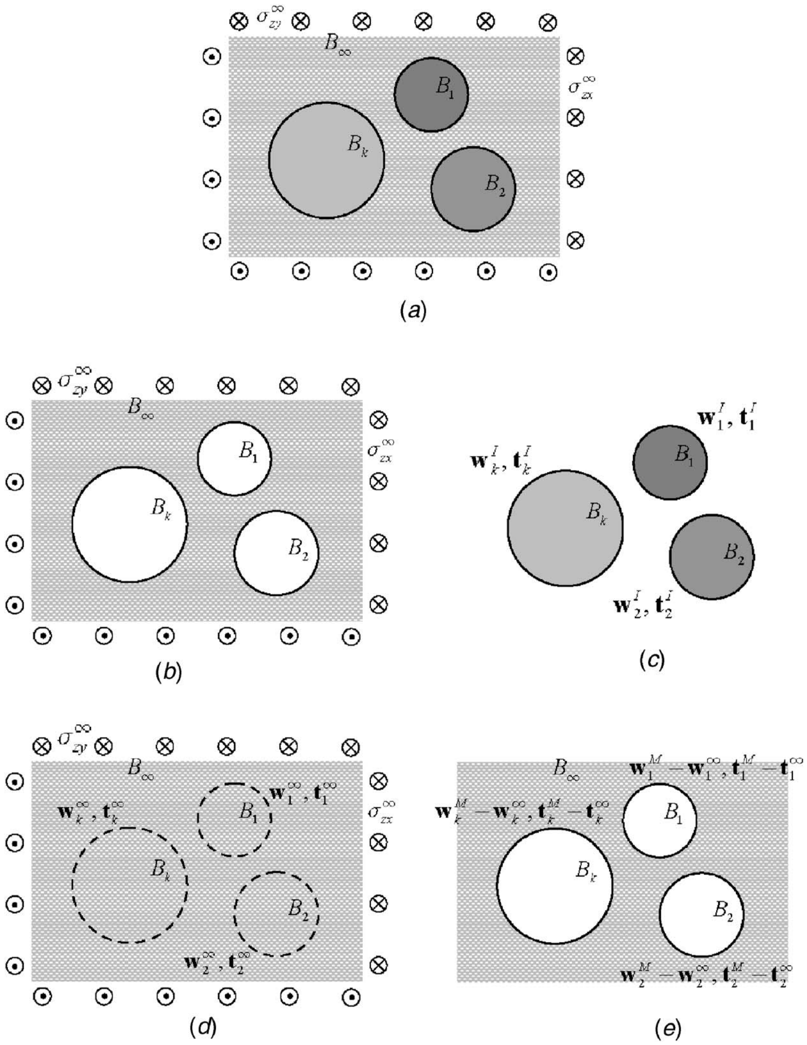


Fig. 1 (***a***) **Infinite antiplane problem with arbitrary circular inclusions under the remote shear,** (***b***) **infinite medium with circular holes under the remote shear,** (***c***) **interior Laplace problems for each inclusion,** (***d***) **infinite medium under the remote shear, and** (***e***) **exterior Laplace problems for the matrix**

source point s located at $R=10.0$ and $\theta=\pi/3$. By setting the origin at o for the observer system, a circle with radius R from the origin o to the source point s is plotted. If the field point x is situated inside the circular region, the degenerate kernel belongs to the interior expression of U^i; otherwise, it is the exterior case. After taking the normal derivative $\partial/\partial R$ with respect to Eq. (12), the $T(s,x)$ kernel yields

$$T(s,x) = \begin{cases} T^i(R,\theta;\rho,\phi) = \dfrac{1}{R} + \displaystyle\sum_{m=1}^{\infty} \left(\frac{\rho^m}{R^{m+1}}\right)\cos m(\theta-\phi), & R > \rho \\ T^e(R,\theta;\rho,\phi) = -\displaystyle\sum_{m=1}^{\infty} \left(\frac{R^{m-1}}{\rho^m}\right)\cos m(\theta-\phi), & \rho > R \end{cases}, \tag{13}$$

and the higher-order kernel functions, $L(s,x)$ and $M(s,x)$, are shown below

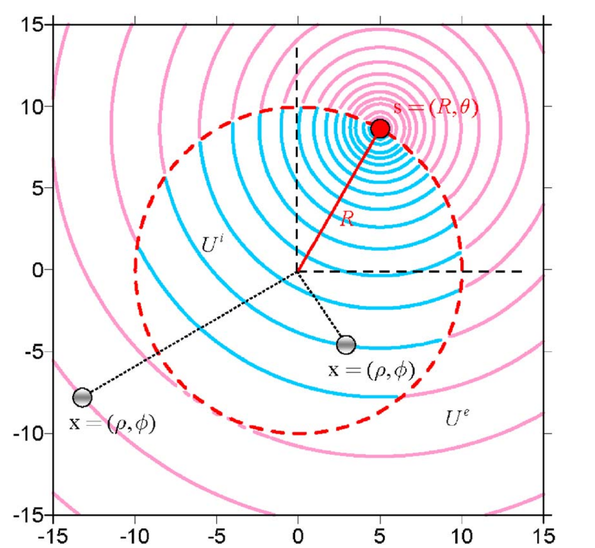


Fig. 2 Graph of the degenerate kernel for the fundamental solution, $s=(10,\pi/3)$

$$L(s,x)=\begin{cases} L^i(R,\theta;\rho,\phi)=-\sum_{m=1}^{\infty}\left(\frac{\rho^{m-1}}{R^m}\right)\cos m(\theta-\phi), & R>\rho \\ L^e(R,\theta;\rho,\phi)=\frac{1}{\rho}+\sum_{m=1}^{\infty}\left(\frac{R^m}{\rho^{m+1}}\right)\cos m(\theta-\phi), & \rho>R \end{cases}, \tag{14}$$

$$M(s,x)=\begin{cases} M^i(R,\theta;\rho,\phi)=\sum_{m=1}^{\infty}\left(\frac{m\rho^{m-1}}{R^{m+1}}\right)\cos m(\theta-\phi), & R\geqslant\rho \\ M^e(R,\theta;\rho,\phi)=\sum_{m=1}^{\infty}\left(\frac{mR^{m-1}}{\rho^{m+1}}\right)\cos m(\theta-\phi), & \rho>R \end{cases}. \tag{15}$$

Since the potentials resulted from $T(s,x)$ and $L(s,x)$ kernels are discontinuous across the boundary, the potentials of $T(s,x)$ and $L(s,x)$ for $R\to\rho^+$ and $R\to\rho^-$ are different. This is the reason why $R=\rho$ is not included for degenerate kernels of $T(s,x)$ and $L(s,x)$ in Eqs. (13) and (14). For problems with circular boundaries, we apply the Fourier series expansions to approximate the potential w and its normal derivative t on the boundary as

$$w(s_k)=a_0^k+\sum_{n=1}^{L}(a_n^k\cos n\theta_k+b_n^k\sin n\theta_k),$$

$$s_k\in B_k,\quad k=0,1,2,\ldots,N, \tag{16}$$

$$t(s_k)=p_0^k+\sum_{n=1}^{L}(p_n^k\cos n\theta_k+q_n^k\sin n\theta_k),$$

$$s_k\in B_k,\quad k=0,1,2,\ldots,N, \tag{17}$$

where $t(s_k)=\partial w(s_k)/\partial n_s$, a_n^k, b_n^k, p_n^k and q_n^k $(n=0,1,2,\ldots L)$ are the Fourier coefficients and θ_k is the polar angle. In the real computation, only $2L+1$ finite terms are considered where L indicates the truncated terms of Fourier series.

3.2 Adaptive Observer System [18,19]. By using the collocation method, the null-field integral equation becomes a set of algebraic equations for the Fourier coefficients. To ensure the stability of the algebraic equations, one has to choose collocating points throughout all the circular boundaries of the inclusions. Since the boundary integral equation is derived from the reciprocal theorem of energy concept. Therefore, the boundary integral equation is frame indifferent due to the objectivity rule. This is the reason why the observer system is adaptively to locate the origin at the center of circle in the boundary integration. The adaptive observer system is chosen to fully employ the property of degenerate kernels. Figures 3(a) and 3(b) show the boundary integration for the circular boundary in the adaptive observer system. It is worth noting that the origin of the observer system is located on the center of the corresponding circle under integration to entirely utilize the geometry of circular boundary for the expansion of degenerate kernels and boundary densities. The dummy variable in the circular integration is the angle (θ) instead of the radial coordinate (R).

3.3 Linear Algebraic System. By moving the null-field point x_m to the kth circular boundary in the limit sense for Eq. (10) in Fig. 3(a), we have

$$\begin{aligned}0=&\sum_{k=0}^{N}\int_{B_k}T(R_k,\theta_k;\rho_m,\phi_m)w(R_k,\theta_k)R_k d\theta_k\\ &-\sum_{k=0}^{N}\int_{B_k}U(R_k,\theta_k;\rho_m,\phi_m)t(R_k,\theta_k)R_k d\theta_k,\\ &x(\rho_m,\phi_m)\in D^c,\end{aligned} \tag{18}$$

where N is the number of circular inclusions and B_0 denotes the outer boundary for the bounded domain. In case of the infinite problem, B_0 becomes B_∞. Note that the kernels $U(s,x)$ and $T(s,x)$ are assumed in the degenerate form given by Eqs. (12) and (13), respectively, while the boundary densities w and t are expressed in terms of the Fourier series expansion forms given by Eqs. (16) and (17), respectively. Then, the integrals multiplied by separate expansion coefficients in Eq. (18) are nonsingular and the limit of the null-field point to the boundary is easily implemented by using appropriate forms of degenerate kernels. Through such an idea, all the singular and hypersingular integrals are well captured. Thus, the collocation point $x(\rho_m,\phi_m)$ in the discretized Eq. (18) can be considered on the boundary B_k, as well as the null-field point. Along each circular boundary, $2L+1$ collocation points are required to match $2L+1$ terms of Fourier series for constructing a square influence matrix with the dimension of $2L+1$ by $2L+1$. In contrast to the standard discretized boundary integral equation formulation with nodal unknowns of the physical boundary densities w and t. Now the degrees of freedom are transformed to Fourier coefficients employed in expansion of boundary densities. It is found that the compatible relationship of the boundary unknowns is equivalent by moving either the null-field point or the domain point to the boundary in different directions using various degenerate kernels as shown in Figs. 3(a) and 3(b). In the B_k integration, we set the origin of the observer system to collocate at the center c_k to fully utilize the degenerate kernels and Fourier series. By collocating the null-field point on the boundary, the linear algebraic system is obtained.

For the exterior problem of matrix, we have

$$[\mathbf{U}^M]\{\mathbf{t}^M-\mathbf{t}^\infty\}=[\mathbf{T}^M]\{\mathbf{w}^M-\mathbf{w}^\infty\}. \tag{19}$$

For the interior problem of each inclusion, we have

$$[\mathbf{U}^I]\{\mathbf{t}^I\}=[\mathbf{T}^I]\{\mathbf{w}^I\}, \tag{20}$$

where the superscripts "M" and "I" denote the matrix and inclusion, respectively. $[\mathbf{U}^M]$, $[\mathbf{T}^M]$, $[\mathbf{U}^I]$, and $[\mathbf{T}^I]$ are the influence matrices with a dimension of $(N+1)(2L+1)$ by $(N+1)(2L+1)$, $\{\mathbf{w}^M\}$, $\{\mathbf{t}^M\}$, $\{\mathbf{w}^\infty\}$, $\{\mathbf{t}^\infty\}$, $\{\mathbf{w}^I\}$, and $\{\mathbf{t}^I\}$ denote the column vectors of Fourier coefficients with a dimension of $(N+1)(2L+1)$ by 1 in which those are defined as follows:

$$[\mathbf{U}^M]=\begin{bmatrix}\mathbf{U}_{00}^M & \mathbf{U}_{01}^M & \cdots & \mathbf{U}_{0N}^M\\ \mathbf{U}_{10}^M & \mathbf{U}_{11}^M & \cdots & \mathbf{U}_{1N}^M\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{U}_{N0}^M & \mathbf{U}_{N1}^M & \cdots & \mathbf{U}_{NN}^M\end{bmatrix},$$

$$[\mathbf{T}^M]=\begin{bmatrix}\mathbf{T}_{00}^M & \mathbf{T}_{01}^M & \cdots & \mathbf{T}_{0N}^M\\ \mathbf{T}_{10}^M & \mathbf{T}_{11}^M & \cdots & \mathbf{T}_{1N}^M\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{T}_{N0}^M & \mathbf{T}_{N1}^M & \cdots & \mathbf{T}_{NN}^M\end{bmatrix}, \tag{21}$$

$$[\mathbf{U}^I]=\begin{bmatrix}\mathbf{U}_{00}^I & \mathbf{U}_{01}^I & \cdots & \mathbf{U}_{0N}^I\\ \mathbf{U}_{10}^I & \mathbf{U}_{11}^I & \cdots & \mathbf{U}_{1N}^I\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{U}_{N0}^I & \mathbf{U}_{N1}^I & \cdots & \mathbf{U}_{NN}^I\end{bmatrix},\quad [\mathbf{T}^I]=\begin{bmatrix}\mathbf{T}_{00}^I & \mathbf{T}_{01}^I & \cdots & \mathbf{T}_{0N}^I\\ \mathbf{T}_{10}^I & \mathbf{T}_{11}^I & \cdots & \mathbf{T}_{1N}^I\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{T}_{N0}^I & \mathbf{T}_{N1}^I & \cdots & \mathbf{T}_{NN}^I\end{bmatrix}, \tag{22}$$

$$\{\mathbf{w}^M\}=\begin{Bmatrix}\mathbf{w}_0^M\\ \mathbf{w}_1^M\\ \mathbf{w}_2^M\\ \vdots\\ \mathbf{w}_N^M\end{Bmatrix},\quad \{\mathbf{t}^M\}=\begin{Bmatrix}\mathbf{t}_0^M\\ \mathbf{t}_1^M\\ \mathbf{t}_2^M\\ \vdots\\ \mathbf{t}_N^M\end{Bmatrix}, \tag{23}$$

$$\{\mathbf{w}^\infty\}=\begin{Bmatrix}\mathbf{w}_0^\infty\\ \mathbf{w}_1^\infty\\ \mathbf{w}_2^\infty\\ \vdots\\ \mathbf{w}_N^\infty\end{Bmatrix},\quad \{\mathbf{t}^\infty\}=\begin{Bmatrix}\mathbf{t}_0^\infty\\ \mathbf{t}_1^\infty\\ \mathbf{t}_2^\infty\\ \vdots\\ \mathbf{t}_N^\infty\end{Bmatrix}, \tag{24}$$

$$\{\mathbf{w}^I\}=\begin{Bmatrix}\mathbf{w}_0^I\\ \mathbf{w}_1^I\\ \mathbf{w}_2^I\\ \vdots\\ \mathbf{w}_N^I\end{Bmatrix},\quad \{\mathbf{t}^I\}=\begin{Bmatrix}\mathbf{t}_0^I\\ \mathbf{t}_1^I\\ \mathbf{t}_2^I\\ \vdots\\ \mathbf{t}_N^I\end{Bmatrix}, \tag{25}$$

where $\{\mathbf{w}^M\}$, $\{\mathbf{t}^M\}$, $\{\mathbf{w}^\infty\}$, $\{\mathbf{t}^\infty\}$, $\{\mathbf{w}^I\}$, and $\{\mathbf{t}^I\}$ are the vectors of Fourier coefficients and the first subscript "j" ($j=0,1,2,\cdots,N$) in $[\mathbf{U}_{jk}^M]$, $[\mathbf{T}_{jk}^M]$, $[\mathbf{U}_{jk}^I]$, and $[\mathbf{T}_{jk}^I]$ denotes the index of the jth circle where the collocation point is located and the second subscript "k" ($k=0,1,2,\ldots,N$) denotes the index of the kth circle when integrating on each boundary data $\{\mathbf{w}_k^M-\mathbf{w}_k^\infty\}$, $\{\mathbf{t}_k^M-\mathbf{t}_k^\infty\}$, $\{\mathbf{w}_k^I\}$, and $\{\mathbf{t}_k^I\}$, N is the number of circular inclusions in the domain and the number L indicates the truncated terms of Fourier series. It is noted that $\{\mathbf{w}^\infty\}$ and $\{\mathbf{t}^\infty\}$ in Fig. 1(d) are the displacement and traction due to the remote shear. The coefficient matrix of the linear algebraic system is partitioned into blocks, and each off-diagonal block corresponds to the influence matrices between two different circular boundaries. The diagonal blocks are the influence matrices due to itself in each individual circle. After uniformly collocating the point along the kth circular boundary, the submatrix can be written as

$$[\mathbf{U}_{jk}^M]=\begin{bmatrix} U_{jk}^{0c}(\phi_1) & U_{jk}^{1c}(\phi_1) & U_{jk}^{1s}(\phi_1) & \cdots & U_{jk}^{Lc}(\phi_1) & U_{jk}^{Ls}(\phi_1)\\ U_{jk}^{0c}(\phi_2) & U_{jk}^{1c}(\phi_2) & U_{jk}^{1s}(\phi_2) & \cdots & U_{jk}^{Lc}(\phi_2) & U_{jk}^{Ls}(\phi_2)\\ U_{jk}^{0c}(\phi_3) & U_{jk}^{1c}(\phi_3) & U_{jk}^{1s}(\phi_3) & \cdots & U_{jk}^{Mc}(\phi_3) & U_{jk}^{Ls}(\phi_3)\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ U_{jk}^{0c}(\phi_{2L}) & U_{jk}^{1c}(\phi_{2L}) & U_{jk}^{1s}(\phi_{2L}) & \cdots & U_{jk}^{Lc}(\phi_{2L}) & U_{jk}^{Ls}(\phi_{2L})\\ U_{jk}^{0c}(\phi_{2L+1}) & U_{jk}^{1c}(\phi_{2L+1}) & U_{jk}^{1s}(\phi_{2L+1}) & \cdots & U_{jk}^{Lc}(\phi_{2L+1}) & U_{jk}^{Ls}(\phi_{2L+1})\end{bmatrix}, \tag{26}$$

$$[\mathbf{T}_{jk}^M]=\begin{bmatrix} T_{jk}^{0c}(\phi_1) & T_{jk}^{1c}(\phi_1) & T_{jk}^{1s}(\phi_1) & \cdots & T_{jk}^{Lc}(\phi_1) & T_{jk}^{Ls}(\phi_1)\\ T_{jk}^{0c}(\phi_2) & T_{jk}^{1c}(\phi_2) & T_{jk}^{1s}(\phi_2) & \cdots & T_{jk}^{Lc}(\phi_2) & T_{jk}^{Ls}(\phi_2)\\ T_{jk}^{0c}(\phi_3) & T_{jk}^{1c}(\phi_3) & T_{jk}^{1s}(\phi_3) & \cdots & T_{jk}^{Lc}(\phi_3) & T_{jk}^{Ls}(\phi_3)\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ T_{jk}^{0c}(\phi_{2L}) & T_{jk}^{1c}(\phi_{2L}) & T_{jk}^{1s}(\phi_{2L}) & \cdots & T_{jk}^{Lc}(\phi_{2L}) & T_{jk}^{Ls}(\phi_{2L})\\ T_{jk}^{0c}(\phi_{2L+1}) & T_{jk}^{1c}(\phi_{2L+1}) & T_{jk}^{1s}(\phi_{2L+1}) & \cdots & T_{jk}^{Lc}(\phi_{2L+1}) & T_{jk}^{Ls}(\phi_{2L+1})\end{bmatrix}, \tag{27}$$

where ϕ_m, $m=1,2,\ldots,2L+1$, is the angle of collocation point along the circular boundary. Although both the matrices in Eqs. (26) and (27) are not sparse, it is found that the higher order harmonics are considered, the lower influence coefficients are obtained in numerical experiments. It is noted that the superscript " $0s$" in Eqs. (26) and (27) disappears since $\sin n\theta=0$ ($n=0$). The element of $[\mathbf{U}_{jk}^M]$ and $[\mathbf{T}_{jk}^M]$ are defined, respectively, as

$$U_{jk}^{nc}(\phi_m)=\int_{B_k} U(s_k,x_m)\cos(n\theta_k)R_k d\theta_k,\quad n=0,1,2,\ldots,L,$$

$$m=1,2,\ldots,2L+1, \tag{28}$$

$$U_{jk}^{ns}(\phi_m)=\int_{B_k} U(s_k,x_m)\sin(n\theta_k)R_k d\theta_k,$$

$$n=1,2,\ldots,L,\quad m=1,2,\ldots,2L+1, \tag{29}$$

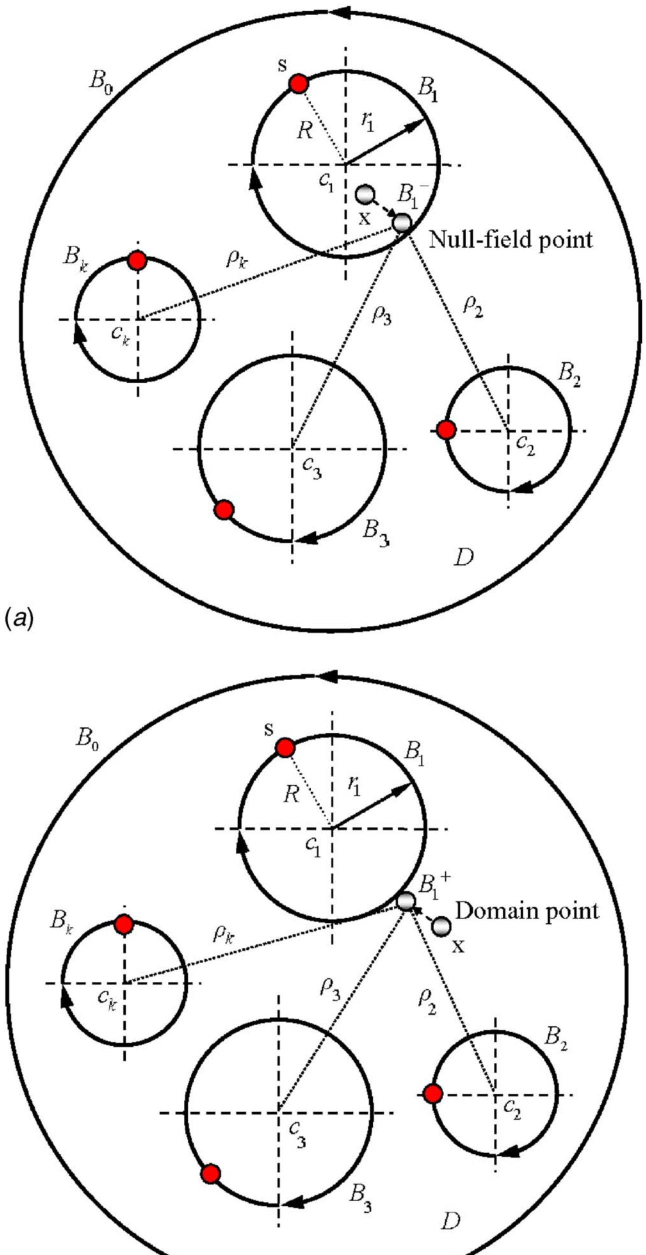
$$T_{jk}^{ns}(\phi_m)=\int_{B_k} T(s_k,x_m)\cos(n\theta_k)R_k d\theta_k,$$

$$n=0,1,2,\ldots,L,\quad m=1,2,\ldots,2L+1, \tag{30}$$

$$T_{jk}^{ns}(\phi_m)=\int_{B_k} T(s_k,x_m)\sin(n\theta_k)R_k d\theta_k,$$

$$n=1,2,\ldots,L,\quad m=1,2,\ldots,2L+1, \tag{31}$$

where k is no sum, $s_k=(R_k,\theta_k)$, and ϕ_m is the angle of collocation point x_m along the boundary. The submatrix $[\mathbf{U}_{jk}^I]$ and $[\mathbf{T}_{jk}^I]$ can be written in a similar way. Equation (18) can be calculated by employing the orthogonal property of trigonometric function in the real computation. Only the finite L terms are used in the summation of Eqs. (16) and (17). The explicit forms of all the boundary integrals for U, T, L, and M kernels are listed in the Table 1. Finite



(*a*)

(*b*)

Fig. 3 (*a*) Sketch of the null-field integral equation for a null-field point in conjunction with the adaptive observer system ($x \notin D, x \to B_k$) and (*b*) sketch of the boundary integral equation for a domain point in conjunction with the adaptive observer system ($x \in D, x \to B_k$)

values of singular and hypersingular integrals are well captured after introducing the degenerate kernel. Besides, the limiting case across the boundary ($R^- \to \rho \to R^+$) is also addressed. The continuous and jump behavior across the boundary is well described. Instead of boundary data in the BEM, the Fourier coefficients become the new unknown degrees of freedom in the formulation. Two cases may be solved in a unified manner using the null-field integral formulation:

(1) One bounded problem of the circular domain in Fig. 1(*c*) becomes the interior problem for each inclusion.
(2) The other is unbounded, i.e., the outer boundary B_0 in Fig. 3(*a*) is B_∞. It is the exterior problem for the matrix as shown in Fig. 1(*e*).

The direction of contour integration should be taken care, i.e., counterclockwise and clockwise directions are for the interior and exterior problems, respectively.

3.4 Match of Interface Conditions. According to the continuity of displacement and equilibrium of traction along the kth interface, we have the two constraints

$$\{\mathbf{w}^M\} = \{\mathbf{w}^I\} \quad \text{on } B_k, \tag{32}$$

$$[\boldsymbol{\mu}_0]\{\mathbf{t}^M\} = -[\boldsymbol{\mu}_\mathbf{k}]\{\mathbf{t}^I\} \quad \text{on } B_k, \tag{33}$$

where $[\boldsymbol{\mu}_0]$ and $[\boldsymbol{\mu}_\mathbf{k}]$ are defined as follows:

$$[\boldsymbol{\mu}_0] = \begin{bmatrix} \mu_0 & 0 & \cdots & 0 \\ 0 & \mu_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_0 \end{bmatrix}, \quad [\boldsymbol{\mu}_\mathbf{k}] = \begin{bmatrix} \mu_k & 0 & \cdots & 0 \\ 0 & \mu_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_k \end{bmatrix}, \tag{34}$$

where μ_0 and μ_k denote the shear modulus of the matrix and the kth inclusion, respectively. By assembling the matrices in Eqs. (19), (20), (32), and (33), we have

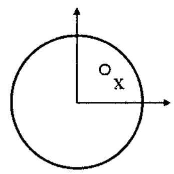
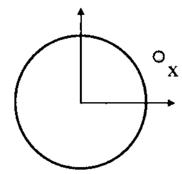
$$\begin{bmatrix} \mathbf{T}^M & -\mathbf{U}^M & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}^I & -\mathbf{U}^I \\ \mathbf{I} & \mathbf{0} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mu}_\mathbf{0} & \mathbf{0} & \boldsymbol{\mu}_\mathbf{k} \end{bmatrix} \begin{Bmatrix} \mathbf{w}^M \\ \mathbf{t}^M \\ \mathbf{w}^I \\ \mathbf{t}^I \end{Bmatrix} = \begin{Bmatrix} \mathbf{a} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{Bmatrix}, \tag{35}$$

where $\{\mathbf{a}\}$ is the forcing term due to the remote shear stress and $[\mathbf{I}]$ is the identity matrix. The calculation for the vector $\{\mathbf{a}\}$ is elaborated on later in Appendix A. After obtaining the unknown Fourier coefficients in Eq. (35), the origin of observer system is set to c_k in the B_k integration as shown in Fig. 3(*b*) to obtain the field potential by employing Eq. (6). The differences between the present formulation and the conventional BEM are listed in Table 2.

3.5 Vector Decomposition Technique for the Potential Gradient in the Hypersingular Equation. In order to determine the stress field, the tangential derivative should be calculated with care. Also Eq. (7) shows the normal derivative of potential for domain points. For the nonconcentric cases, special treatment for the potential gradient should be considered as the source point and field point locate on different circular boundaries. As shown in Fig. 4, the normal direction on the boundary (1, 1′) should be superimposed by those of the radial derivative (3, 3′) and angular derivative (4, 4′) through the vector decomposition technique. According to the concept of vector decomposition technique, the kernel functions of Eqs. (14) and (15) can be modified to

$$L(s,x) = \begin{cases} L^i(R,\theta;\rho,\phi) = -\sum_{m=1}^{\infty}\left(\frac{\rho^{m-1}}{R^m}\right)\cos m(\theta-\phi)\cos(\zeta-\xi) - \sum_{m=1}^{\infty}\left(\frac{\rho^{m-1}}{R^m}\right)\sin m(\theta-\phi)\cos\left(\frac{\pi}{2}-\zeta+\xi\right), & R > \rho \\ L^e(R,\theta;\rho,\phi) = \frac{1}{\rho} + \sum_{m=1}^{\infty}\left(\frac{R^m}{\rho^{m+1}}\right)\cos m(\theta-\phi)\cos(\zeta-\xi) - \sum_{m=1}^{\infty}\left(\frac{R^m}{\rho^{m+1}}\right)\sin m(\theta-\phi)\cos\left(\frac{\pi}{2}-\zeta+\xi\right), & \rho > R \end{cases}, \tag{36}$$

Table 1 Influence coefficients for the singularity distribution on the circular boundary

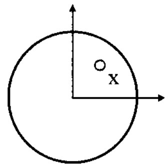
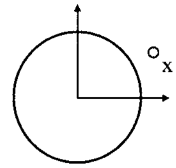
		 $\rho < R$		 $\rho > R$
	Degenerate kernel	$U^i(R,\theta;\rho,\phi) = \ln R - \sum_{m=1}^{\infty} \frac{1}{m}(\frac{\rho}{R})^m \cos m(\theta-\phi),\ R \geq \rho$		$U^e(R,\theta;\rho,\phi) = \ln \rho - \sum_{m=1}^{\infty} \frac{1}{m}(\frac{R}{\rho})^m \cos m(\theta-\phi),\ \rho > R$
$U(\mathrm{s},\mathrm{x})$	Orthogonal process	$\int_0^{2\pi} [U^i][1] R d\theta = 2\pi \ln R,\ R \geq \rho$ $\int_0^{2\pi} [U^i][\cos n\theta] R d\theta = \pi \frac{1}{n} \frac{\rho^n}{R^{n-1}} \cos n\phi,\ R \geq \rho$ $\int_0^{2\pi} [U^i][\sin n\theta] R d\theta = \pi \frac{1}{n} \frac{\rho^n}{R^{n-1}} \sin n\phi,\ R \geq \rho$		$\int_0^{2\pi} [U^e][1] R d\theta = 2\pi \ln \rho,\ \rho > R$ $\int_0^{2\pi} [U^e][\cos n\theta] R d\theta = \pi \frac{1}{n} \frac{R^{n+1}}{\rho^n} \cos n\phi,\ \rho > R$ $\int_0^{2\pi} [U^e][\sin n\theta] R d\theta = \pi \frac{1}{n} \frac{R^{n+1}}{\rho^n} \sin n\phi,\ \rho > R$
	Limiting behavior across the boundary	$\lim_{\rho \to R} 2\pi \ln R = 2\pi \ln R$ $\lim_{\rho \to R} \pi \frac{1}{n} \frac{\rho^n}{R^{n-1}} \cos n\phi = \pi \frac{1}{n} R \cos n\phi$ $\lim_{\rho \to R} \pi \frac{1}{n} \frac{\rho^n}{R^{n-1}} \sin n\phi = \pi \frac{1}{n} R \sin n\phi$	Continuous $(R^- \to R \to R^+)$	$\lim_{\rho \to R} 2\pi \ln \rho = 2\pi \ln R$ $\lim_{\rho \to R} \pi \frac{1}{n} \frac{R^{n+1}}{\rho^n} \cos n\phi = \pi \frac{1}{n} R \cos n\phi$ $\lim_{\rho \to R} \pi \frac{1}{n} \frac{R^{n+1}}{\rho^n} \sin n\phi = \pi \frac{1}{n} R \sin n\phi$
	Degenerate kernel	$T^i(R,\theta;\rho,\phi) = \frac{1}{R} + \sum_{m=1}^{\infty} (\frac{\rho^m}{R^{m+1}}) \cos m(\theta-\phi),\ \ R > \rho$		$T^e(R,\theta;\rho,\phi) = -\sum_{m=1}^{\infty} (\frac{R^{m-1}}{\rho^m}) \cos m(\theta-\phi),\ \ \rho > R$
$T(\mathrm{s},\mathrm{x})$	Orthogonal process	$\int_0^{2\pi} [T^i][1] R d\theta = \frac{2\pi}{R},\ R > \rho$ $\int_0^{2\pi} [T^i][\cos n\theta] R d\theta = \pi (\frac{\rho}{R})^n \cos n\phi,\ R > \rho$ $\int_0^{2\pi} [T^i][\sin n\theta] R d\theta = \pi (\frac{\rho}{R})^n \sin n\phi,\ R > \rho$		$\int_0^{2\pi} [T^e][1] R d\theta = 0,\ \rho > R$ $\int_0^{2\pi} [T^e][\cos n\theta] R d\theta = -\pi (\frac{R}{\rho})^n \cos n\phi,\ \rho > R$ $\int_0^{2\pi} [T^e][\sin n\theta] R d\theta = -\pi (\frac{R}{\rho})^n \sin n\phi,\ \rho > R$
	Limiting behavior across the boundary	$\lim_{\rho \to R} \frac{2\pi}{R} = \frac{2\pi}{R}$ $\lim_{\rho \to R} \pi (\frac{\rho}{R})^n \cos n\phi = \pi \cos n\phi$ $\lim_{\rho \to R} \pi (\frac{\rho}{R})^n \sin n\phi = \pi \sin n\phi$	Jump $(R^- \to R \to R^+)$	$\lim_{\rho \to R} 0 = 0$ $\lim_{\rho \to R} -\pi (\frac{R}{\rho})^n \cos n\phi = -\pi \cos n\phi$ $\lim_{\rho \to R} -\pi (\frac{R}{\rho})^n \sin n\phi = -\pi \sin n\phi$

$$M(s,x) = \begin{cases} M^i(R,\theta;\rho,\phi) = \sum_{m=1}^{\infty} \left(\frac{m\rho^{m-1}}{R^{m+1}}\right) \cos m(\theta-\phi) \cos(\zeta-\xi) - \sum_{m=1}^{\infty} \left(\frac{m\rho^{m-1}}{R^{m+1}}\right) \sin m(\theta-\phi) \cos\left(\frac{\pi}{2} - \zeta + \xi\right), & R \geqslant \rho \\ M^e(R,\theta;\rho,\phi) = \sum_{m=1}^{\infty} \left(\frac{mR^{m-1}}{\rho^{m+1}}\right) \cos m(\theta-\phi) \cos(\zeta-\xi) - \sum_{m=1}^{\infty} \left(\frac{mR^{m-1}}{\rho^{m+1}}\right) \sin m(\theta-\phi) \cos\left(\frac{\pi}{2} - \zeta + \xi\right), & \rho > R \end{cases} \tag{37}$$

where ζ and ξ are shown in Fig. 4. For the special case of confocal, the potential gradient is derived free of special treatment since $\zeta = \xi$.

3.6 Stresses Described in the Polar Coordinate. After obtaining all the unknown Fourier coefficients of w and t for the matrix and inclusions, the stress described in the polar coordinate

Table 1 (***Continued.***)

		 $\rho < R$		 $\rho > R$
$L(\mathrm{s},\mathrm{x})$	Degenerate kernel	$L^i(R,\theta;\rho,\phi) = -\sum_{m=1}^{\infty}(\frac{\rho^{m-1}}{R^m})\cos m(\theta-\phi),\ R>\rho$		$L^e(R,\theta;\rho,\phi) = \frac{1}{\rho}+\sum_{m=1}^{\infty}(\frac{R^m}{\rho^{m+1}})\cos m(\theta-\phi),\ \rho>R$
	Orthogonal process	$\int_0^{2\pi}[L^i][1]Rd\theta = 0,\ R>\rho$ $\int_0^{2\pi}[L^i][\cos n\theta]Rd\theta = -\pi(\frac{\rho}{R})^{n-1}\cos n\phi,\ R>\rho$ $\int_0^{2\pi}[L^i][\sin n\theta]Rd\theta = -\pi(\frac{\rho}{R})^{n-1}\sin n\phi,\ R>\rho$		$\int_0^{2\pi}[L^e][1]Rd\theta = \frac{2\pi}{\rho},\ \rho>R$ $\int_0^{2\pi}[L^e][\cos n\theta]Rd\theta = \pi(\frac{R}{\rho})^{n+1}\cos n\phi,\ \rho>R$ $\int_0^{2\pi}[L^e][\sin n\theta]Rd\theta = \pi(\frac{R}{\rho})^{n+1}\sin n\phi,\ \rho>R$
	Limiting behavior across the boundary	$\lim_{\rho\to R} 0 = 0$ $\lim_{\rho\to R} -\pi(\frac{\rho}{R})^{n-1}\cos n\phi = -\pi\cos n\phi$ $\lim_{\rho\to R} -\pi(\frac{\rho}{R})^{n-1}\sin n\phi = -\pi\sin n\phi$	Jump $(R^- \to R \to R^+)$	$\lim_{\rho\to R}\frac{2\pi}{\rho} = \frac{2\pi}{R}$ $\lim_{\rho\to R}\pi(\frac{R}{\rho})^{n+1}\cos n\phi = \pi\cos n\phi$ $\lim_{\rho\to R}\pi(\frac{R}{\rho})^{n+1}\sin n\phi = \pi\sin n\phi$
$M(\mathrm{s},\mathrm{x})$	Degenerate kernel	$M^i(R,\theta;\rho,\phi) = \sum_{m=1}^{\infty}(\frac{m\rho^{m-1}}{R^{m+1}})\cos m(\theta-\phi),\ R\geq\rho$		$M^e(R,\theta;\rho,\phi) = \sum_{m=1}^{\infty}(\frac{mR^{m-1}}{\rho^{m+1}})\cos m(\theta-\phi),\ \rho>R$
	Orthogonal process	$\int_0^{2\pi}[M^i][1]Rd\theta = 0,\ R\geq\rho$ $\int_0^{2\pi}[M^i][\cos n\theta]Rd\theta = n\pi\frac{\rho^{n-1}}{R^n}\cos n\phi,\ R\geq\rho$ $\int_0^{2\pi}[M^i][\sin n\theta]Rd\theta = n\pi\frac{\rho^{n-1}}{R^n}\sin n\phi,\ R\geq\rho$		$\int_0^{2\pi}[M^e][1]Rd\theta = 0,\ \rho>R$ $\int_0^{2\pi}[M^e][\cos n\theta]Rd\theta = n\pi\frac{R^n}{\rho^{n+1}}\cos n\phi,\ \rho>R$ $\int_0^{2\pi}[M^e][\sin n\theta]Rd\theta = n\pi\frac{R^n}{\rho^{n+1}}\sin n\phi,\ \rho>R$
	Limiting behavior across the boundary	$\lim_{\rho\to R} 0 = 0$ $\lim_{\rho\to R} n\pi\frac{\rho^{n-1}}{R^n}\cos n\phi = n\pi\frac{1}{R}\cos n\phi$ $\lim_{\rho\to R} n\pi\frac{\rho^{n-1}}{R^n}\sin n\phi = n\pi\frac{1}{R}\sin n\phi$	Continuous $(R^- \to R \to R^+)$	$\lim_{\rho\to R} 0 = 0$ $\lim_{\rho\to R} n\pi\frac{R^n}{\rho^{n+1}}\cos n\phi = n\pi\frac{1}{R}\cos n\phi$ $\lim_{\rho\to R} n\pi\frac{R^n}{\rho^{n+1}}\sin n\phi = n\pi\frac{1}{R}\sin n\phi$

can be determined by

$$\sigma_{zr} = \sigma_{zx}\cos\phi + \sigma_{zy}\sin\phi, \tag{38}$$

$$\sigma_{z\theta} = -\sigma_{zx}\sin\phi + \sigma_{zy}\cos\phi, \tag{39}$$

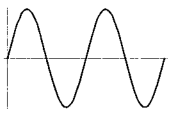
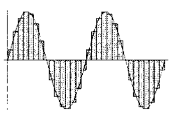
where σ_{zr} and $\sigma_{z\theta}$ are the normal and tangential stresses, respectively. The boundary integral equation for the domain point including the boundary point instead of the null-field formulation is employed to find the stress by employing the appropriate form of degenerate kernels. The flowchart of the present method is shown in Table 3.

4 Numerical Results and Discussions

First, we derive an exact solution for a single inclusion using the present formulation in Appendix B. Symbolic software of MATHEMATICA is employed to solve a $2L+1$ by $2L+1$ sparse matrix by using the induction concept. Then, seven problems solved by previous scholars are revisited by using the present method to show the generality and validity of our formulation. Besides, we demonstrate the problem of interaction of two cavities in case 1 to compare the present method with the conventional BEM.

4.1 Case 1: Two Equal-Sized Holes Lie on the *x* Axis (a Limiting Case) [2,9]. Figure 5(a) shows the geometry of two equal-sized holes in the infinite medium under the remote shear

Table 2 Comparisons of the present method and conventional BEM

	Boundary density discretization	Auxiliary system	Formulation	Observer system	Singularity	Convergence	Boundary-layer effect
Present method	Fourier series 	Degenerate kernel	Null-field integral equation	Adaptive observer system	Disappear after introducing the degenerate kernel	Exponential convergence	Free
Conventional BEM	Constant element 	Fundamental solution	Boundary integral equation	Fixed observer system	Principal values ($C.P.V.$, $R.P.V.$ and $H.P.V.$)	Linear algebraic convergence	Appear

where $C.P.V.$, $R.P.V.$ and $H.P.V.$ are the Cauchy, Riemann and Hadamard principal values, respectively.

$\sigma_{zy}^{\infty}=\tau_{\infty}$. The stress concentration of the problem is illustrated in Fig. 5(b). It indicates that the present result agrees well with the analytical solution of Steif [2] and those obtained by Chao and Young [9] even though the two holes approach each other. Figure 5(c) shows that only few terms of Fourier series can obtain good results. However, more nodes are required by using the conventional BEM to achieve convergence. Our formulation is free of boundary-layer effect instead of appearance by using the conventional BEM when the stress $\sigma_{z\theta}$ near the boundary as shown in Fig. 5(d). Stress concentration factors and errors for various distances between two holes by using the present method and the conventional BEM are listed in Table 4. These results show that the present method is more accurate and effective than those of the conventional BEM. Under the same error tolerance, the CPU time of the present method is fewer than that of the conventional BEM. Besides, it is noted that more terms of Fourier series are required to capture the singular behavior when the two holes approach each other.

4.2 Case 2: Two Identical Inclusions Locating on the x Axis [3]. We consider two identical elastic inclusions of radii $r_1=r_2$ and shear moduli $\mu_1=\mu_2$ embedded in an infinite medium subjected to the remote shear $\sigma_{zx}^{\infty}=\tau_{\infty}$ at infinity as shown in Fig. 6(a). Figure 6(b) shows that stress concentrations diminish when the inclusion spacing increases. We note that the mathematical model of rigid-inclusion problem is equivalent to that of uniform potential flow past two parallel cylinders with no circulation around either cylinder. The remote shear $\sigma_{zx}^{\infty}=\tau_{\infty}$ is similar to the velocity V_{∞} in the x direction at infinity and the velocity field is similar to the stress field [22].

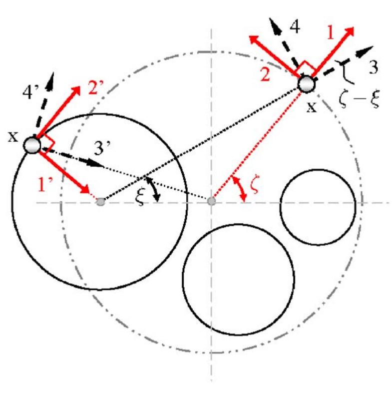


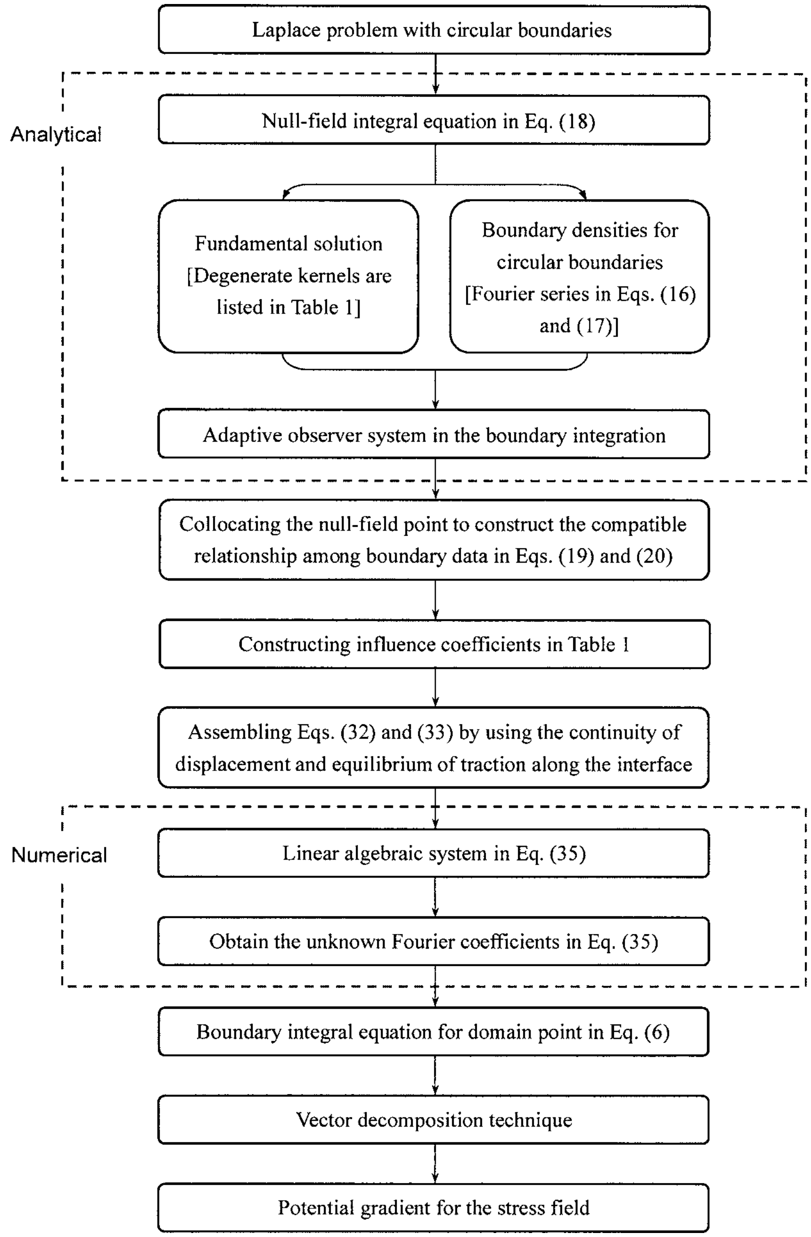
Fig. 4 Vector decomposition for the potential gradient in the hypersingular equation

4.3 Case 3: Two Circular Inclusions Locating on the x Axis [6]. Two inclusions with radii of r_1 and r_2 under the remote shear are considered as shown in Fig. 7(a). The stress distributions in the matrix including the radial component σ_{zr} and the tangential component $\sigma_{z\theta}$ around the circular boundary of radius r_1 are plotted in Figs. 7(b) and 7(c) for various inclusion spacings when the two inclusion radii are equal-sized ($r_1=r_2$). Two limiting cases are considered for rigid inclusions ($\mu_1/\mu_0=\mu_2/\mu_0=\infty$) and for cavities ($\mu_1/\mu_0=\mu_2/\mu_0=0.0$). It can be found that $\sigma_{z\theta}=0$ or $\sigma_{zr}=0$ for rigid inclusions or cavities as predicted for the single inclusion or cavity, respectively. Moreover, the nonzero stress components for these two cases are identical when the stress components at infinity are interchanged, i.e., the stresses around the circular boundary σ_{zr} in one case equals to $\sigma_{z\theta}$ for the other case due to the analogy of mathematical model. It can be seen from Figs. 7(b) and 7(c) that unbounded stresses apparently occur at $\theta=180$ deg under the condition of $\sigma_{zx}^{\infty}=\tau_{\infty}$ for rigid inclusions or $\sigma_{zy}^{\infty}=\tau_{\infty}$ for cavities when two inclusions approach closely or even touch each other. In Figs. 7(d) and 7(e), the variation of stresses around the circular boundary of radius r_1 is shown versus radius r_2 for a fixed separation of $d=0.1r_1$. More terms of Fourier series are required to capture the singular behavior when the two inclusions approach each other as well as the two radii of inclusions are quite different. The present numerical results match very well with those by Goree and Wilson [6].

4.4 Case 4: Two Circular Inclusions Locating on the y Axis [1]. The infinite medium with two elastic inclusions is under the uniform remote shear $\sigma_{zy}^{\infty}=\tau_{\infty}$. The first inclusion centered at the origin of radius r_1 with the shear modulus $\mu_1=2\mu_0/3$ and the other inclusion of radius $r_2=2r_1$ centered on y axis at r_1+r_2+d ($d=0.1r_1$) with the shear modulus $\mu_2=13\mu_0/7$ are shown in Fig. 8(a). In order to be compared with the Honein et al.'s data obtained by using the Möbius transformations [1], the stresses along the boundary of radius r_1 is shown in Fig. 8(b). It satisfies the equilibrium traction along the interface of circular boundary. The stress concentration factor reaches maximum at $\theta=0$ deg in the matrix. Figure 8(c) shows that only few terms of Fourier series can yield acceptable results. Figures 8(d) and 8(e) indicates that our formulation is free of boundary-layer effect since stresses σ_{zr} and $\sigma_{z\theta}$ near the boundary can be smoothly predicted, respectively. The key to eliminate the boundary-layer effect is that we introduce the degenerate kernel to describe the jump function for interior and exterior regions as shown in Table 1.

4.5 Case 5: Two Inclusions Located on the x Axis Under the Two-Direction Shear [8]. In Fig. 9(a), the parameters used in the calculation are taken as $r_1=r_2$, $\sigma_{zx}^{\infty}=\sigma_{zy}^{\infty}=\tau_{\infty}$, $\mu_0=0.185$, and $\mu_1=\mu_2=4.344$. Figure 9(b) shows stress distributions σ_{zx} and σ_{zy}

Table 3 Flowchart of the present method



along the x axis when $d=0.1$. It can be seen that the stress component σ_{zx} is continuous across the interface between two different materials and has a peak value between two inclusions. The stress component σ_{zy} is discontinuous across the interface of two different materials. Figures 9(c) and 9(d) illustrate stress distributions of σ_{zx} and σ_{zy} along the x axis when $d=0.4$ and $d=1.0$, respectively. Both figures indicate that stress components σ_{zx} and σ_{zy} have similar changing curves to those of Fig. 9(b). However, it should be noted that the maximum value of stress component σ_{zx} drops when the distance d between the two inclusions increases. Figure 9(e) illustrates the normal stress σ_{zr} distributions along the contour $(1.001, \theta)$ for various cases of $d=0.1$, 0.5, and 1.0. It shows that the shear stress σ_{zr} increases as the distance d between the two inclusions decreases at the point where two inclusions approach each other. However, the distance d has a slight effect on σ_{zr} when the angle is in the range of 90 deg$<\theta<$320 deg. Figure 9(f) illustrates the tangential stress $\sigma_{z\theta}$ distributions along the contour $(1.001, \theta)$ for various distances of $d=0.1$, 0.5, and 1.0. It should be noted that the absolute value of tangential stress $\sigma_{z\theta}$ is very small in comparison with that of σ_{zr}. Figure 9(g) illustrates the variation of stress components σ_{zx} and σ_{zy} in the matrix at the point (1.001,0 deg) versus the distance d between the two inclusions. From the figure, it can be seen that stress components σ_{zx} and σ_{zy} have higher values when the two inclusions approach each other. However, stress components σ_{zx} and σ_{zy} tend smoothly to the constant when the two inclusions separate away. Figure 9(h) shows stress distributions σ_{zx} and σ_{zy} along the x axis when the two inclusions touch each other. It can be seen that the shear stress σ_{zx} has a peak value at the touched point. For the increasing value of x, σ_{zx} tends to match the remote shear τ_∞. Besides, the stress component σ_{zy} is continuous at the tangent point $(x/r_1=1.0)$ and has a discontinuous jump on the interface between the matrix and inclusion $(x/r_1=3.0)$. The present results in Figs. 9(b)–9(h) agree very well with the Wu's data [8]. Only the stress component σ_{zx} at the touched point is lower than the Wu's data as shown in

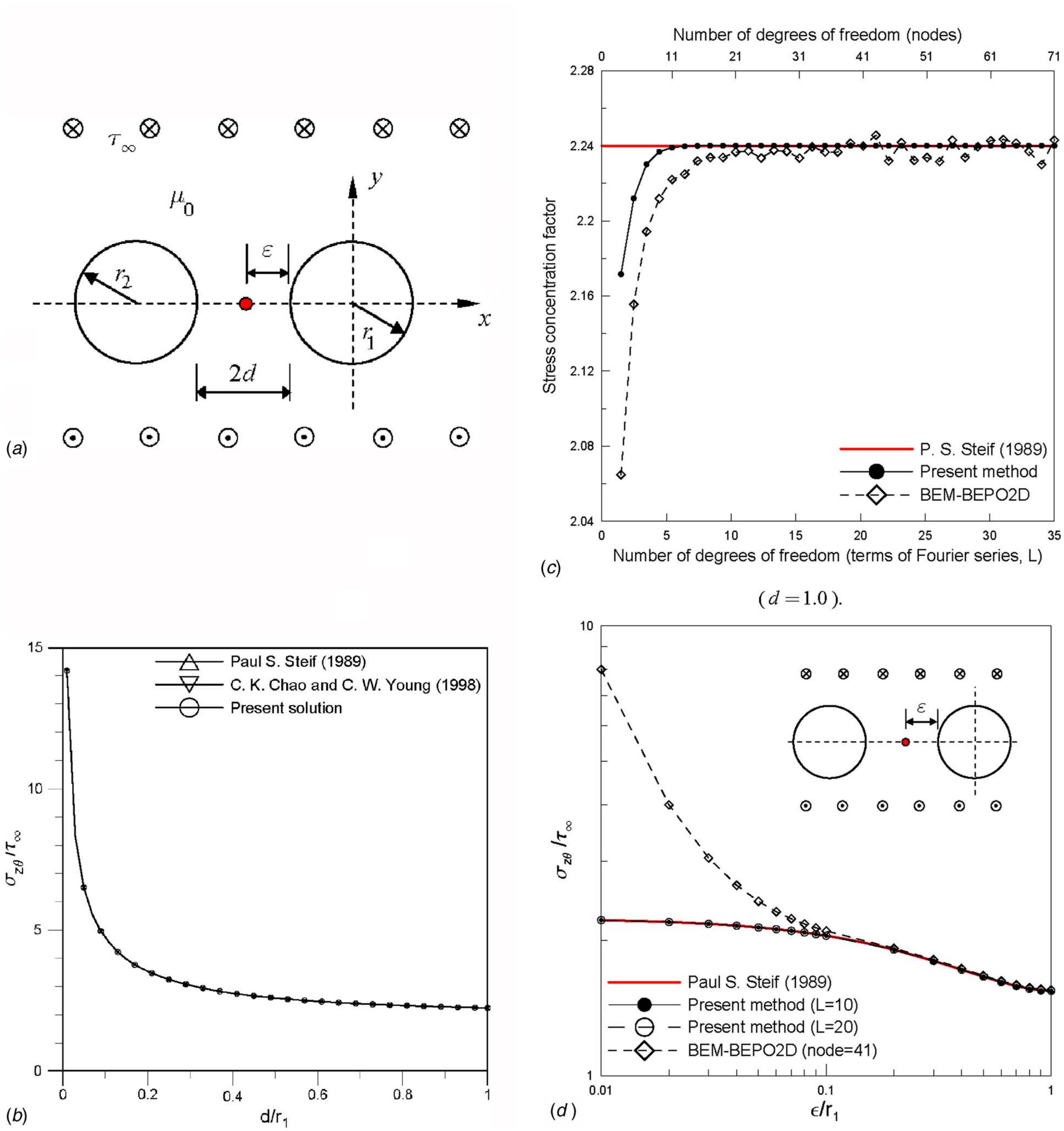


Fig. 5 (*a*) Two equal-sized holes ($r_1 = r_2$) with centers on the *x* axis, (*b*) stress concentration of the problem containing two equal-sized holes, (*c*) convergence test of the problem containing two equal-sized holes (d=1.0), and (*d*) tangential stress in the matrix near the boundary (d=1.0)

Fig. 9(*h*), since separate Fourier expansions are described for the touched inclusions in our formulation.

4.6 Case 6: One Hole Surrounded by Two Circular Inclusions [9]. Figure 10(*a*) shows that a circular hole centered at the origin of radius r_1 is surrounded by two circular inclusions ($d/r_1 = 1.0$) with equal radius $r_2 = r_3 = 2r_1$ and equal shear modulus $\mu_2 = \mu_3$ under the remote shear $\sigma_{zx}^{\infty} = \tau_{\infty}$. We solved the distribution of the tangential stress along the circular hole influenced by the surrounding inclusions when they are arrayed in parallel (β =0 deg) or perpendicular (β=90 deg) to the direction of uniform shear as shown in Figs. 10(*b*) and 10(*c*). It is found that, when a hole and two inclusions are arrayed parallel to the applied load (β=0 deg), the stress concentration factor, reaching maximum at θ=90 deg along a circular hole, increases (or decreases) as the neighboring hard (or soft) inclusions approach a circular hole as shown in Figs. 10(*b*) and 10(*d*). On the contrary, when a hole and two inclusions are perpendicular to the applied load (β=90 deg), the stress concentration factor, reaching maximum at θ=90 deg, decreases (or increases) as the neighboring hard (or soft) inclusions approach a circular hole as shown in Figs. 10(*c*) and 10(*e*). Our numerical results match very well with the Chao and Young's results [9].

4.7 Case 7: Three Identical Inclusions Forming an Equilateral Triangle [10]. Figure 11(*a*) shows that three identical inclusions ($r_1 = r_2 = r_3$) subjected to the uniform shear stress $\sigma_{zy}^{\infty} = \tau_{\infty}$ at infinity. The three inclusions form an equilateral triangle

Table 4 Stress concentration factors and errors for various distances between two holes using the present approach and BEM

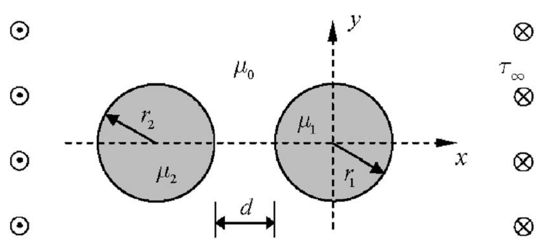
	d/r_1		0.01	0.2	0.4	0.6	0.8	1.0
Stress concentration factor	Analytical solution [2]		14.2247	3.5349	2.7667	2.4758	2.3274	2.2400
	Present method	$L=10$	10.5096 (26.12%)	3.5306 (0.12%)	2.7664 (0.01%)	2.4758 (0.00%)	2.3274 (0.00%)	2.2400 (0.00%)
		$L=20$	13.3275 (6.31%)	3.5349 (0.00%)	2.7667 (0.00%)	2.4758 (0.00%)	2.3274 (0.00%)	2.2400 (0.00%)
	BEM BEPO2D	$node=21$	7.2500 (49.03%)	3.4532 (2.31%)	2.738 (1.04%)	2.4639 (0.48%)	2.3168 (0.46%)	2.2366 (0.15%)
		$node=41$	10.2008 (28.29%)	3.5188 (0.46%)	2.7619 (0.17%)	2.4747 (0.04%)	2.3312 (0.16%)	2.2398 (0.01%)

Data in parentheses denote error.

and are placed at a distance $4r_1$ apart. Besides, the distance the hoop stress $\sigma_{z\theta}$ in the matrix around the boundary of the inclusion located at the origin as shown in Fig. 11(*b*). Good agreement is obtained between the Gong's results [10] and ours. It is obvious that the limiting case of circular holes ($\mu_1/\mu_0=\mu_2/\mu_0=\mu_3/\mu_0=0.0$) leads to the maximum stress concentration at $\theta=0$ deg, which is larger than 2 of a single hole due to the interaction effect. It is also interesting to note that the stress component $\sigma_{z\theta}$ vanishes in the case of rigid inclusions ($\mu_1/\mu_0=\mu_2/\mu_0=\mu_3/\mu_0=\infty$), which can be explained by a general analogy between solutions for traction-free holes and those involving rigid inclusions [2].

5 Conclusions

A semi-analytical formulation for multiple circular inclusions with arbitrary radii, moduli, and locations using degenerate kernels and Fourier series in the adaptive observer system was developed to ensure the exponential convergence. Generally speaking, only ten terms of Fourier series ($L=10$) can obtain the acceptable and accurate results. More terms of Fourier series are required to capture the singular behavior when the two inclusions approach each other as well as the two radii of inclusions are quite different. The singularity and hypersingularity were avoided after introducing the concept of degenerate kernels for interior and exterior regions. Besides, the boundary-layer effect for the stress calculation is eliminated since the degenerate kernel can describe the jump behavior for interior and exterior domains, respectively. The exact solution for a single inclusion was also rederived by using the present formulation. Several examples investigated by Steif [2], Budiansky and Carrier [3], Goree and Wilson [6], Honein et al. [1], Wu [8], Chao and Young [9], and Gong [10] were revisited, respectively. Good agreements were made after comparing with the previous results. Regardless of the number, size, and the position of circular inclusions and cavities, the proposed method can offer good results. Moreover, our method presented here can be applied to Laplace problems with circular boundaries, e.g.,



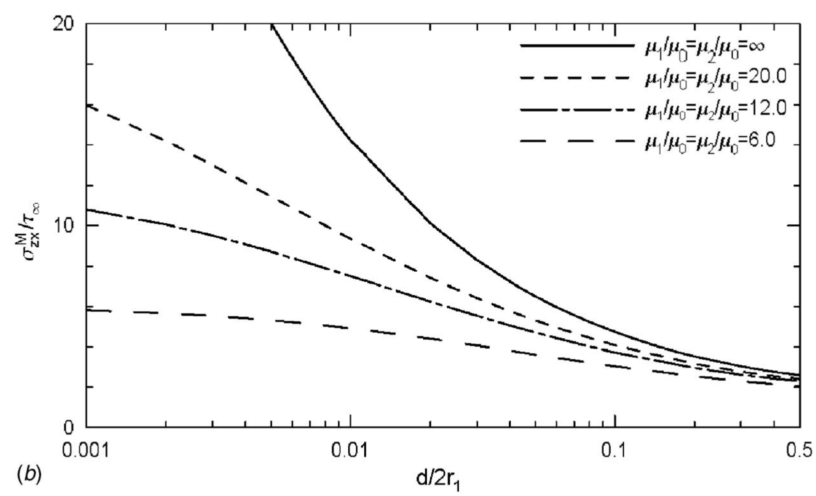


Fig. 6 (*a*) Two identical inclusions with centers on the *x* axis and (*b*) average shear stress of inclusion versus fiber spacing

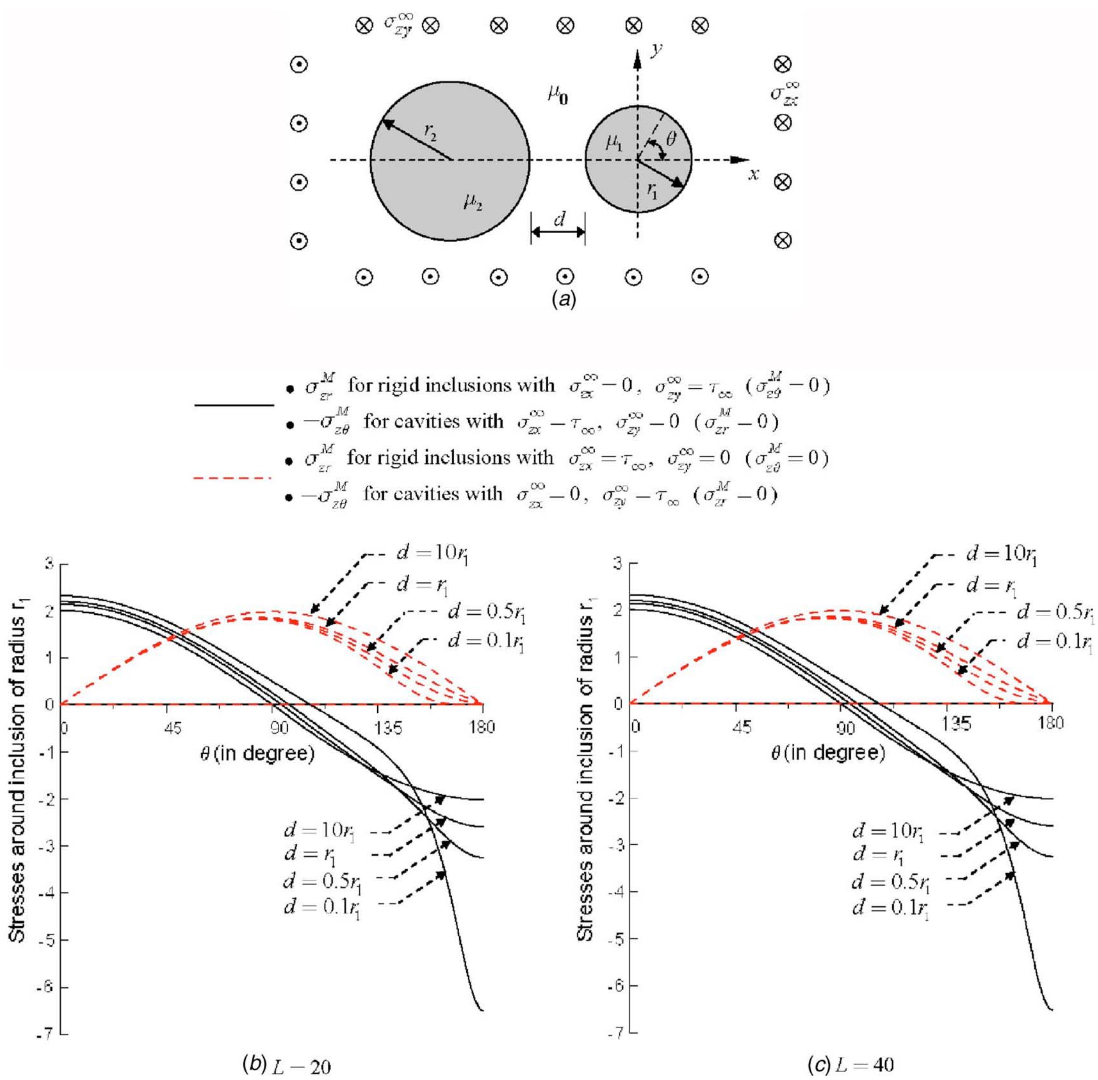


Fig. 7 (*a*) Two circular inclusions with centers on the *x* axis, (*b*) effects of spacing on the stresses around the boundary of radius r_1 for two equal-sized inclusions (L=20), (*c*) effects of spacing on the stresses around the boundary of radius r_1 for two equal-sized inclusions (L=40), (*d*) effects of the size of neighboring inclusion on the stresses around the boundary of radius r_1 with d=0.1r_1 (L=80), and (*e*) effects of the size of neighboring inclusion on the stresses around the boundary of radius r_1 with d=0.1r_1 (L=100)

electrostatic and magnetic problems. Besides, extensions to Helmholtz and biharmonic operators as well as 3D problems are straightforward.

Acknowledgment

The second author (A. C. Wu) would like to thank Wen-Cheng Shen for valuable discussions and the student scholarship from the China Engineering Consultants, Inc., Taiwan. The present work was supported by the National Science Council of Taiwan, through Grant No. NSC94-2211-E-019-009.

Appendix A: Calculation for the Forcing Term {a}

According to Eqs. (2) and (3), the displacement and traction fields in the infinite medium due to the remote shear σ_{zx}^{∞} and σ_{zy}^{∞} in Fig. 1(d) are

$$w^{\infty}=\frac{\sigma_{zx}^{\infty}}{\mu_0}x+\frac{\sigma_{zy}^{\infty}}{\mu_0}y, \tag{A1}$$

$$t^{\infty}=\frac{\partial w^{\infty}}{\partial n}=-\left(\frac{\sigma_{zx}^{\infty}}{\mu_0}n_x+\frac{\sigma_{zy}^{\infty}}{\mu_0}n_y\right), \tag{A2}$$

where the unit outward normal vector on the boundary is $n=(n_x,n_y)$. By comparing Eq. (19) with the first low of Eq. (35), we have

$$\{\mathbf{a}\}=[\mathbf{T}^{M}]\{\mathbf{w}^{\infty}\}-[\mathbf{U}^{M}]\{\mathbf{t}^{\infty}\}. \tag{A3}$$

For the circular boundary which the original system is located, the boundary conditions due to the remote shear are

$$w_1^{\infty}=\frac{\sigma_{zx}^{\infty}}{\mu_0}r_1\cos\theta_1+\frac{\sigma_{zy}^{\infty}}{\mu_0}r_1\sin\theta_1, \tag{A4}$$

$$t_1^{\infty}=-\left(\frac{\sigma_{zx}^{\infty}}{\mu_0}\cos\theta_1+\frac{\sigma_{zy}^{\infty}}{\mu_0}\sin\theta_1\right). \tag{A5}$$

Considering the boundary condition, due to the remote shear, on the kth circular boundary with respect to the observer system, we have

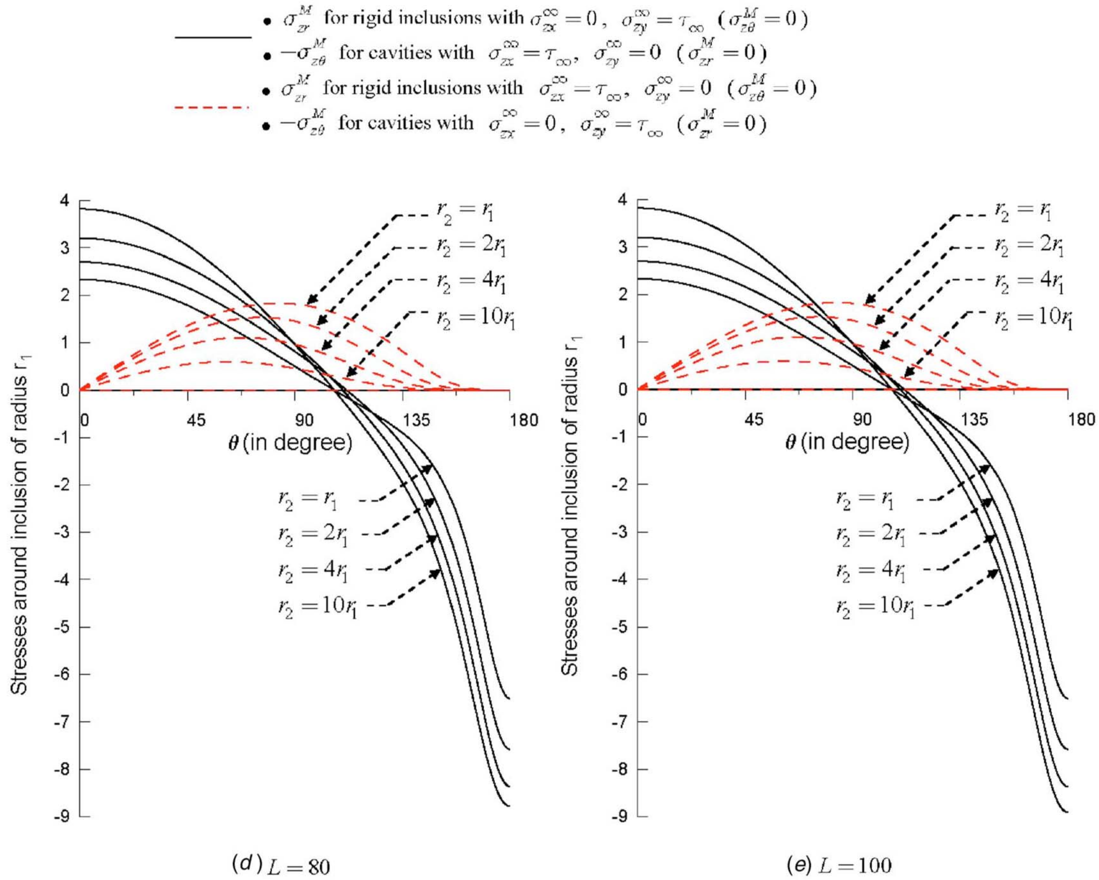


Fig. 7 (***Continued***).

$$w_k^{\infty}=\frac{\sigma_{zx}^{\infty}}{\mu_0}(e_x+r_k\cos\theta_k)+\frac{\sigma_{zy}^{\infty}}{\mu_0}(e_y+r_k\sin\theta_k), \tag{A6}$$

$$t_k^{\infty}=-\left(\frac{\sigma_{zx}^{\infty}}{\mu_0}\cos\theta_k+\frac{\sigma_{zy}^{\infty}}{\mu_0}\sin\theta_k\right), \tag{A7}$$

where e_x and e_y, respectively, denote the eccentric distance of kth inclusion in the x and y direction. By comparing Eq. (A5) with Eq. (A7), we find that t^{∞} can be described in any observer system without any change, where θ_k denotes the polar angle in the adaptive observer coordinate system.

Appendix B: Derivation of the Exact Solution for a Single Inclusion

We derive the exact solution for antiplane problem with a single inclusion under the remote shear using the present formulation. The infinite medium under the shear stress $\sigma_{zx}^{\infty}=0$ and $\sigma_{zy}^{\infty}=\tau_{\infty}$ at infinity is considered. The Fourier coefficients in Eq. (24) can be written as

$$\{\mathbf{w}^{\infty}\}=\begin{Bmatrix}0\\0\\\dfrac{\tau_{\infty}r_1}{\mu_0}\\\vdots\\0\\0\end{Bmatrix}_{(2L+1)\times1},\quad \{\mathbf{t}^{\infty}\}=\begin{Bmatrix}0\\0\\-\dfrac{\tau_{\infty}}{\mu_0}\\\vdots\\0\\0\end{Bmatrix}_{(2L+1)\times1}, \tag{B1}$$

where r_1 is the radius of the single inclusion. By substituting the appropriate degenerate kernels in Eqs. (12) and (13) into Eqs. (19) and (20) and employing the continuity of displacement and equilibrium of traction along the interface in Eqs. (32) and (33), the unknown boundary data in Eqs. (23) and (25) can be obtained using the symbolic software MATHEMATICA as shown below

$$\{\mathbf{w}^{M}\}=\begin{Bmatrix}0\\0\\\dfrac{2\tau_{\infty}r_1}{\mu_0+\mu_1}\\\vdots\\0\\0\end{Bmatrix}_{(2L+1)\times1},\quad \{\mathbf{t}^{M}\}=\begin{Bmatrix}0\\0\\\dfrac{-2\tau_{\infty}\mu_1}{\mu_0(\mu_0+\mu_1)}\\\vdots\\0\\0\end{Bmatrix}_{(2L+1)\times1}, \tag{B2}$$

$$\{\mathbf{w}^{I}\}=\begin{Bmatrix}0\\0\\\dfrac{2\tau_{\infty}r_1}{\mu_0+\mu_1}\\\vdots\\0\\0\end{Bmatrix}_{(2L+1)\times1},\quad \{\mathbf{t}^{I}\}=\begin{Bmatrix}0\\0\\\dfrac{2\tau_{\infty}}{\mu_0+\mu_1}\\\vdots\\0\\0\end{Bmatrix}_{(2L+1)\times1}. \tag{B3}$$

After substituting Eqs. (B1) and (B2) into the boundary integral equation for the domain point in Eq. (6), we obtain the total stress fields in the matrix

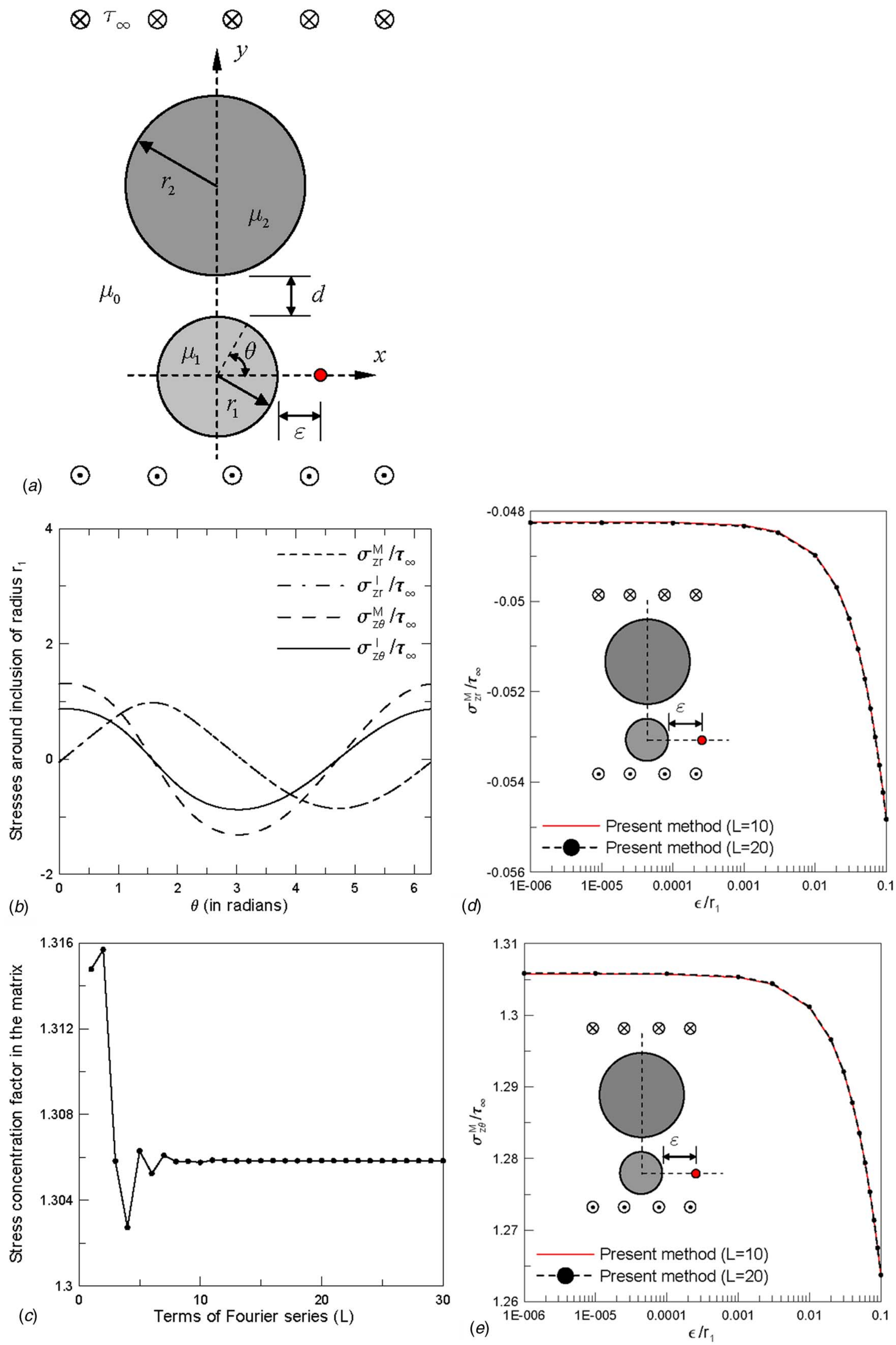


Fig. 8 (*a*) Two circular inclusions with centers on the *y* axis, (*b*) stresses around the circular boundary of radius r_1, (*c*) convergence test of the two-inclusions problem, (*d*) radial stress in the matrix near the boundary, and (*e*) tangential stress in the matrix near the boundary

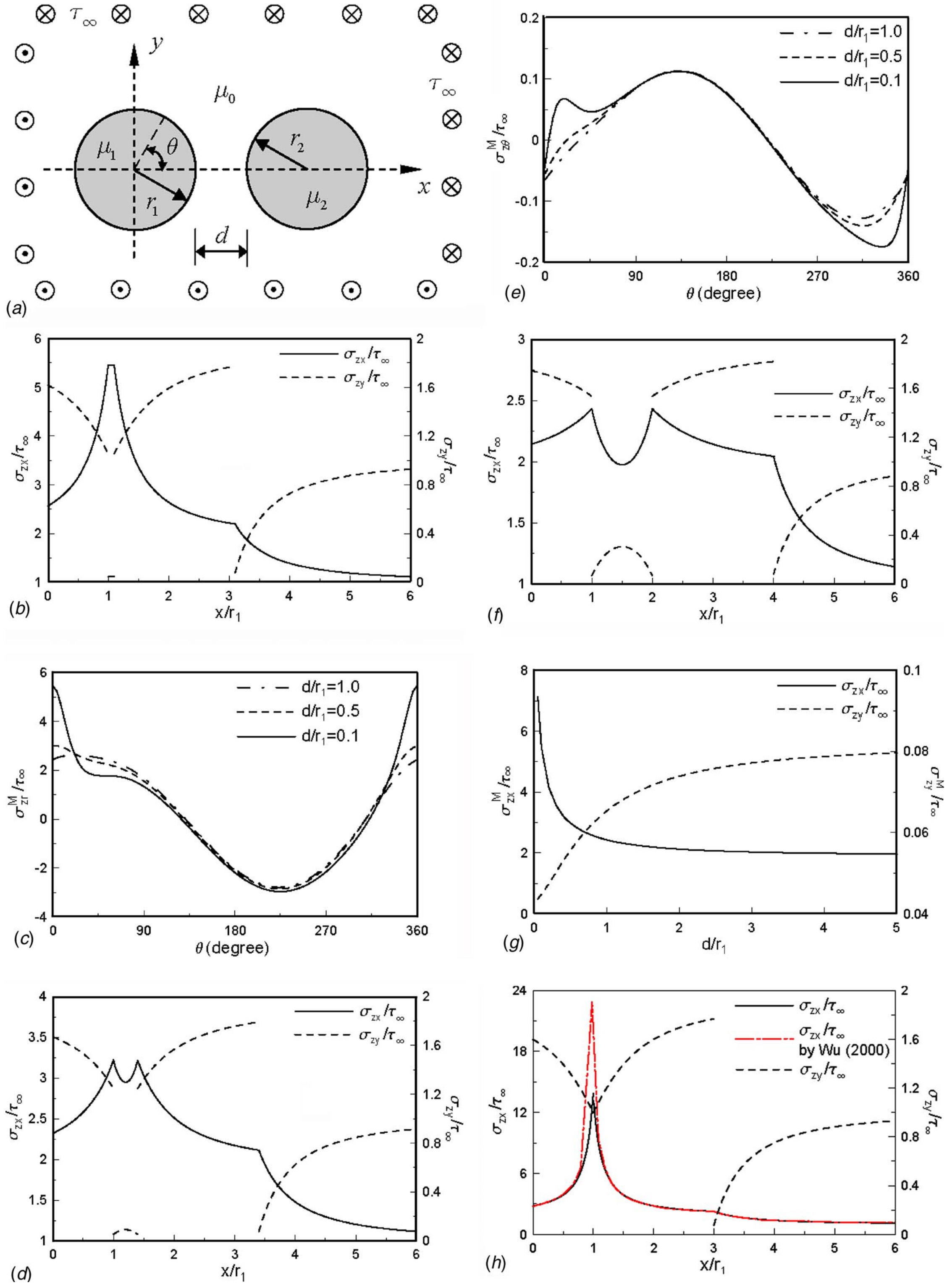


Fig. 9 (*a*) **Two circular inclusions embedded in a matrix under the remote antiplane shear in two directions,** (*b*) **stress distributions along the *x* axis when $d=0.1$,** (*c*) **stress distributions along the *x* axis when $d=0.4$,** (*d*) **stress distributions along the *x* axis when $d=1.0$,** (*e*) **normal stress distributions along the contour $(1.001, \theta)$,** (*f*) **tangential stress distributions along the contour $(1.001, \theta)$,** (*g*) **variations of stresses at the point (1.001,0 deg),** **and** (*h*) **stress distributions along the *x* axis when the two inclusions touch each other**

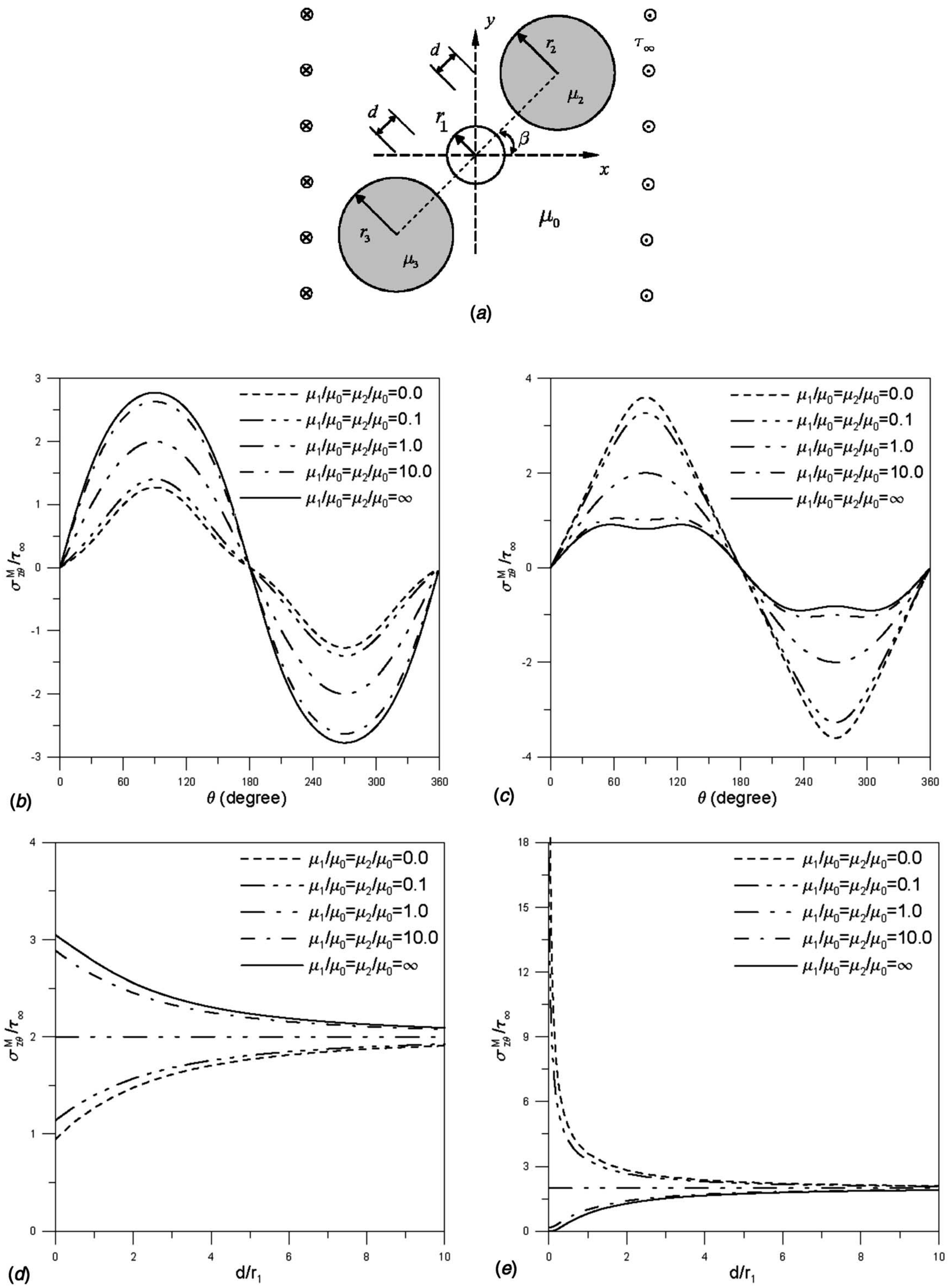


Fig. 10 (*a*) One hole surrounded by two circular inclusions, (*b*) tangential stress distribution along the hole boundary with $\beta=0$ deg, (*c*) tangential stress distribution along the hole boundary with $\beta=90$ deg, (*d*) stress concentration as a function of the spacing d/r_1 with $\beta=0$ deg, and (*e*) stress concentration as a function of the spacing d/r_1 with $\beta=90$ deg

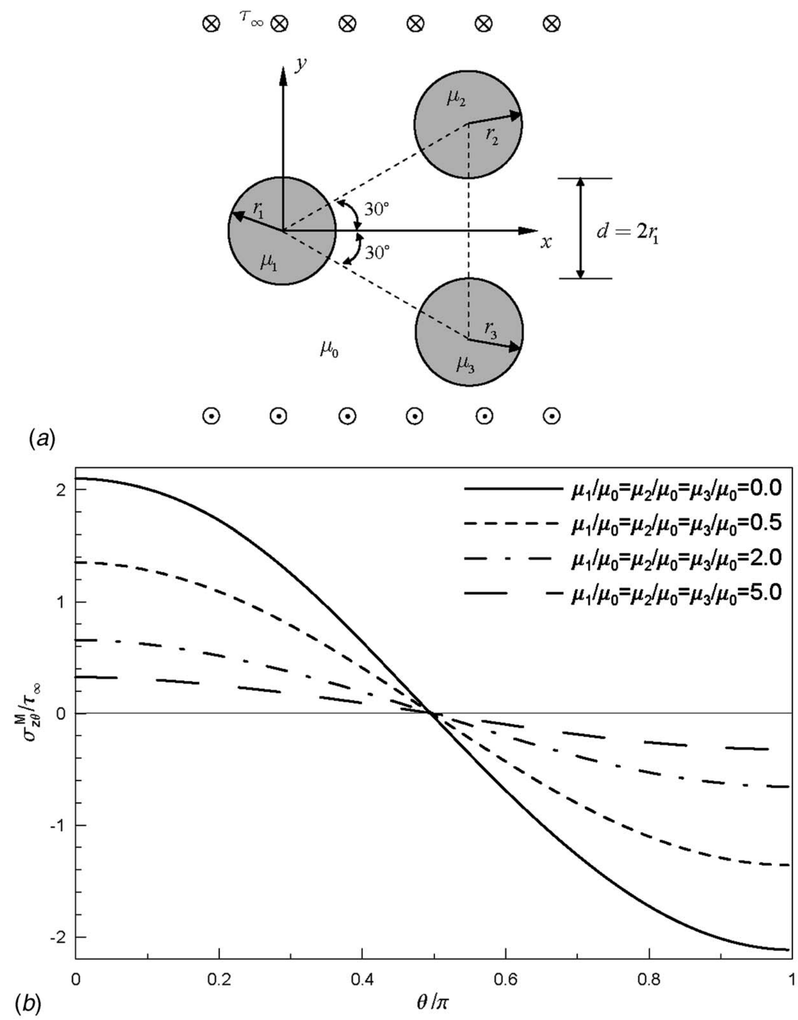


Fig. 11 (*a*) Three identical inclusions forming an equilateral triangle, and (*b*) tangential stress distribution around the inclusion located at the origin

$$\sigma^M_{zx} = \mu_0 \frac{\partial w^M}{\partial x} + \sigma^\infty_{zx} = -2\tau_\infty \frac{r_1^2}{\rho^2} \frac{\mu_0 - \mu_1}{\mu_0 + \mu_1} \sin\phi \cos\phi,$$

$$r_1 \leqslant \rho \leqslant \infty, \quad 0 \leqslant \phi \leqslant 2\pi, \tag{B4}$$

$$\sigma^M_{zy} = \mu_0 \frac{\partial w^M}{\partial y} + \sigma^\infty_{zy} = \tau_\infty \frac{r_1^2}{\rho^2} \frac{\mu_0 - \mu_1}{\mu_0 + \mu_1} (\cos^2\phi - \sin^2\phi) + \tau_\infty,$$

$$r_1 \leqslant \rho \leqslant \infty, \quad 0 \leqslant \phi \leqslant 2\pi. \tag{B5}$$

After substituting Eq. (B3) into the boundary integral equation for the domain point in Eq. (6), we have the total stress fields in the inclusion

$$\sigma^I_{zx} = \mu_1 \frac{\partial w^I}{\partial x} = 0, \quad 0 \leqslant \rho \leqslant r_1, \quad 0 \leqslant \phi \leqslant 2\pi, \tag{B6}$$

$$\sigma^I_{zy} = \mu_1 \frac{\partial w^I}{\partial y} = 2\tau_\infty \frac{\mu_1}{\mu_0 + \mu_1}, \quad 0 \leqslant \rho \leqslant r_1, \quad 0 \leqslant \phi \leqslant 2\pi. \tag{B7}$$

Finally, the stress components σ_{zr} and $\sigma_{z\theta}$ in Eqs. (38) and (39) can be superimposed by using σ_{zx} and σ_{zy} as shown below

$$\sigma^M_{zr} = 2\tau_\infty \frac{r_1^2}{\rho^2} \frac{\mu_1}{\mu_0 + \mu_1} \sin\phi, \quad r_1 \leqslant \rho \leqslant \infty, \quad 0 \leqslant \phi \leqslant 2\pi, \tag{B8}$$

$$\sigma^M_{z\theta} = 2\tau_\infty \frac{r_1^2}{\rho^2} \frac{\mu_0}{\mu_0 + \mu_1} \cos\phi, \quad r_1 \leqslant \rho \leqslant \infty, \quad 0 \leqslant \phi \leqslant 2\pi, \tag{B9}$$

$$\sigma^I_{zr} = 2\tau_\infty \frac{\mu_1}{\mu_0 + \mu_1} \sin\phi, \quad 0 \leqslant \rho \leqslant r_1, \quad 0 \leqslant \phi \leqslant 2\pi, \tag{B10}$$

$$\sigma^I_{z\theta} = 2\tau_\infty \frac{\mu_1}{\mu_0 + \mu_1} \cos\phi, \quad 0 \leqslant \rho \leqslant r_1, \quad 0 \leqslant \phi \leqslant 2\pi. \tag{B11}$$

It is obvious to see that the maximum stress concentration occurs at $\rho = r_1$ and $\phi = 0$. The stress concentration factor is reduced due to the inclusion in comparison with that of cavity ($\mu_1 = 0$) as shown in Eq. (B9). Besides, it is noted that σ^M_{zr} coincides with σ^I_{zr} as required by the traction equilibrium on the interface between the matrix and inclusion. The exact solution for a single inclusion using the present formulation matches well with the previous one obtained by employing the complex-variable formulation [1].

References

[1] Honein, E., Honein, T., and Herrmann, G., 1992, "On Two Circular Inclusions in Harmonic Problems," Q. Appl. Math., **50**, pp. 479–499.

[2] Steif, P. S., 1989, "Shear Stress Concentration between Holes," ASME J. Appl. Mech., **56**, pp. 719–721.

[3] Budiansky, B., and Carrier, G. F., 1984, "High Shear Stresses in Stiff-Fiber Composites," ASME J. Appl. Mech., **51**, pp. 733–735.

[4] Zimmerman, R. W., 1988, "Second-Order Approximation for the Compression of an Elastic Plate Containing a Pair of Circular Holes," Z. Angew. Math. Mech., **68**, pp. 575–577.

[5] Sendeckyj, G. P., 1971, "Multiple Circular Inclusion Problems in Longitudinal Shear Deformation," J. Elast., **1**(1), pp. 83–86.

[6] Goree, J. G., and Wilson, H. B., 1967, "Transverse Shear Loading in an Elastic Matrix Containing Two Circular Cylindrical Inclusions," ASME J. Appl. Mech., **34**, pp. 511–513.

[7] Bird, M. D., and Steele, C. R., 1992, "A Solution Procedure for Laplace's Equation on Multiply-Connected Circular Domains," ASME J. Appl. Mech., **59**, pp. 398–404.

[8] Wu, L. Z., 2000, "Interaction of Two Circular Cylindrical Inhomogeneities under Anti-Plane Shear," Compos. Sci. Technol., **60**, pp. 2609–2615.

[9] Chao, C. K., and Young, C. W., 1998, "On the General Treatment of Multiple Inclusions in Antiplane Elastostatics," Int. J. Solids Struct., **35**, pp. 3573–3593.

[10] Gong, S. X., 1995, "Antiplane Interaction among Multiple Circular Inclusions," Mech. Res. Commun., **22**(3), pp. 257–262.

[11] Chou, S. I., 1997, "Stress Field around Holes in Antiplane Shear Using Complex Variable Boundary Element Method," ASME J. Appl. Mech., **64**, pp. 432–435.

[12] Ang, W. T., and Kang, I., 2000, "A Complex Variable Boundary Element Method for Elliptic Partial Differential Equations in a Multiple-Connected Region," Int. J. Comput. Math., **75**, pp. 515–525.

[13] Mogilevskaya, S. G., and Crouch, S. L., 2001, "A Galerkin Boundary Integral Method for Multiple Circular Elastic Inclusions," Int. J. Numer. Methods Eng., **52**, pp. 1069–1106.

[14] Kress, R., 1989, *Linear Integral Equations*, Springer-Verlag, Berlin.

[15] Chen, J. T., Kuo, S. R., and Chen, K. H., 1999, "A Nonsingular Integral Formulation for the Helmholtz Eigenproblems of a Circular Domain," J. Chin. Inst. Eng., **22**(6), pp. 729–739.

[16] Chen, J. T., and Kuo, S. R., 2000, "On Fictitious Frequencies Using Circulants for Radiation Problems of a Cylinder," Mech. Res. Commun., **27**(1), pp. 49–58.

[17] Chen, J. T., Kuo, S. R., and Lin, J. H., 2002, "Analytical Study and Numerical Experiments for Degenerate Scale Problems in the Boundary Element Method for Two-Dimensional Elasticity," Int. J. Numer. Methods Eng., **54**(12), pp. 1669–1681.

[18] Chen, J. T., Shen, W. C., and Wu, A. C., 2006, "Null-Field Integral Equations for Stress Field around Circular Holes under Antiplane Shear," Eng. Anal. Boundary Elem., **30**, pp. 205–217.

[19] Chen, J. T., Hsiao, C. C., and Leu, S. Y., 2006, "Null-Field Integral Equation Approach for Plate Problems with Circular Boundaries," ASME J. Appl. Mech., **73**, pp. 679–693.

[20] Chen, J. T., and Hong, H.-K., 1999, "Review of Dual Boundary Element Methods with Emphasis on Hypersingular Integrals and Divergent Series," Appl. Mech. Rev., **52**, pp. 17–33.

[21] Chen, J. T., and Chiu, Y. P., 2002, "On the Pseudo-Differential Operators in the Dual Boundary Integral Equations Using Degenerate Kernels and Circulants," Eng. Anal. Boundary Elem., **26**, pp. 41–53.

[22] Lebedev, N. N., Skalskaya, I. P., and Uflyand, Y. S., 1979, *Worked Problem in Applied Mathematics*, Dover, New York.

Surface Energy for Creation of Multiple Curved Cracks in Rubbery Materials

J. H. Chang
Professor

J. S. Lin
Graduate Research Assistant

Department of Civil Engineering,
National Central University,
Chungli 32001,
Taiwan

A fracture parameter M_c is proposed for evaluation of the surface energy associated with the creation of multiple curved cracks in 2D rubbery solids under the action of large deformation. Based on the concept of the M-integral, the parameter is developed by performing the integration along a closed contour enclosing all the cracks and with respect to a reference coordinate system originated at the geometric center of all the crack tips. The integration is shown to be path-independent so that the complicated singular stress field in the near-tip areas need not be involved in the calculation. It is thus suggested that M_c be possibly used as a fracture parameter for describing the degradation of material and/or structural integrity caused by irreversible evolution of multiple curved cracks in a rubbery media. [DOI: 10.1115/1.2338058]

1 Introduction

For those materials with mechanical behavior remaining (nonlinearly) elastic at large strains, we often call them rubbery. Rubbery materials find applications in many engineering components, where the structural integrity is substantially limited by the growth of a system of distributed cracks rather than a single, continuous crack. Such multi-cracked fracture conditions range from the presence of a finite number of macrocracks, to the formation of densely distributed microcracks with random location and orientation. Also, the cracks may have arbitrary curved shapes and so the associated fracture behavior depends significantly on the crack geometry. For rubbery problems involving nonlinear responses under large deformations, the local stress state in the region around these strongly interacted crack tips becomes quite complicated and difficult to describe analytically. Investigation on the use of proper energy parameters in characterizing the global fracture state corresponding to the multi-cracked damage is, therefore, in need.

For rubbery material problems containing a single crack, the J_k-integrals ($k=1,2$) have been successfully used for evaluating the energy release rates corresponding to extension and kinking of the crack tip either analytically or numerically (e.g., Rivlin and Thomas [1], Pidaparti, et al. [2], Chang and Yeh [3], etc.). However, due to their "local" nature associated with a single crack tip, J_k are not feasible for description of the "global" multi-cracked fracture state.

In addition to J_k, the well-known energy conservation contour integrals derived from Noether's theorem in plane elasticity also include the M- and L-integrals (Knowles and Sternberg [4], Budiansky and Rice [5], Eshelby [6], and Freund [7]). While application of the M-integral is not as common as J_k, it has been used in linearly elastic problems containing either an isolated singular point or a straight crack (e.g., Herrmann and Herrmann [8], Eischen and Herrmann [9], Seed [10], etc.). The integration contours for M in these applications were defined in a "global" way, i.e., M is evaluated along a counterclockwise contour enclosing the whole single defect. Such a global feature implies the applicability of the M-integral to fracture analysis for multi-cracked problems, provided that the integration contours are suitably chosen.

Due to its global feature, the M-integral has recently been applied to characterizing the fracture state associated with the presence of multiple straight cracks. A series of studies were presented in evaluating the material damage level for uniformly loaded microcracking linearly elastic solids, where the calculation is carried out by including all the microcracks inside the integration contour (Chen [11], Wang and Chen [12], Tian [13], etc.). Also, a problem-invariant parameter M_c is proposed by the authors and suggested as an energy fracture parameter for describing the degradation of material and/or structural integrity caused by the irreversible evolution of multiple cracks in linearly elastic solids (Chang and Chien [14], Chang and Wu [15]) and rubbery materials (Chang and Peng [16]), respectively.

In this paper, an energy parameter M_c is proposed for evaluating the surface energy corresponding to the creation of a multiple-cracked system, in which each crack may be of an arbitrary curved shape. The formulation is based on the concept of the M-integral and considered to be suited for fracture analysis in rubbery material problems subjected to large elastic deformation. The outline of this paper is organized as follows. In Sec. 2, an M_c-integral is defined by suitably selecting the integration contour and originating the coordinate system at the geometric center of the enclosed curved cracks. Subsequently in Sec. 3, the property of path-independence for M_c is established. In Sec. 4, the physical meaning of M_c in the context of large deformation is interpreted. In Sec. 5, the feasibility of M_c is illustrated via numerical examples conducted by using finite elements. Attention is addressed to illustration on its use for the study of the material and/or structural behavior degraded due to presence of the curved cracks.

2 The M_c-Integral Under Large Deformation

In this section, the basic concept of the M_c-integral is introduced for rubbery material problems containing curved cracks and subjected to large elastic deformation. The integral is defined by taking the conventional M-integral with specific selection of integration contour and reference coordinate system. In the following study, the single-crack case is first considered, and then followed by the multi-cracked condition.

2.1 A Single Curved Crack. In the context of large elastic deformation, the following contour integral is applied with state variables reinterpreted with respect to a specific reference configuration. We consider a homogeneous rubbery body in a 2D

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received July 29, 2005; final manuscript received March 24, 2006. Review conducted by K. Ravi Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

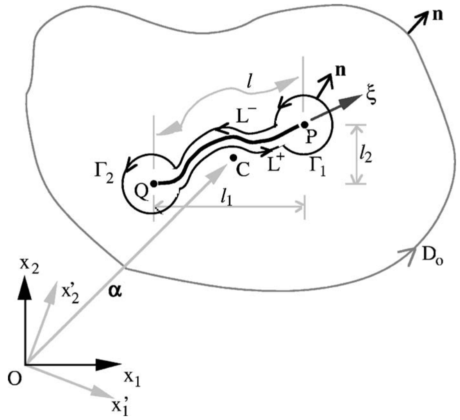


Fig. 1 A rubbery body containing a two-tip curved crack at its original state

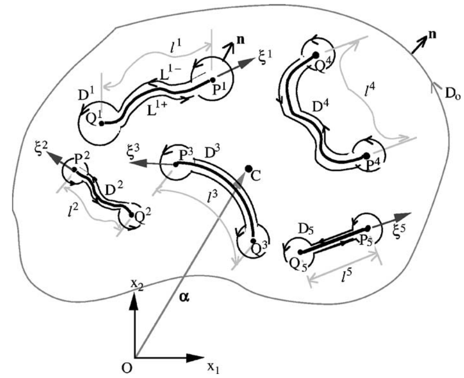


Fig. 2 A homogeneous rubbery body containing *N* distributed curved cracks at its original state (*N*=5 in this figure)

field, containing a single curved crack with two tips P and Q at its original (unstressed and undeformed) state (Fig. 1). The crack is of arbitrary curved shape and of length l. In this reference configuration, a coordinate system originating at an arbitrarily chosen point O is introduced and, with no loss of generality, the crack lying tangentially parallel to the x_1-direction (asymptotically at tip Q). The position vector at the geometric center C of the two tips is denoted $\boldsymbol{\alpha}$. The body is then subjected to a system of external loads and reaches its current deformed state. When the body is subjected to no body forces, the M-integral with respect to O is defined as

$$M = \int_D \left[W n_i x_i - T_k \left(\frac{\partial u_k}{\partial x_i} \right) x_i \right] ds \tag{1}$$

The integral is performed in the undeformed reference configuration, where D is a counterclockwise closed contour around the whole crack and consists of four parts $L^+ + \Gamma_1 + L^- + \Gamma_2$, W is the strain energy density of the material (per unit undeformed volume), $\mathbf{T}$ is the traction vector resulting from the first Piola-Kirchhoff stress tensor along the contour, $\mathbf{n}$ is the outward unit vector normal to D, s is the arc length along D, $\mathbf{x}$ is the position vector of the integration point along D, and $\mathbf{u}$ is the displacement vector. For rubbery materials modeled with hyperelasticity, the first Piola-Kirchhoff stress $\boldsymbol{\sigma}$ is defined by the following constitutive relation as:

$$\sigma_{ij} = \frac{\partial W}{\partial (u_{i,j})} \tag{2}$$

with $u_{ij} = \partial u_i / \partial x_j$,

By definition, the integration is carried out by taking the limiting case where Γ_1 and Γ_2 are shrunk onto the crack tips $\mathbf{P}$ and Q, respectively, and L^+ and L^- are lying along the curved crack surfaces (this limiting case is not shown in Fig. 1). When the crack surfaces are traction-free, it can be shown (Appendix A) that M is related to the J_k-integrals as

$$M = \alpha_k \left[(J_k^P - J_k^Q) + \left(\int_{L^+ + L^-} W n_k \, d\xi \right) \right] + \frac{l_k}{2} (J_k^P + J_k^Q) + \left[\int_{L^+ + L^-} \left(\xi_k - \frac{l_k}{2} \right) W n_k \, d\xi \right] \tag{3}$$

where J_k^P and J_k^Q are the J_k integrals evaluated along Γ_1 and Γ_2, respectively, l_k is the length of the projection of l in the x_k-direction, ξ is the local curvilinear coordinate system lying along the crack curve and originated at tip Q, and ξ_k is the length of the projection of ξ in the x_k-direction. The J_k-integrals (say, at point P) are defined as

$$J_k^P = \lim_{\Gamma_1 \to 0} \int_{\Gamma_l} \left[W n_k - T_i \left(\frac{\partial u_i}{\partial x_k} \right) \right] ds \quad k = 1,2 \tag{4}$$

Again, by definition, the J_k-integrals are evaluated by taking the limiting case in which Γ_1 and Γ_2 are shrunk towards P and Q. Equation (4) shows that, while J_k remains invariant for a given boundary value problem, the value of M varies with the selection of origin O and appear to depend linearly upon the components of the corresponding position vector $\boldsymbol{\alpha}$. Nevertheless, by locating the origin at the geometric center C, i.e., by taking $\boldsymbol{\alpha} = (0,0)$, we then define a problem-invariant parameter M_c as

$$M_C \equiv M|_{\boldsymbol{\alpha}=(0,0)} = \frac{l_k}{2} (J_k^P + J_k^Q) + \left[\int_{L^+ + L^-} \left(\xi_k - \frac{l_k}{2} \right) W n_k \, d\xi \right] \tag{5}$$

Further, it can easily be shown that the quantities in Eqs. (3) and (5) are independent of the orientation of the coordinate system. Such a characteristic implies that M, as well as M_c, remains unchanged when they are evaluated with respect to an arbitrarily oriented system, e.g., $x_1{}' - x_2{}'$ depicted in Fig. 1.

For the special case corresponding to a straight crack, the results of M and M_c reduce to

$$M = \alpha_1 (J_1^P - J_1^Q) + \alpha_2 (J_2^P - J_2^Q) + \frac{l}{2} (J_1^P + J_1^Q) + \alpha_2 \int_{L^+ + L^-} W n_2 \, ds \tag{6}$$

and

$$M_C = \frac{l}{2} (J_1^P + J_1^Q) \tag{7}$$

respectively, where l is the length of the crack.

2.2 Multiple Curved Cracks. Consider the 2D homogeneous rubbery body containing N distributed curved cracks at its reference configuration (i.e., the original undeformed state), each curve of length l^r $(r=1,\ldots,N)$ and with random location and orientation, as shown in Fig. 2 (N=5 in this figure). The geometric center of all the crack tips (i.e., P^r and Q^r, $r=1,\ldots,N$) is positioned and denoted C.

The M-integral associated with the N curved cracks is defined as

$$M = \sum_{r=1}^{N} \int_{D^r} \left[W n_i x_i - T_k \left(\frac{\partial u_k}{\partial x_i} \right) x_i \right] ds \tag{8}$$

where D^r is the counterclockwise closed contour associated with the rth crack. Still, when the body is subjected to large deformation, the above contour integral is applied with all state variables reinterpreted with respect to this reference configuration. Again, by definition, the integration is performed by taking the limiting case in which the portions of D^r are shrunk onto the crack tips and lying along the crack surfaces. Also, note that the value of M varies with respect to different selections of origin O. Therefore, by further locating the origin at the geometric center C, the problem-invariant parameter M_c can thus be defined.

3 Path-Independence

In this section, the property of path-independence for M_c is established. We first consider the instance containing a single curved crack, as shown in Fig. 1. When the crack surfaces are traction-free, it can easily be shown that the M-integral defined in Eq. (1) is path-independent and has the same value for any remote contour, say D_o, around the whole crack. As a consequence, the M_c-integral defined in Eq. (5) is also path-independent in the same manner.

Next, we consider the condition that contains N distributed curved cracks, as shown in Fig. 2. When all the crack surfaces are traction-free, it can be shown that the M-integral defined in Eq. (8) is path-independent and has the same value for any remote contour D_o, i.e.,

$$M = \int_{D_o} \left[W n_i x_i - T_k \left(\frac{\partial u_k}{\partial x_i} \right) x_i \right] ds \tag{9}$$

The outer contour can be arbitrarily chosen (except for the requirements to be inside the body, enclose all the N cracks, and contain no other singularity in it). The property of path-independence implies that good accuracy of the integration can be achieved by adopting a contour remote from the tips of the cracks, where complicated singular stress behavior dominates. Note that the M_c-integral is also path-independent in the same manner.

4 Physical Interpretation

In the following two subsections, the physical meaning for M_c in the context of large deformation is discussed. A single straight crack is considered in the first subsection, and then followed by multiple curved cracks in the second subsection. The content in each subsection is outlined as follows. First, the concept of origin-independence, a special property for M when the body is homogeneously stressed, is illustrated. Subsequently, the physical meaning for M_c is interpreted. Finally, the feasibility of M_c for use in characterizing both the material and the structural fracture behavior due to evolution of multiple curved cracks is illustrated.

4.1 A Single Straight Crack. For problems containing a single straight crack, the results of M and M_c can be related to the J_k-integrals as shown in Eqs. (6) and (7), respectively. To illustrate the property of origin-independence, an interesting feature under the special condition when the body is homogeneously stressed, we take the rubbery body as an infinite medium and subjected to a uniform remote loading system. Under this condition, it is observed that $J_k^P = J_k^Q$ and the integration of Wn_2 along the crack surfaces vanishes due to symmetry. Therefore, the M-integral in Eq. (6) and the M_c-integral in Eq. (7) become equivalent and both reduce to

$$M = M_C = lJ_1^{\,P} = lJ_1^{\,Q} \tag{10}$$

Equation (10) implies that the value of M remains unchanged with different selections of origin O of the reference coordinate system, i.e., it is origin-independent in a homogeneously stressed infinite medium.

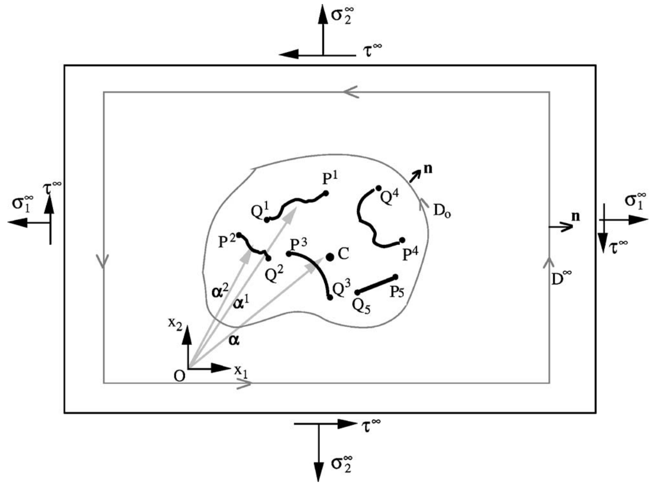


Fig. 3 The integration contours for an infinite medium subjected to uniform remote loads

Under the aforementioned special homogeneously stressed condition, it is well known that the result of M $(=M_c)$ in linear elasticity can be physically related to the energy release rate due to self-similar expansion of the crack (Herrmann and Herrmann [8]). Furthermore, as illustrated in the authors' previous work (Chang and Peng [16]) under a more general condition (i.e., nonlinear elastic media subjected to nonhomogeneous large deformation), the M_c-integral appears to be equal to twice the surface energy required for formation of the whole crack. Since the latter physical interpretation is derived under such a general condition it is, therefore, anticipated that M_c be feasible in characterizing the structural fracture behavior due to presence of the crack.

4.2 Multiple Curved Cracks. We first take the special condition when the N distributed curved cracks are located in an infinite medium and subjected to uniform remote loads, as shown in Fig. 3. It can be shown (Appendix B) that, under such a homogeneously stressed condition, the result of the M-integral is independent of the origin O provided that the integration is performed by including all cracks in a single remote contour (say, D_o in Fig. 3). This also indicates that $M = M_c$ under this special condition. Nevertheless, it should be noted that the value of M does in general vary with respect to different selections of origin when the body is nonhomogeneously stressed.

Still, the physical meaning of M_c can be interpreted in two different aspects. First, it is well known that evaluation of M_c yields summation of the driving work conjugate to expansion of each crack tip relative to the geometric center C. This feature, nevertheless, is associated with simultaneous expansion of all crack tips and rarely occurs in reality. On the other hand, it can alternatively be shown (Appendix C) that the result of M_c is equivalent to twice the surface energy required for formation of all the curved cracks (i.e., the change of potential energy due to creation of the cracks). The latter interpretation appears to be more applicable in practice. Furthermore, it is noted that this characteristic is derived under the general nonhomogeneously stressed condition. This indicates that M_c can be used for characterizing the degradation of structural integrity corresponding to formation of the whole set of curved cracks.

5 Numerical Examples

Two numerical examples are presented in the following two subsections. A uniformly loaded specimen (material characteristic) is considered in the first subsection, and subsequently a non-uniformly loaded problem (structural characteristic) is presented

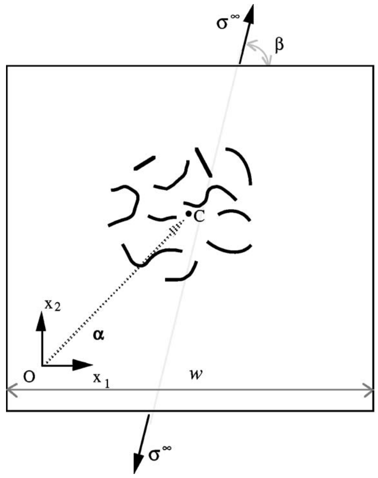


Fig. 4 A plane stress rubbery specimen containing multiple curved cracks (Problem 1)

in the next subsection. The rubbery material behavior is modeled with a neo-Hookean isotropic incompressible strain energy density function W (Marusak [17]) as

$$W = 97.2(I_1 - 3) \quad \text{(unit: Pa)} \tag{11}$$

where I_1 is the first principal invariants associated with the Cauchy-Green strain tensor. In the following calculations, quadratic finite elements are used for interpolation of the displacement field. Also, since the material is treated as incompressible, by directly using quadratic elements for interpolation of the pressure field will result in an overconstrained numerical system. In order to satisfy the requirement for numerical stability (usually termed the "discretized LBB condition" [18]), the pressure field is thus interpolated in terms of nonconforming linear elements with reduced integration. In addition, note that no particular singular elements are used throughout the study.

Although M_c can be physically related both to the energy release rate due to crack expansion and to the surface energy associated with crack formation, only the validity of the latter will be addressed.

5.1 Problem 1—Material Characteristic (Homogeneously Stressed Condition). The object of this problem is to verify the feasibility of using the origin-independent M (or, equivalently, M_c) as a fracture parameter for the description of the material degradation in a homogeneously stressed rubbery medium. To this end, we consider a plane stress rubbery specimen containing a number of densely (and randomly) distributed curved cracks, each with a different length and orientation, as shown in Fig. 4. The crack length is relatively small compared with the width w and length B so that the finite width effect of the specimen can be neglected. The specimen is subjected to an oblique uniaxial tensile stress σ^∞ along the boundaries.

The analysis in this problem is organized as follows. First, the effect of the local finite element modeling around the crack tips is investigated. Next, the property of path-independence is examined. Subsequently, the origin-independence property of M in this homogeneously stressed medium is investigated. Finally, the physical interpretation of M is verified by comparing the result of the integral with the energy difference due to formation of the cracks.

For accuracy, it is required to compute the M-integral with a high degree of precision. To investigate the effect of finite element modeling around the crack tips, we construct various finite element meshes by successive local h-refinements in the near-tip region, with the size of the tip-element ranging from 2% to 10% of the crack length. Although not depicted in detail here, the results obtained from the above different meshes show very good convergence.

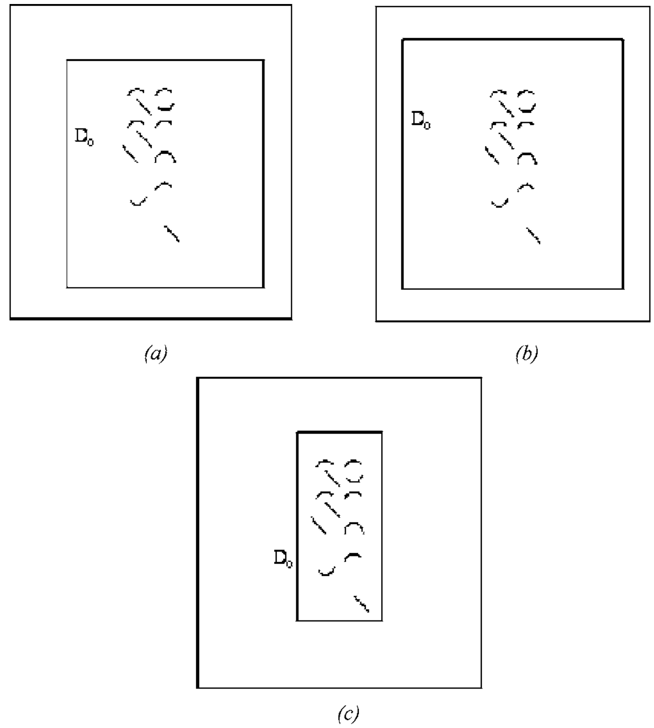


Fig. 5 Three contours, each encloses different portion of the specimen, are used for the integration

Three integration paths, each enclosing a different portion of the same finite element mesh, are used in order to verify path-independence of the calculation. The associated exterior contours D_o's are depicted in Fig. 5. The results with respect to $\boldsymbol{\alpha} = (7.5, 8)m$ under two loading conditions (σ^∞, β) are shown in Table 1. While the M-integral is analytically path-independent, rather minor differences are observed in the numerical results from the three paths. This is because the equilibrium equations are satisfied only weakly in the finite element formulation. The property of path-independence, therefore, only holds weakly in finite element calculations.

In order to show the property of origin-independence, three coordinate systems located at different origins are selected. The values of M with respect to the three choices of $\boldsymbol{\alpha}$ under two loading conditions are shown in Table 2. The numerical results indicate that, even when the origin is chosen to be far away from the cracks, deviations of M are observed to be under 2%. The validity of origin-independence is thus verified.

To illustrate the physical meaning of the M-integral, we consider the same specimen and take its uncracked state as a reference configuration. The values of the potential energy Π associated with both the cracked and uncracked configurations are then calculated by using finite elements. The surface energy due to

Table 1 The results of M from three integration paths for Problem 1 (Pa-m^2) (Note: w=150 m, B=150 m, α(7.5,8)m, N=12)

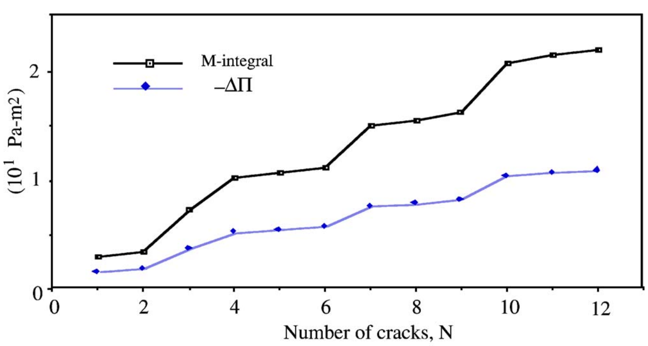
	path 1 (Fig. 5(*a*))	path 2 (Fig. 5(*b*))	path 3 (Fig. 5(*c*))
$(\sigma^\infty, \beta) = (400\ \text{Pa}, 60°)$	22.72	22.75	22.91
$(\sigma^\infty, \beta) = (100\ \text{Pa}, 90°)$	1.632	1.635	1.644

Table 2 The results of M wrt different α's for Problem 1 (Pa-m^2) (Note: w=150 m, B=150 m, N=12)

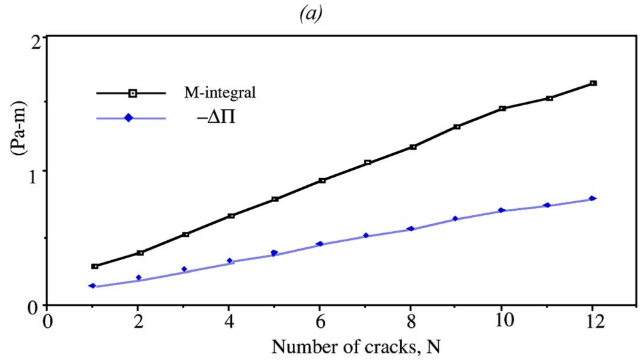
α	(7.5,8)m	(15,15)m	(1000,1000)m
$(\sigma^\infty,\beta)=(400\text{ Pa},60°)$	22.72	22.68	22.43
$(\sigma^\infty,\beta)=(100\text{ Pa},90°)$	1.632	1.628	1.602

formation of the cracks is calculated by evaluating the difference of potential energy $\Delta\Pi$. The values of $\Delta\Pi$ versus the number of cracks N, under two loading conditions, are depicted in Figs. 6(*a*) and 6(*b*), respectively. Also shown in the figures is the FE results of the corresponding M-integral. By comparing the values of M and $-\Delta\Pi$, the validity of the physical meaning of M as associated with twice the surface energy is thus well demonstrated. As an aside, it is noted that, while the value of M in the figures is monotonically increasing with the number of cracks, the contribution from each crack is quite different in that the surface energy depends upon the crack size, the crack orientation, and also the stressed condition. Therefore, the relation between M and N does not necessarily correspond to a smooth curve.

The validity of the correspondence between the origin-independent M-integral ($=M_c$) and the surface energy for multiple curved cracks in a homogeneously stressed infinite rubbery field is numerically supported by the results from above. The results in Fig. 6 reveal that M is a monotonically increasing function of the crack density (i.e., the number of cracks in the specimen). Also, M essentially characterizes the effect due to interactions of the densely distributed cracks. These observed features indicate that application of M would facilitate the understanding of the degradation of materials induced by the above fracture factors. Thus, the M-integral furnishes the above information needed to obtain the fracture toughness, which could be a material characteristic and used to describe loss of material integrity caused by the irreversible evolution of curved cracks in a rubbery medium under the action of large deformation. This implies that, with the concept of M, a possible measure of damage parameters corresponding to characterization for evolution of the cracks can, therefore, be properly developed. Detailed discussions on such related issues are, however, beyond the scope of this study.



(a)



(b)

Fig. 6 The values of M ($=M_c$) versus the number of cracks (Problem 1). (*a*) $(\sigma^\infty,\beta)=(400\text{ Pa},60°)$, (*b*) $(\sigma^\infty,\beta)=(100\text{ Pa},90°)$. (Note: w=150 *m*, B=150 m.)

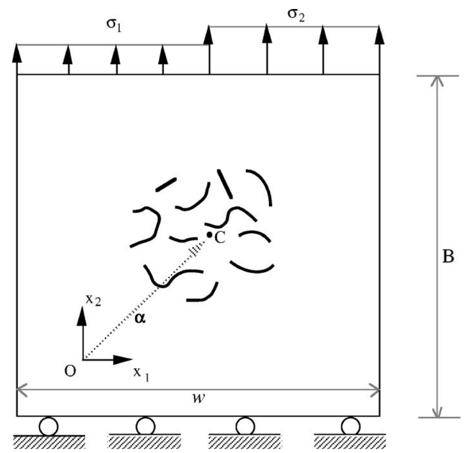
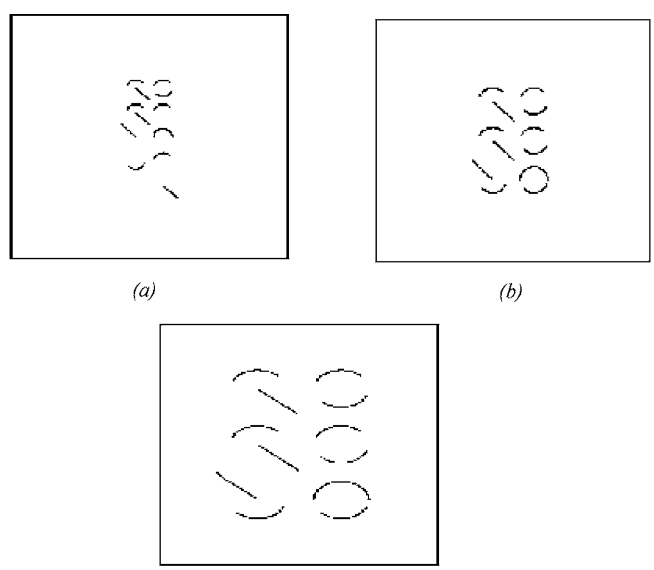


Fig. 7 A plane stress rubbery specimen containing *N* cracks under nonhomogeneously stressed state (Problem 2)

5.2 Problem 2—Structural Characteristic (Nonhomogeneously Stressed Condition). The applicability of M_c for use in nonhomogeneously stressed problems is illustrated in this example. To this end, we consider a plane stress rubbery specimen containing a number of densely (and randomly) distributed curved cracks and subjected to a nonuniform loading system, as shown in Fig. 7. Note that, as addressed in Appendix C, the relation between M_c and the surface energy is valid under the condition when the crack length is relatively small compared with the specimen size. In order to investigate the size effect, three geometric instances are presented in this problem. In the first instance (Fig. 8(*a*)), the crack size is relatively small compared with the size of the specimen so that the boundary effect of the specimen can be neglected. On the other hand, the specimen contains relatively



(a) *(b)* *(c)*

Fig. 8 Three geometric instances containing cracks of different relative sizes (Problem 2)

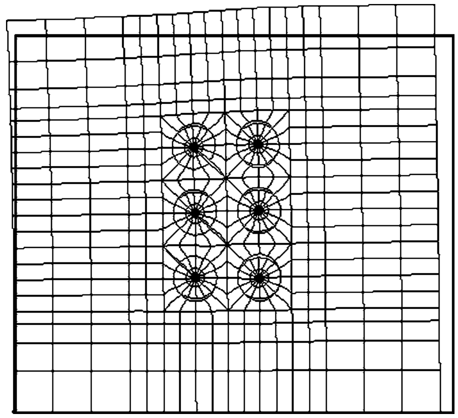


Fig. 9 The deformed mesh for the specimen in Fig. 8(*b*)

large curved cracks in the second the third instances (Figs. 8(*b*) and 8(*c*)). The deformed finite element mesh in Fig. 9 shows the response of the mid-sized specimen (Fig. 8(*b*)) under the action of $(\sigma_1, \sigma_2) = (15, 30)$Pa, where the undeformed configuration is depicted in heavy lines.

We first examine the characteristic of origin-dependence. By considering the case of five cracks (i.e., $N=5$), the results of M with respect to various coordinate origins for the three geometric instances are shown in Table 3. These results show that the values of M do depend upon $\boldsymbol{\alpha}$, as anticipated. Particularly, the M_c-integral (i.e., $\boldsymbol{\alpha}=(0,0)$) is evaluated by locating the origin of the coordinate system at the geometric center of all the crack tips.

Further, although not shown in detail here, the finite element results for M_c appear to be path-independent. Also, the accuracy of the computation is evident by good convergence observed in the study of finite element meshes.

Numerical results of M_c corresponding to a different number of cracks N for the three geometric instances (i.e., Figs. 8(*a*)–8(*c*)) are shown in Figs. 10(*a*)–10(*c*)), respectively. Also included in the figures is the surface energy $-\Delta\Pi$ corresponding to the formation of the curved cracks. It is observed that, for the instances of relatively small and mid-sized cracks (Figs. 10(*a*) and 10(*b*)), the values of the surface energy are well consistent with (half) those of M_c. This indicates that the surface energy can be appropriately extracted from the result of the M_c-integral. As to the instance when the crack lengths are relatively large (Fig. 10(*c*)), it is also evident that M_c and $-\Delta\Pi$ bear a very similar trend in characterizing the effect due to the increase of N. Although some deviations between the surface energy and (half) of M_c are observed (within 10%), the M_c-integral still furnishes reasonable approximations in predicting the surface energy.

The relation between M_c and $\Delta\Pi$ for a nonhomogeneously stressed body containing multiple curved cracks is evident by the numerical results presented above. It is particularly noted that this feature holds for problems containing strongly interacting cracks. The feasibility of the proposed integral is thus appropriately demonstrated. Based on this characteristic, we suggest M_c be used as a fracture parameter for describing the degradation of structural integrity caused by irreversible evolution of curved cracks in rubbery media. As an aside, note that the physical interpretation for

Table 3 The FE results of M wrt different α's for Problem 2 (10^1 Pa-m^2). (Note: $(\sigma_1, \sigma_2) = (15,30)$Pa, N=5.)

$\boldsymbol{\alpha}$	(0,0)m(,M$_c$)	(8.867,6.781)m	(93.87,9.178)m
M (Fig. 8(*a*))	1.749	1.231	−3.846
M (Fig. 8(*b*))	1.782	1.026	−6.539
M (Fig. 8(*c*))	2.005	0.746	−11.74

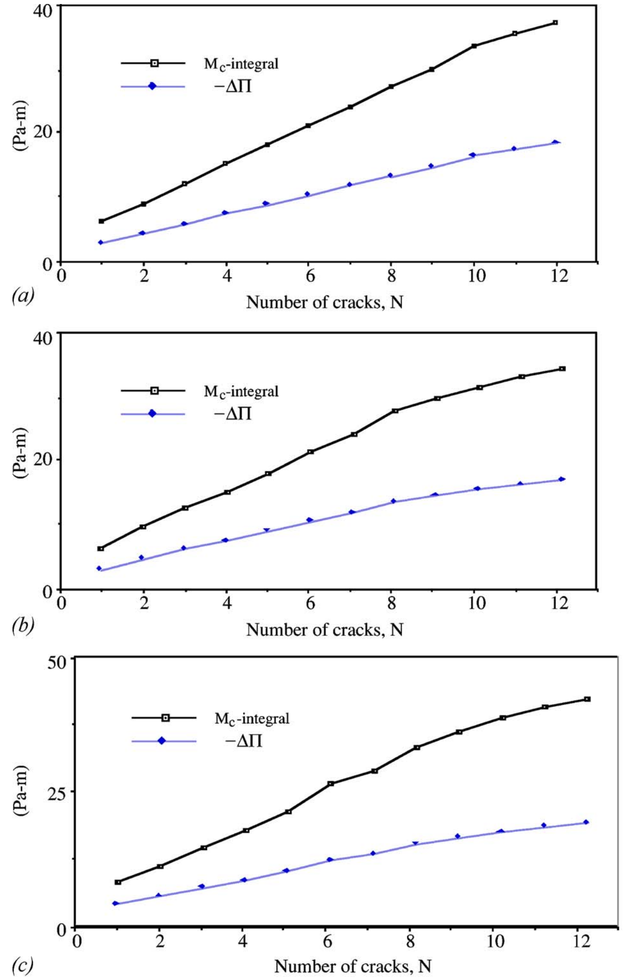


Fig. 10 The values of M_c versus the number of cracks for Figs. 8(*a*)–8(*c*), respectively (Problem 2)

M_c is exactly valid for problems containing relatively small cracks, and approximately satisfied for those with large cracks.

6 Conclusions

A fracture parameter M_c is presented for evaluating the surface energy associated with the formation of multiple curved cracks in a 2D rubbery body under the action of large deformation. The parameter is defined by performing the conventional M-integral along a contour enclosing all the cracks in the body, and originating the coordinate system at the geometric center of all the crack tips. Physically, the results generated by M_c can be interpreted in two different aspects. A well-known, but rarely occurred in reality, feature is corresponding to the energy release rate conjugate to expansion of each crack tip relative to the geometric center C. Alternatively, M_c is shown to be equivalent to twice the surface energy required for creation of all the curved cracks. The latter interpretation, being more applicable in practice, is verified both analytically and numerically. Note that, the relation is exactly valid when the crack length is small compared with the size of the specimen, and approximately satisfied for problems containing relatively large cracks. Based on this characteristic it is, therefore, suggested that M_c be practically used as a fracture parameter for describing the degradation of structural integrity caused by the irreversible evolution of curved cracks in rubbery media.

Due to path-independence, the M_c-integral can be performed along an arbitrary outer contour, which is chosen to be far from

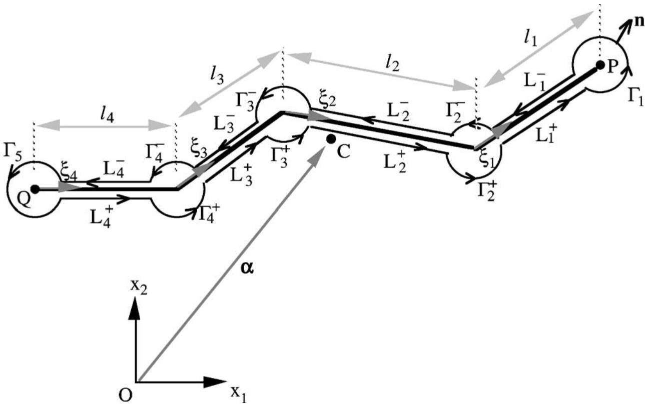


Fig. 11 The curved crack in Fig. 1 is approximated by a winding kinked crack with m segments (m=4 in this figure)

the crack tips. With this property, the complicated singular stress field in the near-tip areas can be avoided in the calculation.

For the special condition when the body corresponds to a homogeneously stressed infinite medium, the M-integral is independent of the coordinate origin. In this case, M is equivalent to M_c and can be used for characterizing the material fracture behavior associated with formation of the curved cracks.

Acknowledgment

This work has been partially supported by National Science Council Grant No. NSC 93-2211-E-008-017 to National Central University.

Appendix A: Relation Between M- and J_k-Integrals

In this section, the relation between M and J_k for a single curved crack is established. To this end, we first approximate the curved crack in Fig. 1 by considering a winding kinked configuration with two tips P and Q. The kinked crack consists of m segments, each of length l_i ($i=1,\ldots,m$) and with different orientations, as shown in Fig. 11 (m=4 in this figure). Still, with no loss of generality, the last segment of crack lying parallel to the x_1-direction of the coordinate system originating at O. The position vector at the geometric center C of the two tips is denoted $\boldsymbol{\alpha}$.

The M-integral associated with the whole crack is then defined by performing Eq. (1) along the counterclockwise closed contour D, which consists of 4 m curve segments as $\Gamma_1+L_1^-+[\sum_{i=2}^{m}(\Gamma_i^-+L_i^-)]+\Gamma_{m+1}+[\sum_{i=0}^{m-2}(L_{m-i}^{+}+\Gamma_{m-i}^{+})]+L_1^+$. Note that, by definition, the curve segments Γ_1, Γ_{m+1}, Γ_i^-, $\Gamma_i^+(i=2,\ldots,m)$ are shrunk onto the two crack tips (i.e., P and Q) and the kinked points, respectively. Also, L_i^- and L_i^+ ($i=1,\ldots,m$) are lying along the crack surfaces (this limiting condition is not shown in Fig. 11). When the crack surfaces are traction-free, it can be shown (Chang and Wu [15]) that M is related to the J_k-integrals as

$$M=\alpha_k\left[(J_k^P-J_k^Q)+\left(\sum_{i=1}^{m}\int_{L_i^++L_i^-}Wn_k\,d\xi_i\right)\right]+\frac{1}{2}\left(\sum_{i=1}^{m}l_{ik}\right)(J_k^P+J_k^Q)+\sum_{j=2}^{m}\left[\left(\sum_{i=1}^{j-1}l_{i1}-\frac{1}{2}\sum_{i=1}^{m}l_{i1}\right)\int_{L_j^++L_j^-}Wn_l\,d\xi_j\right]\qquad\text{(A1)}$$

where ξ_i is the local coordinate system lying along the ith segment and originated at the ith kinked point (particularly, ξ_m originated at tip Q), l_{ik} is the length of the projection of l_i in the x_k-direction. By observing Eq. (A1), it is noted that integration along the circular segments Γ_i^- and Γ_i^+ ($i=2,\ldots,m$) at the kinked points makes no contribution to the result of M. This vanishing feature explains why the kinked points are not included in positioning the geometric center C.

By taking the number of segments m into infinity, the kinked crack approaches asymptotically to the curved crack. Since the kinked points are not involved in evaluating M, Eq. (A2) can then be written as Eq. (3) in such a limiting case. Also, it is noted that the geometric center for the curved crack is positioned in the same manner as that for the kinked crack, i.e., at the center of the two tips P and Q.

Appendix B: Origin-Independence of M Under Uniform Remote Loads

The property of origin-independence for the M-integral has been established for problems containing multiple straight cracks in the authors″ previous work (Chang and Peng [16]). In this section, the validity of this property for curved cracks in homogeneously stressed rubbery media will be demonstrated.

For the infinite multi-cracked medium under uniform remote loads, by arbitrarily originating the coordinate system at point O and choosing a closed contour D_o enclosing all the N cracks (e.g., N=5 in Fig. 3), the M-integral (with respect to O) evaluated along D_o can be related to the J_k-integrals as

$$M=\sum_{r=1}^{N}\left\{\alpha_k^r\left[(J_k^{Pr}-J_k^{Qr})+\left(\int_{L^{r+}+L^{r-}}Wn_k^r\,d\xi^r\right)\right]+\frac{l_k^r}{2}(J_k^{Pr}-J_k^{Qr})+\left[\int_{L^{r+}+L^{r-}}\left(\xi_k^r-\frac{l_k^r}{2}\right)Wn_k^r\,d\xi^r\right]\right\}\qquad\text{(B1)}$$

where $\boldsymbol{\alpha}^r$ is the position vector for the rth crack (which is positioned at the geometric center of the two tips, P^r and Q^r) with respect to O, L^{r+} and L^{r-} are the curve segments along the rth crack surfaces, ξ^r is the local curvilinear coordinate system originated at and originated at tip Q^r, l_k^r is the length of the projection of l^r in the x_k-direction, and n_k^r is the outward unit vector normal to $L^{r+}+L^{r-}$. Note that, due to the property of path-independence, the integration can also be performed along another remote rectangular contour D^∞, with the value remaining unchanged.

Alternatively, by taking the geometric center of the first crack as the reference point and applying the concept of path-independence, we can rewrite Eq. (B1)

$$M=\sum_{r=2}^{N}\left\{(\alpha_k^r-\alpha_k^l)\left[(J_k^{Pr}-J_k^{Qr})+\left(\int_{L^{r+}+L^{r-}}Wn_k^r\,d\xi^r\right)\right]\right\}+\sum_{r=1}^{N}\left\{\frac{l_k^r}{2}(J_k^{Pr}+J_k^{Qr})+\left[\int_{L^{r+}+L^{r-}}\left(\xi_k^r-\frac{l_k^r}{2}\right)Wn_k^r\,d\xi^r\right]\right\}+\alpha_k^l\int_{D^\infty}\left[Wn_k-\sigma_{ij}n_j\left(\frac{\partial u_i}{\partial x_k}\right)\right]ds\qquad\text{(B2)}$$

In the first term on the RHS of Eq. (B2), the summation (from the second to the Nth cracks) accounts for the interaction between the first crack and the others. Further, since D^∞ is chosen to be far from the cracked region, the stresses and the strain energy density W are thus homogeneous along the contour. Consequently, the derivatives of the remote displacements with respect to the coordinate components, $u_{i,j}$, are also uniform. Therefore, the last term of Eq. (B2) (i.e., the integration along D^∞) intrinsically vanishes for both k=1 and 2 in that all the state variables are homogeneously distributed. This indicates that the result of M depends only on the relative locations of the center of each crack, rather than their absolute positions with respect to the origin O.

The origin-independent property can be further evidenced by arbitrarily taking another point (say, the geometric center of all the crack tips, C) as the reference. With this, Eq. (B1) can then be alternatively written as

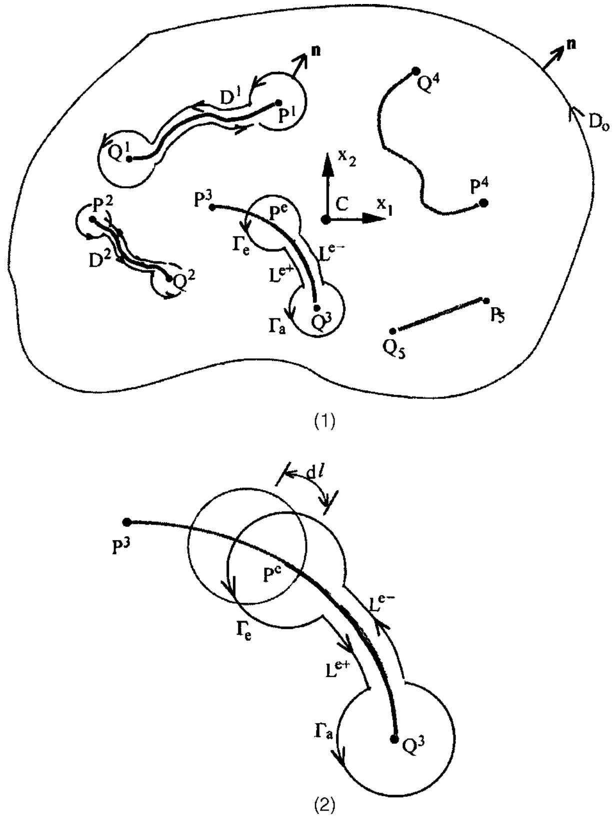


Fig. 12 (*a*) An intermediate state during evolution of the multi-cracked system shown in Fig. 2. (*b*) The enlarged portion of the extended configuration in the local near-tip area.

$$
\begin{aligned}
M = & \sum_{r=1}^{N}\left\{(\alpha_k{}^r - \alpha_k)\left[(J_k{}^{Pr} - J_k{}^{Qr}) + \left(\int_{L^{r+}+L^{r-}} Wn_k{}^r \, d\xi^r\right)\right]\right\} \\
& + \sum_{r=1}^{N}\left\{\frac{l_k{}^r}{2}(J_k{}^{Pr} + J_k{}^{Qr}) + \left[\int_{L^{r+}+L^{r-}}\left(\xi_k{}^r - \frac{l_k{}^r}{2}\right)Wn_k{}^r d\xi^r\right]\right\} \\
& + \alpha_k \int_{D^\infty}\left[Wn_k - \sigma_{ij}n_j\left(\frac{\partial u_i}{\partial x_k}\right)\right]ds
\end{aligned} \tag{B3}
$$

where $\boldsymbol{\alpha}$ is the position vector of C. Again, the last term of Eq. (B3) vanishes for both $k=1$ and 2. The first term, which takes summation from the first to the Nth cracks, accounts for the interaction between the N cracks. In fact, it can easily be shown that the first terms of Eqs. (B2) and (B3) are both equivalent. The property of origin-independence for Eq. (B2) is, therefore, demonstrated.

Appendix C: Physical Interpretation of M_c

The object here is to show that the result of the M_c-integral is equivalent to twice the surface energy required for formation of multiple curved cracks in a nonhomogeneously stressed rubbery medium under large elastic deformation.

By considering the 2D rubbery body containing N distributed curved cracks ($N=5$ in Fig. 2), we arbitrarily choose an intermediate state during the process of crack evolution, as shown in Fig. 12(a). At this stage, the crack system is instantly terminated at, say, point P^e of the third crack. In Fig. 12(a), the portions of the crack system that remain to be formed are depicted in dashed lines. The associated energy release rate G, due to a unit advance at the point P^e along the original direction of the crack segment, can be expressed as a decrease of the potential energy Π as

$$
G = -\frac{d\Pi}{dl} = \lim_{\Gamma_e \to 0}\int_{\Gamma_e}\left[Wn_k - T_i\left(\frac{\partial u_i}{\partial x_k}\right)\right]ds \tag{C1}
$$

where l is the original crack length, Γ_e is the integration contour enclosing and shrinking onto P^e, and x_k is the orientation tangential to the crack. By introducing a weighting function q, equation (C1) can further be rewritten as

$$
-\frac{d\Pi}{dl} = \int_{D_e}\left[Wn_k - T_i\left(\frac{\partial u_i}{\partial x_k}\right)\right]q\, ds + \int_{L^{e-}+L^{e+}} T_i\left(\frac{\partial u_i}{\partial x_k}\right)q\, ds \tag{C2}
$$

where $D_e = D^1 + D^2 + \Gamma_a + L^{e-} + \Gamma_e + L^{e+}$. Here, q is a sufficiently smooth scalar function taken as unity on Γ_e (i.e., onto P^e in the limiting case) and vanishing on D^1, D^2, and Γ_a. Also, note that the last term on the RHS of equation (C2) vanishes when the crack surfaces are traction-free.

In order to accommodate the extension of crack length by an amount of dl at P^e, we define an extended configuration relative to the intermediate state of the cracked system. In this extended state, all material points on (and within) Γ_e, L^{e-}, and L^{e+} translated by dl along the crack surfaces, with all other points remaining in their original positions. The enlarged portion of these two states in the local near-tip region is shown in Fig. 12(b). We also assume that there exists a one-to-one invertible coordinate mapping between these two states. Based on the mapping, a scalar function q can then be defined as

$$
q = \frac{dx_k}{dl} \tag{C3}
$$

By taking Eq. (C3) and noting that the geometric and boundary conditions remaining unchanged during the continuously varying process of crack evolution, we thus conduct integration of Eq. (C2) throughout the process of crack evolution as

$$
-\Delta\Pi = \int_{-L/2}^{L/2}\left[\int_{D_e}\left(Wn_k - T_i\frac{\partial u_i}{\partial x_k}\right) - \frac{dx_k}{dl}ds\right]dl \tag{C4}
$$

where L is the total crack length, i.e., $L=\Sigma_{r=1}^{N} l^r$. Since the measure of the crack system is taken with respect to the geometric center C, the integration is thus carried out by evaluation of G with respect to l from $-L/2$ to $L/2$. Furthermore, we note that the variation of energy release rate dG appears to be directly proportional to the variation of crack length dl in the crack system. Such linear relationship between dG and dl has been observed for rubbery problems containing a single straight crack (e.g.,[1]). In order to further investigate the effect of dl for problems with multiple curved cracks, a series of finite element computations is conducted in this research, where the values of G with respect to varying original crack length are calculated. Although not depicted in detail here, the validity of this linear relation is evident as long as the crack length is relatively small compared with the size of the specimen.

Based on the linear relation between dG and dl, the integration of Eq. (C4) then becomes

$$
-\Delta\Pi = \frac{1}{2}\int_{D}\left(Wn_k - T_i\frac{\partial u_i}{\partial x_k}\right)\Delta x_k\, ds \tag{C5}
$$

where $D=\Sigma_{r=1}^{N} D^r$ and Δx_k denotes the relative position of the integration point with respect to the geometric center C Equation (C5) indicates that the M_c-integral can thus be related to the energy difference $\Delta\Pi$ as

$$
M_C = -2\Delta\Pi \tag{C6}
$$

where $\Delta\Pi$ corresponds to the surface energy associated with creation of all the cracks.

References

[1] Rivlin, R. S., and Thomas, A. G., 1983, "The Incipient Characteristic Tearing Energy for an Elastomer Crosslinked Under Strain," J. Polym. Sci., Polym. Phys. Ed., **21**, pp. 1807–1814.

[2] Pidaparti, R. M. V., Yang, T. Y., and Soedel, W., 1989, "A Plane Stress Finite Element Method for the Prediction of Rubber Fracture," Int. J. Fract., **39**(4), pp. 255–268.

[3] Chang, J. H., and Yeh, J. B., 1998, "Calculation of *J*2-Integral due to a 2D Crack in Rubbery Materials," Eng. Fract. Mech., **59**(5), pp. 683–695.

[4] Knowles, J. K., and Sternberg, E., 1972, "On a Class of Conservation Laws in Linearized and Finite Elastostatics," Arch. Ration. Mech. Anal., **44**, pp. 187–211.

[5] Budiansky, B., and Rice, J. R., 1973, "Conservation Laws and Energy-Release Rates," ASME J. Appl. Mech., **40**, pp. 201–203.

[6] Eshelby, J. D., 1974, "Calculation of Energy Release Rate," *Prospects Fracture Mechanics*, G. C. Sih, H. C. Van Elst, and P. Broek, eds., Noordhoff, Groningen, pp. 69–84.

[7] Freund, L. B., 1978, "Stress-Intensity Factor Calculations Based on a Conservation Integral," Int. J. Solids Struct., **14**, pp. 241–250.

[8] Herrmann, A. G., and Herrmann, G., 1981, "On Energy Release Rates for a Plane Crack," ASME J. Appl. Mech., **48**, pp. 525–528.

[9] Eischen, J. W., and Herrmann, G., 1987, "Energy Release Rates and Related Balance Laws in Linear Elastic Defect Mechanics," ASME J. Appl. Mech., **54**, pp. 388–392.

[10] Seed, G. M., 1997, "The Boussinesq Wedge and the J_k, L, and M Integrals," Fatigue Fract. Eng. Mater. Struct., **20**(6), pp. 907–916.

[11] Chen, Y. H., 2001, "*M*-integral Analysis for Two-Dimensional Solids With Strongly Interacting Microcracks—Part I: In an Infinite Brittle Solid," Int. J. Solids Struct., **38**, pp. 3193–3212.

[12] Wang, D. F., and Chen, Y. H., 2002, "*M*-Integral Analysis for a Two-Dimensional Metal/Ceramic Bimaterial Solid With Extending Subinterface Microcracks," Arch. Appl. Mech., **72**, pp. 588–598.

[13] Tian, W. Y., 2004, "Microcracking Damage Analysis in Piezoelectric Bimaterial Solids," Int. J. Fatigue, **26**(8), pp. 873–888.

[14] 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.

[15] Chang, J. H., and Wu, W. H., 2005, "Evaluation of Surface Energy Associated With Formation of Multiple Kinked Cracks in Anisotropic Elastic Solids" (in preparation).

[16] Chang, J. H., and Peng, D. J., 2004, "On the Use of *M*-Integral for Rubbery Material Problems Containing Multiple Defects," ASCE J. Eng. Mech., **130**(5), pp. 589–598.

[17] Marusak, R. E., 1988, "Determining the Parameters of Elastic Strain Energy Functions." Master thesis, Univ. of Texas at Austin.

[18] Oden, J. T., and Carey, G. F., 1983, *Finite Elements IV: Mathematical Aspects*, Prentice-Hall, Englewood Cliffs, NJ.

The Lure of the Mean Axes

Leonard Meirovitch[1]
Department of Engineering Science and Mechanics,
Virginia Polytechnic Institute and State University,
MS 0219, Blacksburg, VA 24061
e-mail: lmeirovi@vt.edu

Ilhan Tuzcu
Department of Aerospace Engineering and Mechanics,
The University of Alabama,
Tuscaloosa, AL 35487-0280
e-mail: ituzcu@coe.eng.ua.edu

A variety of aerospace structures, such as missiles, spacecraft, aircraft, and helicopters, can be modeled as unrestrained flexible bodies. The state equations of motion of such systems tend to be quite involved. Because some of these formulations were carried out decades ago when computers were inadequate, the emphasis was on analytical solutions. This, in turn, prompted some investigators to simplify the formulations beyond all reasons, a practice continuing to this date. In particular, the concept of mean axes has often been used without regard to the negative implications. The allure of the mean axes lies in the fact that in some cases they can help decouple the system inertially. Whereas in the case of some space structures this may mean complete decoupling, in the case of missiles, aircraft, and helicopters the systems remain coupled through the aerodynamic forces. In fact, in the latter case the use of mean axes only complicates matters. With the development of powerful computers and software capable of producing numerical solutions to very complex problems, such as MATLAB and MATHEMATICA, there is no compelling reason to insist on closed-form solutions, particularly when undue simplifications can lead to erroneous results. [DOI: 10.1115/1.2338060]

1 Introduction

A variety of aerospace structures can be modeled as unrestrained flexible bodies. Examples of these are missiles, spacecraft, helicopters, and aircraft. In formulating the equations of motion, the methodology used depended largely on the preference of the investigators. Because some of these investigations were carried out decades ago when computers were inadequate, there has been a tendency to simplify the formulations to an undue extent. At the other extreme, in relatively recent days the tendency has been to rely on large computer codes written for purposes other than for the problem under consideration, resulting in very cumbersome and inefficient solutions.

Surprisingly, the literature on dynamics of unrestrained flexible bodies is not very abundant. Using essentially a Newtonian approach extended to finite bodies, Bisplinghoff and Ashley [1] derived scalar equations of motion for flexible missiles. Making a number of assumptions, they arrived at entirely decoupled equations for the rigid-body translations, greatly simplified equations for the rigid-body rotations and decoupled equations in terms of the "natural modes" for the elastic deformations. Among the factors contributing to the simplification of the equations of motion we note the use of principal axes as body axes, the assumption that the elastic deformations have a negligible effect on the mass moments of inertia and that the independent equations for the elastic deformation are the same as for an undamped multi-degree-of-freedom system. The scalar equations of motion of Bisplinghoff and Ashley can be expressed in the compact vector-matrix form

$$m(\dot{\mathbf{V}} + \tilde{\omega}\mathbf{V}) = \mathbf{F}, \quad J\dot{\boldsymbol{\omega}} + \tilde{\omega}J\boldsymbol{\omega} + \Phi^*\dot{\boldsymbol{\eta}} + \tilde{\omega}\Phi^*\boldsymbol{\eta} = \mathbf{M},$$

$$M(\dot{\boldsymbol{\eta}} + \Omega^2\boldsymbol{\xi}) = \mathbf{X} \tag{1}$$

where m is the total mass of the missile, $\mathbf{V}$ the velocity vector of the origin O of the body axes xyz, $\boldsymbol{\omega}$ the angular velocity vector of xyz, $\tilde{\omega}$ a skew symmetric matrix corresponding to $\boldsymbol{\omega}$ (Ref. [2]), $J=\text{diag}[J_{xx}\ J_{yy}\ J_{zz}]$ the inertia matrix, $\Phi^*=\int_m \tilde{r}\Phi dm$, in which $\tilde{r}$ is a skew symmetric matrix corresponding to the radius vector $\mathbf{r}$ from O to a typical mass element dm and Φ is a matrix of natural modes of vibration (referred to as shape functions in this paper, as explained later), $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are vectors of generalized displacements and velocities, respectively, $M=\text{diag}[M_{11}\ M_{22}\cdots M_{nn}]$ is a matrix of "generalized masses" and $\mathbf{F}$, $\mathbf{M}$, and $\mathbf{X}$ are a force vector acting on the whole missile, a moment vector about O acting on the whole missile and a vector of generalized forces acting on the generalized coordinates. We note that $\mathbf{F}$, $\mathbf{M}$, and $\mathbf{X}$ include all the forces acting on the missile. These include the very important aerodynamic forces and the thrust force during the powered flight. The effect of the aerodynamic forces is to introduce dissipative effects into the system, rendering the concept of natural modes meaningless, as natural modes exist only in undamped, and hence conservative systems. A comparison of Eqs. (1) with the rigorous equations of motion of an unrestrained flexible body given later in this paper will reveal glaring omissions.

Although the developments of Ref. [1] used a flexible missile as a model, the implication was that the developments applied equally well to flexible aircraft. Because the aerodynamic forces result from a body moving through air, these forces are most conveniently expressed in terms of local coordinates moving with the body, i.e., in terms of the body axes xyz, which explains why body axes were used in Ref. [1]. Body axes can be used as a reference frame in the case of flexible spacecraft as well. In the case of spacecraft, however, the translation of the origin of the reference frame, generally coinciding with the system center of mass, follows a predetermined orbit, so that there are no translational rigid-body degrees of freedom. Considering a spacecraft consisting of a rigid core with flexible appendages simulating antennas, Meirovitch and Nelson [3] defined the rotational motions and elastic deformations by means of a reference frame attached to the rigid core, and hence to the undeformed spacecraft, in essence using body axes. However, the use of body axes is not a uniform practice in the case of flexible spacecraft. Indeed, contending that in general spacecraft do not possess rigid cores, Canavin and Likins [4], argued in favor of using a "floating reference frame," referred to as a "Tisserand frame," or "mean axes frame." This argument is not very convincing, however, as there is no denying that all flexible bodies possess an undeformed state. At any rate, the mean axes were defined by Canavin and Likins by setting the internal angular momentum relative to the origin of the reference frame, made to coincide with the system mass center, to zero. In terms of our notation, this amounts to

[1]Author to whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received July 26, 2005; final manuscript received April 18, 2006. Review conducted by N. Sri Namachchivaya. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after fianl publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

$$\int_m \tilde{r}\mathbf{u}dm = \mathbf{0} \tag{2}$$

where $\tilde{r}$ is a skew symmetric matrix corresponding to the radius vector $\mathbf{r}$ and $\mathbf{u}$ is the elastic displacement vector. They correctly identified Eq. (2) as three scalar constraint equations to be imposed on $\mathbf{u}$. Then, they discuss at length the nature of the constraints and ways of satisfying them. The example model consists of two beams connected at one end to a hinge with a torsional spring and free at the other end. Because the model is not really representative of spacecraft and the discussion of the constraints satisfaction is rather laborious, we do not pursue the subject at this point but return to it later in this paper.

Whereas one can make a case for the use of mean axes for flexible spacecraft, provided the constraint equations are handled correctly, it should be pointed out that spacecraft operate in a much more benign environment than aerial vehicles. Indeed, they do not have to cope with one of the most significant problems in aerial vehicles, namely, the aerodynamic forces. In the case of aerial vehicles, any advantage that may accrue due to the inertial decoupling resulting from the use of mean axes is negated by the persistent coupling caused by the aerodynamic forces.

Worthy of notice due to the use of mean axes is a report by Dusto et al. [5], leading to a system of computer programs, named FLEXSTAB, for evaluating the stability of arbitrary flexible aircraft configurations in subsonic and supersonic flights. In addition to the common body axes, they used mean axes for deriving the equations of motion and "fluid axes" for describing the aerodynamic forces. They defined the mean axes by

$$\int_m \mathbf{u}dm = \mathbf{0}, \quad \int_m \tilde{r}\mathbf{u}dm = \mathbf{0} \tag{3}$$

and recognized that Eqs. (3) represented six constraint equations. To satisfy Eqs. (3), they used Lagrange's multipliers [2]. It should be noted that one of the disadvantages of Lagrange's multipliers method is that it introduces new unknowns into the equations of motion, in the case at hand six of them. There are two issues casting doubts on the validity of the results. The first is the fact that they used two different reference frames, one for the aircraft motions and one for the aerodynamic forces, without transforming from one to the other, and the second concerns the way Lagrange's multipliers method was applied. Of course, these issues are dwarfed by the fact that any decoupling resulting from the use of Eqs. (3) is only illusory in the presence of aerodynamic forces.

Using different reasoning, Eqs. (3) can lead to similar results in Ref. [1] as those obtained in Refs. [4,5]. Indeed, without any reference to mean axes, Ref. [1] interprets Eqs. (3) to mean that the natural modes of vibration are orthogonal to the translational and rotational rigid-body modes. In Refs. [4] they imply that the angular momentum due to the elastic deformations is zero, and in Ref. [5] that the linear and angular momenta are zero. Of course, they all imply that the system is undamped and that all motions, rigid and elastic, are small. This is hardly the case in general, as the aerodynamic forces are dissipative in nature and the rigid-body motions tend to be large.

The concept of mean axes was eagerly embraced by a variety of investigators in the belief that its use is likely to result in significant simplification of the equations of motion. In particular, it was believed that the use of mean axes would reduce coupling, thus resulting in independent equations for the rigid body translations, rigid body rotations, and elastic deformations. This belief turns out to be mistaken, as the disadvantages resulting from the requirement to satisfy the constraint equations, Eqs. (3), far outweigh any apparent advantages that might accrue from the elimination of some inertial coupling terms. Indeed, as pointed out above, there is no real decoupling, as the equations remain coupled through the aerodynamic forces. In fact, as shown later in this paper, the use of mean axes not only makes matters more complicated but in some cases it can render the formulation at odds with the physical reality.

Whatever advantages or disadvantages accrue from the use of mean axes, the procedures used in [4,5] are essentially correct. Unfortunately, the same cannot be said about the subsequent investigations, as at some later time the concept of mean axes was invoked but not really used. Indeed, some later investigators latched onto the notion of mean axes to drop most inertial coupling terms from the equations of motion while failing to enforce the constraints embodied by Eqs. (3). In fact, some of the investigators did not even give the definition of the mean axes.

Typical of investigators regarding the mean axes, Eqs. (3), as a vehicle to drop the coupling terms with impunity are Nydick and Friedmann [6] and Friedmann, McNamara, Thuruthimattam and Nydick [7], where the first is a conference presentation and the second is basically a journal version of the first. Indeed, professing to use the mean axes, when in fact they did not, as well as invoking a variety of other assumptions, and confining themselves to the case of steady level cruise, they obtained the greatly simplified equations of motion

$$m(\dot{V}_{0x} + qV_{0z}) = X - mg\sin\theta,$$

$$m(\dot{V}_{0z} - qV_{0x}) = Z + mg\cos\theta, \quad j_{yy}^{(0)}\dot{q} = M$$

$$M^g\ddot{\boldsymbol{\eta}} + C^g\dot{\boldsymbol{\eta}} + \left(K^g - \int_D \rho\Phi^T\tilde{\omega}^T\tilde{\omega}\Phi dD\right)\boldsymbol{\eta} = \int_D \rho\tilde{r}^T\tilde{\omega}^T\tilde{\omega}\Phi dD + \hat{Q} \tag{4}$$

in which m is the total aircraft mass, V_{0x} the forward velocity, V_{0z} the plunge velocity, q the pitch velocity, θ the pitch angle, j_{yy}^{0} the mass moment of inertia about the pitch axis y, assumed to be constant, X, Z, and M are associated forces and moment, including those due to aerodynamics, M^g, C^g, and K^g are "generalized" mass, damping and stiffness matrices, respectively, $\boldsymbol{\eta}$ is a vector of generalized elastic coordinates, Φ a "modal" matrix, $\tilde{\omega}$ a skew symmetric matrix derived from the angular velocity vector $\boldsymbol{\omega}$ $=[0\ q\ 0]^T$ and $\hat{Q}$ a generalized force density vector. Equations (4) are even simpler than they may seem, and we note that the first three are scalar equations and the fourth is a vector equation, as they can be solved independently for the rigid-body translations, rigid-body rotion, and elastic displacements. Indeed, first the third of Eqs. (4) can be solved for the pitch velocity q, and hence for the pitch angle θ, then the first and second can be solved for V_{0x} and V_{0z} and finally the fourth can be solved for $\boldsymbol{\eta}$. However, there are several problems with this proposition. In the first place, although the authors do list the definition of the mean axes, Eqs. (3), they make no attempt to enforce the constraints imposed by them, which can have disturbing implications. Moreover, there is no indication that the aerodynamic forces were ever expressed in terms of mean axes components, nor is there any hint that the aerodynamic forces keep the equations coupled, thus preventing independent solutions of the equations for the rigid body translations, rigid body rotations and elastic deformations.

It is clear from [6,7] that the objective is an aeroelasticity analysis, which calls for very simple equations of motion. Yet, the papers begin with a formulation capable of describing the dynamics and control of maneuvering flexible aircraft [8,9], a very complex problem, and proceed to strip away the elaborate formulation reducing it to Eqs. (4). In the process of reducing the equations of motion in terms of quasi-coordinates, the authors of [6,7] get tripped by the misuse of mean axes. It appears that the authors of [6,7] would have been better served by beginning directly with aeroelasticity equations rather than with the complex formulation of Ref. [8]. To indicate where [6,7], went wrong and to highlight some possible negative implication of the use of mean axes, the correct use of the mean axes is discussed later in this paper.

The use, and sometimes abuse, of the concept of mean axes is not confined to missiles, spacecraft, and aircraft. Indeed, abuse of the concept can be found in the case of helicopters as well. From a dynamicist's point of view, helicopters can be regarded broadly as consisting of three parts, the main rotor, the tail rotor and the fuselage, all acting together as a single dynamical system. Of these, the main rotor is by far the most critical part, although a helicopter could not function properly without a tail rotor. Indeed, in the absence of a tail rotor to counteract the angular momentum of the main rotor, the fuselage would spin in a sense opposite to that of the main rotor. The equations of motion for helicopters can be derived by modifying the approach used for aircraft, modifications made necessary by the fact that helicopters possess spinning rotors. However, this is not the approach used by Cribbs, Friedmann and Chiu [10], who consider the problem of a "coupled helicopter rotor/flexible fuselage aeroelasticity model" in a very unorthodox way as far as dynamics is concerned. In particular, *they concentrated on the fuselage alone* and invoked the use of mean axes to derive decoupled equations for the fuselage rigid-body translations, rigid-body rotations and elastic deformations, as follows:

$$m_f \ddot{\mathbf{x}}_{cm} = \mathbf{F}, \quad J_f \dot{\boldsymbol{\omega}} + \tilde{\omega} J_f \boldsymbol{\omega} = \mathbf{M}_{cm}, \quad M\ddot{\mathbf{q}}_e + K\mathbf{q}_e = \mathbf{Q} \tag{5}$$

where m_f is the total mass of the fuselage, $\mathbf{x}_{cm}$ the displacement vector of the fuselage mass center, $\mathbf{F}$ the resultant vector of all external forces acting on the fuselage, J_f the inertia matrix of the fuselage, $\boldsymbol{\omega}$ the angular velocity vector of the fuselage, $\mathbf{M}_{cm}$ the resultant vector of all external moments about the fuselage mass center, M and K are mass and stiffness matrices for the fuselage, $\mathbf{q}_e$ is a vector of generalized coordinates and $\mathbf{Q}$ a vector of generalized forces for the fuselage. Equations (5) are very strange in several respects. In the first place, they are for the fuselage alone, while insisting that the main rotor was coupled to the fuselage. To this end, they referred to some "equations of equilibrium" derived for the main rotor in a NASA CR dated 22 years earlier, without explaining how equations of equilibrium derived separately for the main rotor, equations not even given in Ref. [10], are coupled to equations of motion derived for the fuselage alone. Moreover, the tail rotor is not even mentioned. Finally, it is clear that Eqs. (5) are for a fictitious fuselage at best, as they were derived invoking the use of mean axes, when in fact the mean axes were not used at all.

Later in this paper, it is shown how the problem of helicopter dynamics is to be approached.

2 Proper Formulation for the Dynamics of Aerospace Structures

As discussed in the preceding section, the mean axes have been portrayed as a useful tool in the treatment of such diverse aerospace structures as missiles, spacecraft, aircraft, and helicopters. A closer examination of pertinent investigations, however, paints a different picture. Improper formulations of the equations of motion and/or misuse of the concept of mean axes do not inspire much confidence in the usefulness of the approach. In this section, we propose to contrast the approach based on misuse of mean axes with a proper formulation of the same problem.

From Meirovitch [8], the motions of a flexible body can be described by means of a reference frame xyz (Fig. 1) fixed in the *undeformed body* and known as *body*. Then, the rigid-body motions are defined as three translations and three rotations of the body axes relative to the inertial space XYZ and the elastic deformations as the displacements of points on the body relative to the body axes. When expressed in terms of components along the body axes, the rotational velocities are referred to as *quasi-velocities*, and the corresponding vector is denoted by $\boldsymbol{\omega}$. It is shown in Ref. [8] that, when expressed in terms of body-axes components, the translational velocities can also be treated as quasi-velocities; they are arranged in the vector $\mathbf{V}$. Note that, unlike ordinary velocities, quasi-velocities cannot be integrated to obtain displacements [2].

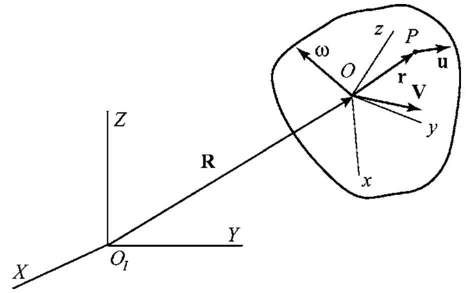


Fig. 1 Flexible body in space

Denoting by $\mathbf{u}$ and $\mathbf{v}$ the three-dimensional elastic displacement and elastic velocity vectors, respectively, where $\mathbf{u}$ and $\mathbf{v}$ are measured relative to the body axes xyz, it is shown by Meirovitch (Ref. 8) that the hybrid dynamical equations of motion in terms of quasi-coordinates can be written in the compact vector-matrix form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \mathbf{V}}\right) + \tilde{\omega}\frac{dL}{\partial \mathbf{V}} - C\frac{\partial L}{\partial \mathbf{R}} = \mathbf{F}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \boldsymbol{\omega}}\right) + \tilde{V}\frac{\partial L}{\partial \mathbf{V}} + \tilde{\omega}\frac{\partial L}{\partial \boldsymbol{\omega}} - (E^T)^{-1}\frac{\partial L}{\partial \boldsymbol{\theta}} = \mathbf{M}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \hat{T}}{\partial \mathbf{v}}\right) - \frac{\partial \hat{T}}{\partial \mathbf{u}} + \frac{\partial \hat{\mathcal{F}}}{\partial \mathbf{v}} + \mathcal{L}\mathbf{u} = \hat{\mathbf{U}} \tag{6}$$

where explicit provision was made for damping, in which L is the system Lagrangian, $\tilde{V}$ and $\tilde{\omega}$ are skew symmetric matrices derived from $\mathbf{V}$ and $\boldsymbol{\omega}$, respectively, C is a matrix of direction cosines between xyz and XYZ, $\mathbf{R}=[R_X\ R_Y\ R_Z]^T$ is the radius vector from the origin O_I of XYZ to the origin O of xyz, E is a matrix relating the symbolic angular velocity vector $\dot{\boldsymbol{\theta}}=[\dot{\theta}_1\ \dot{\theta}_2\ \dot{\theta}_3]^T$ to the angular quasi-velocity vector $\boldsymbol{\omega}$, $\hat{T}$ is the kinetic energy density of the body, $\hat{\mathcal{F}}$ is Rayleigh's dissipation density function [11] and $\mathcal{L}$ is a 3×3 matrix of stiffness differential operators. Moreover, $\mathbf{F}$, $\mathbf{M}$, and $\hat{\mathbf{U}}$ are generalized force vectors, which *must be expressed in terms of the same body axes components used to express the motion variables*. They can be obtained from the actual distributed force vector $\mathbf{f}(\mathbf{r},t)$ and the discrete force vectors $\mathbf{F}_i(t)$ acting at the points $\mathbf{r}=\mathbf{r}_i (i=1,2,\ldots,p)$ by means of the virtual work expression. Discrete forces can be treated as distributed by writing them in the form $\mathbf{F}_i(t)\delta(\mathbf{r}-\mathbf{r}_i)$, where $\delta(\mathbf{r}-\mathbf{r}_i)$ are spatial Dirac delta functions [11], so that the virtual work can be written in the form

$$\delta W = \int_D \left[\mathbf{f}^T(\mathbf{r},t) + \sum_{i=1}^{p} \mathbf{F}_i^T(t)\delta(\mathbf{r}-\mathbf{r}_i)\right]\delta\mathbf{R}_P^* dD \tag{7}$$

where $\delta\mathbf{R}_P^*$ is a virtual displacement vector whose expression can be obtained by considering the velocity vector of a typical point P in the body in terms of components along the body axes as follows:

$$\mathbf{v}_P = \mathbf{V} + \boldsymbol{\omega}\times(\mathbf{r}+\mathbf{u}) + \mathbf{v} = \mathbf{V} + (\tilde{r}+\tilde{u})^T\boldsymbol{\omega} + \mathbf{v} \tag{8}$$

where $\mathbf{r}$ is the radius vector from O to P and $\tilde{r}+\tilde{u}$ is the skew symmetric matrix derived from the vector $\mathbf{r}+\mathbf{u}$. Using the analogy with Eq. (8) with the term $\tilde{u}^T\boldsymbol{\omega}$ ignored as small compared to $\tilde{r}^T\boldsymbol{\omega}$, we obtain

$$\delta\mathbf{R}_P^* = \delta\mathbf{R}^* + \tilde{r}^T\delta\boldsymbol{\alpha} + \delta\mathbf{u} \tag{9}$$

which is the virtual displacement vector of point P (Fig. 1), in which $\delta\mathbf{R}^*$ is the virtual displacement vector of the origin of xyz, $\delta\boldsymbol{\alpha}$ is the virtual angular displacement vector of xyz and $\delta\mathbf{u}$ is the virtual elastic displacement vector of point P relative to xyz, all in terms of body-axes components. Inserting Eq. (9) into Eq. (7), we obtain

$$\delta W = \int_D \left[\mathbf{f}^T(\mathbf{r},t) + \sum_{i=1}^p \mathbf{F}_i^T(t)\delta(\mathbf{r}-\mathbf{r}_i)\right](\delta\mathbf{R}^* + \tilde{r}^T\delta\boldsymbol{\alpha} + \delta\mathbf{u})dD$$
$$= \mathbf{F}^T\delta\mathbf{R}^* + \mathbf{M}^T\delta\boldsymbol{\alpha} + \int_D \hat{\mathbf{U}}^T\delta\mathbf{u}dD \tag{10}$$

where

$$\mathbf{F} = \int_D \mathbf{f}dD + \sum_{i=1}^p \mathbf{F}_i, \quad \mathbf{M} = \int_D \tilde{r}\mathbf{f}dD + \sum_{i=1}^p \tilde{r}_i\mathbf{F}_i,$$
$$\hat{\mathbf{U}} = \mathbf{f} + \sum_{i=1}^p \mathbf{F}_i\delta(\mathbf{r}-\mathbf{r}_i) \tag{11}$$

are the desired generalized forces.

Closed-form solutions of hybrid sets of differential equations describing the dynamics of flexible aircraft is not within the state of the art, so that one must be content with approximate solutions. This implies invariably spatial discretization of the elastic variables [11]. We consider spatial discretization of the system by assuming that the elastic variables can be approximated to a reasonable degree of accuracy by series of space-dependent shape functions multiplied by time-dependent generalized coordinates, as follows:

$$\mathbf{u} = \Phi\boldsymbol{\xi}, \quad \mathbf{v} = \Phi\boldsymbol{\eta} \tag{12}$$

where $\Phi=\Phi(\mathbf{r})$ is a $3\times n$ matrix of shape functions [11], in which n is the number of elastic degrees of freedom, $\boldsymbol{\xi}$ is an n-vector of generalized coordinates and $\boldsymbol{\eta}$ is an n-vector of generalized velocities. A common misconception is to refer to the entries of Φ as "modes" when in fact they are merely shape functions. Of course, one can always try to choose the shape functions as the eigenfunctions of a somewhat related system, but that does not make them modes. It should be noted that the system described by Eqs. (6) is nonlinear and subject to viscous damping. Moreover, in the case in which Eqs. (6) represent the equations of motion for a flexible aircraft, $\mathbf{F}$, $\mathbf{M}$, and $\hat{\mathbf{U}}$ include aerodynamic forces. As a result, even after linearization, any modes are likely to be complex, whereas the entries of Φ are real functions. In the spirit of Rayleigh-Ritz, it is shown by Meirovitch [11] that any set of functions capable of describing any possible elastic deformation of the system to a given desired degree of accuracy represents an acceptable set of shape functions. However, the use of shape functions represents a mere spatial discretization process resulting in a set of ordinary differential equations.

Before discretizing Eqs. (6), we consider Eqs. (8) and (12) and write the two forms of the kinetic energy

$$T = \frac{1}{2}\int_D \rho\mathbf{v}_P^T\mathbf{v}_P dD = \frac{1}{2}m\mathbf{V}^T\mathbf{V} + \mathbf{V}^T\tilde{S}^T\boldsymbol{\omega} + \mathbf{V}^T\int_D \rho\mathbf{v}dD$$
$$+ \boldsymbol{\omega}^T\int_D \rho(\tilde{r}+\tilde{u})\mathbf{v}dD + \frac{1}{2}\boldsymbol{\omega}^T J\boldsymbol{\omega}$$
$$+ \frac{1}{2}\int_D \rho\mathbf{v}^T\mathbf{v}dD$$
$$\cong \frac{1}{2}m\mathbf{V}^T\mathbf{V} + \mathbf{V}^T\tilde{S}^T\boldsymbol{\omega} + \mathbf{V}^T\bar{\Phi}\boldsymbol{\eta}$$
$$+ \boldsymbol{\omega}^T\int_D \rho(\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}})\Phi dD\boldsymbol{\eta} + \frac{1}{2}\boldsymbol{\omega}^T J\boldsymbol{\omega}$$
$$+ \frac{1}{2}\boldsymbol{\eta}^T\bar{\bar{\Phi}}\boldsymbol{\eta} \tag{13}$$

where

$$\bar{\Phi} = \int_D \rho\Phi dD, \quad \tilde{S} = \int_D \rho(\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}})dD,$$
$$J = \int_D \rho(\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}})(\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}})^T dD, \quad \bar{\bar{\Phi}} = \int_D \rho\Phi^T\Phi dD \tag{14}$$

in which $\tilde{S}$ and J are the matrix of first moments of inertia and the inertia matrix, respectively, of Rayleigh's dissipation function

$$\mathcal{F} = \int_D \hat{\mathcal{F}}dD = \frac{1}{2}\int_D c\mathbf{v}^T\mathbf{v}dD \cong \frac{1}{2}\boldsymbol{\eta}^T\int_D c\Phi^T\Phi dD\boldsymbol{\eta} = \frac{1}{2}\boldsymbol{\eta}^T C\boldsymbol{\eta} \tag{15}$$

where

$$C = \int_D c\Phi^T\Phi dD \tag{16}$$

is a damping matrix, in which c is a damping density function, and of the strain energy

$$V_{\text{str}} = \frac{1}{2}\int_D \mathbf{u}^T\mathcal{L}\mathbf{u}dD \cong \frac{1}{2}\boldsymbol{\xi}^T\int_D \Phi^T\mathcal{L}\Phi dD\boldsymbol{\xi} = \frac{1}{2}\boldsymbol{\xi}^T K\boldsymbol{\xi} \tag{17}$$

where

$$K = \int_D \Phi^T\mathcal{L}\Phi dD \tag{18}$$

is a stiffness matrix. Both C and K are symmetric. Then assuming that the Lagrangian does not depend on $\mathbf{R}$ and $\boldsymbol{\theta}$, which is true for flying aircraft, the discrete counterpart of Eqs. (6) are simply

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\mathbf{V}}\right) + \tilde{\omega}\frac{\partial L}{\partial\mathbf{V}} = \mathbf{F}, \quad \frac{d}{dt}\left(\frac{\partial L}{\partial\boldsymbol{\omega}}\right) + \tilde{V}\frac{\partial L}{\partial\mathbf{V}} + \tilde{\omega}\frac{\partial L}{\partial\boldsymbol{\omega}} = \mathbf{M}$$
$$\frac{d}{dt}\left(\frac{\partial L}{\partial\boldsymbol{\eta}}\right) - \frac{\partial T}{\partial\boldsymbol{\xi}} + C\boldsymbol{\eta} + K\boldsymbol{\xi} = \mathbf{X} \tag{19}$$

in which, from Eq. (13)

$$\frac{\partial T}{\partial\boldsymbol{\xi}} = -\bar{\Phi}^T\tilde{V}^T\boldsymbol{\omega} + \int_D \rho\Phi^T\widetilde{\Phi\boldsymbol{\eta}}dD\boldsymbol{\omega} - \int_D \rho\Phi^T\tilde{\omega}^2(\mathbf{r}+\Phi\boldsymbol{\xi})dD \tag{20}$$

and

$$\mathbf{X} = \int_D \Phi^T\hat{\mathbf{U}}dD \tag{21}$$

is the discretized generalized elastic force vector.

Finally, including some obvious kinematical identities and carrying out the indicated operations, we obtain the discretized set of state equations

$$\dot{\mathbf{R}} = C^T\mathbf{V}, \quad \dot{\boldsymbol{\theta}} = E^{-1}\boldsymbol{\omega}, \quad \dot{\boldsymbol{\xi}} = \boldsymbol{\eta}$$

$$m\dot{\mathbf{V}} + \tilde{S}^T\dot{\boldsymbol{\omega}} + \bar{\Phi}\dot{\boldsymbol{\eta}} = \left(2\int_D \rho\widetilde{\Phi\boldsymbol{\eta}}dD + m\tilde{V} + \tilde{\omega}\tilde{S}\right)\boldsymbol{\omega} + \mathbf{F}$$

$$\tilde{S}\dot{\mathbf{V}} + J\dot{\boldsymbol{\omega}} + \int_D \rho(\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}})\Phi dD\,\dot{\boldsymbol{\eta}} = \left\{\int_D \rho[\widetilde{\Phi\boldsymbol{\eta}}(\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}}) + (\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}})\widetilde{\Phi\boldsymbol{\eta}}]dD + \tilde{V}\tilde{S} + \widetilde{\tilde{S}\mathbf{V}} - \tilde{\omega}J\right\}\boldsymbol{\omega} + \tilde{\omega}\int_D \rho(\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}})^T\Phi\boldsymbol{\eta}dD + \mathbf{M}$$

$$\bar{\Phi}^T\dot{\mathbf{V}} + \int_D \rho\Phi^T(\tilde{r}+\widetilde{\Phi\boldsymbol{\xi}})^T dD\dot{\boldsymbol{\omega}} + \bar{\bar{\Phi}}\dot{\boldsymbol{\eta}} = -\bar{\Phi}^T\tilde{V}^T\boldsymbol{\omega} - \int_D \rho\Phi^T\tilde{\omega}^2(\mathbf{r} + \Phi\boldsymbol{\xi})dD - 2\int_D \rho\Phi^T\widetilde{\Phi\boldsymbol{\eta}}^T dD\boldsymbol{\omega} - C\boldsymbol{\eta} - K\boldsymbol{\xi} + \mathbf{X} \quad (22)$$

Next, we address the problem of helicopter dynamics. To this end, we consider the typical helicopter configuration shown in Fig. 2 and assume that the main rotor consists of n_m equally spaced blades and the tail rotor consists of n_τ equally spaced blades. Then, by analogy with Eqs. (6) and using the notation of Fig. 2, the helicopter equations of motion can be expressed in the generic form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \mathbf{V}_O}\right) + \tilde{\omega}_O\frac{\partial L}{\partial \mathbf{V}_O} - C_O\frac{\partial L}{\partial \mathbf{R}_O} = \mathbf{F}_O$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \boldsymbol{\omega}_O}\right) + \tilde{\mathbf{V}}_O\frac{\partial L}{\partial \mathbf{V}_O} + \tilde{\omega}_O\frac{\partial L}{\partial \boldsymbol{\omega}_O} - (E_O^T)^{-1}\frac{\partial L}{\partial \boldsymbol{\theta}_O} = \mathbf{M}_O$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \hat{T}_f}{\partial \mathbf{v}_f}\right) - \frac{\partial \hat{T}_f}{\partial \mathbf{u}_f} + \frac{\partial \hat{\mathcal{F}}_f}{\partial \mathbf{v}_f} + \mathcal{L}_f\mathbf{u}_f = \hat{\mathbf{U}}_f$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \hat{T}_{mi}}{\partial \mathbf{v}_{mi}}\right) - \frac{\partial \hat{T}_{mi}}{\partial \mathbf{u}_{mi}} + \frac{\partial \hat{\mathcal{F}}_{mi}}{\partial \mathbf{v}_{mi}} + \mathcal{L}_{mi}\mathbf{u}_{mi} = \hat{\mathbf{U}}_{mi}, \quad i = 1,2,\ldots,n_m$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \hat{T}_{\tau j}}{\partial \mathbf{v}_{\tau j}}\right) - \frac{\partial \hat{T}_{\tau j}}{\partial \mathbf{u}_{\tau j}} + \frac{\partial \hat{\mathcal{F}}_{\tau j}}{\partial \mathbf{v}_{\tau j}} + \mathcal{L}_{\tau j}\mathbf{u}_{\tau j} = \hat{\mathbf{U}}_{\tau j}, \quad j = 1,2,\ldots,n_\tau \quad (23)$$

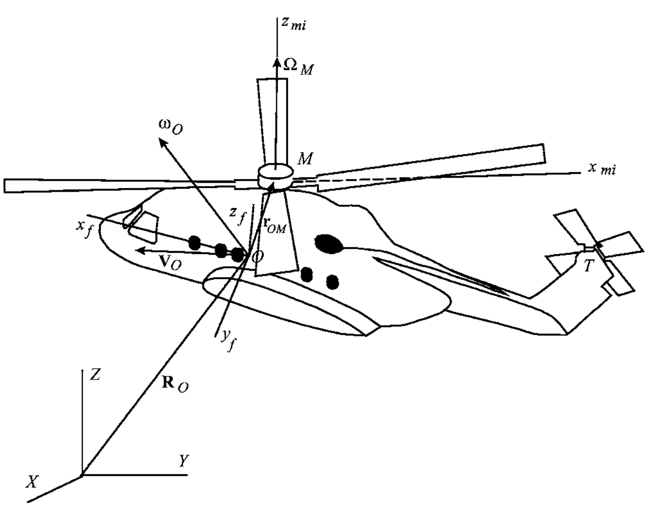


Fig. 2 Flexible helicopter

The derivation of explicit equations of motion would require a great deal of perserverance and patience, and is beyond the scope of this paper. To develop an appreciation of the nature of the equations, however, we will outline some of the steps involved in their derivation. The kinetic energy has the general expression

$$T = \frac{1}{2}\int_{m_f}\mathbf{V}_f^T\mathbf{V}_f dm_f + \frac{1}{2}\sum_{i=1}^{n_m}\int_{m_i}\mathbf{V}_{mi}^T\mathbf{V}_{mi}dm_i + \frac{1}{2}\sum_{j=1}^{n_\tau}\int_{m_j}\mathbf{V}_{\tau j}^T\mathbf{V}_{\tau j}dm_j \quad (24)$$

where, following an orderly kinematical procedure, the velocity vectors of the individual components are given by

$$\mathbf{V}_f(\mathbf{r}_f,t) = \mathbf{V}_O(t) + [\tilde{r}_f + \tilde{u}_f(\mathbf{r}_f,t)]^T\boldsymbol{\omega}_O(t) + \mathbf{v}_f(\mathbf{r}_j,t)$$

$$\mathbf{V}_{mi}(\mathbf{r}_{mi},t) = C_{mi}\mathbf{V}_M + [\tilde{r}_{mi} + \tilde{u}_{mi}(\mathbf{r}_{mi},t)]^T(C_{mi}\boldsymbol{\omega}_O + \Omega_M) + \mathbf{v}_{mi}(\mathbf{r}_{mi},t), \quad i = 1,2,\ldots,n_m$$

$$\mathbf{V}_{\tau j}(\mathbf{r}_{\tau j},t) = C_{\tau j}\mathbf{V}_T + [\tilde{r}_{\tau j} + \tilde{u}_{\tau j}(\mathbf{r}_{\tau j},t)]^T(C_{\tau j}\boldsymbol{\omega}_O + \Omega_T) + \mathbf{v}_{\tau j}(\mathbf{r}_{\tau j},t), \quad j = 1,2,\ldots,n_\tau \quad (25)$$

in which

$$\mathbf{V}_M(t) = \mathbf{V}_f(\mathbf{r}_{OM},t) = \mathbf{V}_O(t) + [\tilde{r}_{OM} + \tilde{u}_f(\mathbf{r}_{OM},t)]^T\boldsymbol{\omega}_O(t) + \mathbf{v}_f(\mathbf{r}_{OM},t)$$

$$\mathbf{V}_T(t) = \mathbf{V}_f(\mathbf{r}_{OT},t) = \mathbf{V}_O(t) + [\tilde{r}_{OT} + \tilde{u}_f(\mathbf{r}_{OT},t)]^T\boldsymbol{\omega}_O(t) + \mathbf{v}_f(\mathbf{r}_{OT},t) \quad (26)$$

are the velocity vectors of the main rotor hub M and tail rotor hub T, obtained by evaluating $\mathbf{V}_f$ at M and T, respectively. Moreover, C_{mi} and $C_{\tau j}$ are matrices of direction cosines between the main rotor blade body axes $x_{mi}y_{mi}z_{mi}$ and the fuselage body axes $x_f y_f z_f$ and the tail rotor blade body axes $x_{\tau j}y_{\tau j}z_{\tau j}$ and $x_f y_f z_f$ due to the spin of the rotors; both C_{mi} and $C_{\tau j}$ depend explicitly on time. The kinetic energy is obtained by inserting Eqs. (25) and (26) into Eq. (24) and carrying out the indicated operations. A cursory examination of Eqs. (25) and (26) will reveal that the fuselage, main rotor and tail rotor are all inertially coupled and so are the equations of motion. Furthermore, it is futile to look for mean axes capable of changing this fact. Clearly, the use of mean axes for the fuselage alone, which is the least critical part of a helicopter, cannot be justified.

Although we expressed the fuselage stiffness in terms of a stiffness operator matrix $\mathcal{L}_f$, this was merely symbolic because it is not feasible to generate a differential operator for such a complex structure as the fuselage. In practice, it is necessary to express the fuselage stiffness, in the context of a spatial discretization process, by means of a stiffness matrix generated by the finite element method. A typical main rotor blade can be modeled as a thin beam undergoing torsion about the longitudinal axis x_{mi} and bending about axes y_{mi} and z_{mi}. Denoting the displacement vector for blade m_i by $\mathbf{u}_{mi}=[\psi_{xmi}\ u_{ymi}\ u_{zmi}]^T$, the corresponding stiffness operator matrix can be shown to have the form [11]

$$\mathcal{L}_{mi} = \begin{bmatrix} -\frac{\partial}{\partial x_{mi}}\left[GJ_{mi}(x_{mi})\frac{\partial}{\partial x_{mi}}\right] & 0 & 0 \\ 0 & \frac{\partial^2}{\partial x_{mi}^2}\left[EI_{zmi}(x_{mi})\frac{\partial^2}{\partial x_{mi}^2}\right] & 0 \\ 0 & 0 & \frac{\partial^2}{\partial x_{mi}^2}\left[EI_{ymi}(x_{mi})\frac{\partial^2}{\partial x_{mi}^2}\right] - \frac{\partial}{\partial x_{mi}}\left[P_{mi}(x_{mi})\frac{\partial}{\partial x_{mi}}\right] \end{bmatrix} \quad (27)$$

where

$$P_{mi}(x_{mi}) = \int_{x_{mi}}^{L_{mi}} m_i(\xi)(C_{mi}\boldsymbol{\omega}_O + \Omega_M)^2_{zmi}\xi\, d\xi \quad (28)$$

is the axial force on blade m_i due to the centrifugal force caused by the spin of the main rotor hub. The operator matrix $\mathcal{L}_{\tau j}$ for the tail rotor can be obtained from Eqs. (27) and (28) in an analogous fashion.

Some of the other quantities, such as Rayleigh's dissipation function densities $\hat{\mathcal{F}}_f$, $\hat{\mathcal{F}}_{mi}$, and $\hat{\mathcal{F}}_{\tau j}$ and the generalized forces $\mathbf{F}_0$, $\mathbf{M}_0$, $\hat{\mathbf{U}}_f$, $\hat{\mathbf{U}}_{mi}$, and $\hat{\mathbf{U}}_{\tau j}$ can be obtained by analogy with those for the aircraft, where the forces include the aerodynamic forces, which are much more complicated than those for aircraft. In this regard, it must be pointed out that the degree of complexity increases significantly from hover to forward flight. Indeed, in forward flight the blade velocity due to the hub rotation adds to the fuselage velocity during half of the rotation and subtracts from it during the other half. To compensate for this, the blade is made to pitch accordingly.

Generic equations of motion for whole flexible helicopters can be obtained by inserting Eqs. (24)–(28) into Eqs. (23) and carrying out the indicated operations, which would involve a great deal of symbolic manipulations, well in excess of those for flexible aircraft. It is not difficult to see that the resulting equations are likely to be extremely complex. Contrasting Eqs. (23)–(28) with the formulation of Ref. [10], Eqs. (5) of the present paper, one concludes that the formulation of Ref. [10] is badly flawed, and no sensible simplification of Eqs. (23)–(28) would reduce the equations of motion to an extent that would make them resemble Eqs. (5), not even vaguely. Hence, the contention that Eqs. (5) reflect a "Coupled Helicopter Rotor/Flexible Fuselage Aeroelastic Model…," as the title of Ref. [10] implies, is more than questionable.

3 The Use of Mean Axes

In deriving Eqs. (22), a reference frame embedded in the undeformed body was used to define the rigid-body and elastic motions, as well as the forces, moments and distributed forces. This choice seems only natural. As can be expected, the equations for the rigid-body translations, rigid-body rotations and elastic deformations are all coupled. There seems to be a belief that a different choice of reference frame is able to reduce the coupling, and hence the complexity of the formulation. We wish to examine this proposition.

The inertial terms can be simplified to some extent by choosing the body axes as the principal axes with the origin at the mass center. In this case, we have

$$\int_D \rho\mathbf{r}dD = \mathbf{0}, \quad \int_D \rho\tilde{r}\tilde{r}^T dD = J^{(0)} \quad (29)$$

where $J^{(0)}$ represents the diagonal matrix of the principal moments of inertia of the undeformed body. Perhaps more extensive simplifications can be achieved, at least at first sight, by using the mean axes, defined by

$$\int_D \rho\mathbf{u}dD \cong \int_D \rho\Phi\boldsymbol{\xi}dD = \bar{\Phi}\boldsymbol{\xi} = \mathbf{0},$$

$$\int_D \rho\mathbf{r}\times\mathbf{u}dD = \int_D \rho\tilde{r}\mathbf{u}dD \cong \int_D \rho\tilde{r}\Phi\boldsymbol{\xi}dD = \Phi^*\boldsymbol{\xi} = \mathbf{0} \quad (30)$$

Inserting Eqs. (29) and (30) into the second half of Eqs. (22), the dynamical part of the state equations reduces to

$$m\dot{\mathbf{V}} + \bar{\Phi}\dot{\boldsymbol{\eta}} = \left(2\int_D \rho\widetilde{\Phi\boldsymbol{\eta}}dD + m\tilde{V}\right)\boldsymbol{\omega} + \mathbf{F}$$

$$J\dot{\boldsymbol{\omega}} + \int_D \rho\widetilde{\Phi\boldsymbol{\xi}}\Phi dD\,\dot{\boldsymbol{\eta}} = \left\{\int_D \rho[\widetilde{\Phi\boldsymbol{\eta}}(\tilde{r} + \widetilde{\Phi\boldsymbol{\xi}}) + (\tilde{r} + \widetilde{\Phi\boldsymbol{\xi}})\widetilde{\Phi\boldsymbol{\eta}}]dD - \tilde{\omega}J\right\}\boldsymbol{\omega} + \tilde{\omega}\int_D \rho(\tilde{r} + \widetilde{\Phi\boldsymbol{\xi}})^T\Phi dD\,\boldsymbol{\eta} + \mathbf{M}$$

$$\bar{\Phi}^T\dot{\mathbf{V}} + \int_D \rho\Phi^T\widetilde{\Phi\boldsymbol{\xi}}^T dD\,\dot{\boldsymbol{\omega}} + \bar{\bar{\Phi}}\dot{\boldsymbol{\eta}} = \bar{\Phi}^T\tilde{V}^T\boldsymbol{\omega} - \int_D \rho\Phi^T\tilde{\omega}^2(\mathbf{r} + \Phi\boldsymbol{\xi})dD - 2\int_D \rho\Phi^T\widetilde{\Phi\boldsymbol{\eta}}^T dD\boldsymbol{\omega} - C\boldsymbol{\eta} - K\boldsymbol{\xi} + \mathbf{X} \quad (31)$$

Contrasting Eqs. (31) with the simplified equations of Refs. [6,7], Eqs. (4) of the present paper, we conclude that Eqs. (4) do not describe any real aircraft but some fictitious one that does not exist, so that any analysis based on Eqs. (4) is an exercise in futility.

4 Implications of Constraints Enforcement

The mean axes defined by Eqs. (30) represent an abstraction in the sense that the axes are not as easy to visualize as the body axes embedded in the undeformed body, let alone to determine them. More important, however, is the fact that, from a dynamical point of view, Eqs. (30) represent *constraints* imposed on the displacement vector function $\mathbf{u}(\mathbf{r},t)$, or the generalized coordinate vector $\boldsymbol{\xi}$. Indeed, Eqs. (30) represent *six equations of constraint*. Less appreciated and seldom mentioned, if ever, is the fact that *these constraints must be enforced*. Enforcement of constraints has certain implications well-worth exploring. There is nothing to be gained by enforcing the constraints on Eqs. (4) because these equations are flawed. Indeed, we can only gain insight into the problem by enforcing the constraints on valid equations, namely, Eqs. (31).

For convenience, we confine ourselves to the discrete version of Eqs. (30); they represent constraint equations to be satisfied by the n-dimensional generalized coordinate vector $\boldsymbol{\xi}$. The implication is that the components of $\boldsymbol{\xi}$ are not independent but subject to the six scalar equations

$$\sum_{j=1}^{n} \bar{\Phi}_{ij}\xi_j = 0, \quad \sum_{j=1}^{n} \Phi^*_{ij}\xi_j = 0, \quad i = 1,2,3 \tag{32}$$

The six constraint equations, Eqs. (32), can be enforced by means of Lagrange's multipliers (Refs. [2,5]), a somewhat tedious process. They can be more conveniently enforced by means of a coordinate transformation reflecting the fact that only $n-6$ of the components $\xi_1, \xi_2, \ldots, \xi_n$ are independent. To this end, we rewrite Eqs. (32) in the form

$$A\boldsymbol{\xi}_{\text{ind}} + B\boldsymbol{\xi}_{\text{dep}} = \mathbf{0} \tag{33}$$

where

$$A = \begin{bmatrix} \bar{\Phi}_{11} & \bar{\Phi}_{12} & \cdots & \bar{\Phi}_{1,n-6} \\ \bar{\Phi}_{21} & \bar{\Phi}_{22} & \cdots & \bar{\Phi}_{2,n-6} \\ \cdots & \cdots & \cdots & \cdots \\ \Phi^*_{31} & \Phi^*_{32} & \cdots & \Phi^*_{3,n-6} \end{bmatrix}, \quad B = \begin{bmatrix} \bar{\Phi}_{1,n-5} & \bar{\Phi}_{1,n-4} & \cdots & \bar{\Phi}_{1n} \\ \bar{\Phi}_{2,n-5} & \bar{\Phi}_{2,n-4} & \cdots & \bar{\Phi}_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \Phi^*_{3,n-5} & \Phi^*_{3,n-4} & \cdots & \Phi^*_{3n} \end{bmatrix} \tag{34}$$

are $6\times(n-6)$ and 6×6 matrices and $\boldsymbol{\xi}_{\text{ind}}=[\xi_1\xi_2\ldots\xi_{n-6}]^T$ and $\boldsymbol{\xi}_{\text{dep}}=[\xi_{n-5}\xi_{n-4}\ldots\xi_n]^T$ are $(n-6)$-dimensional and six-dimensional vectors of independent and dependent components, respectively. Then, regarding the original n-dimensional vector $\boldsymbol{\xi}$ as a *constrained vector*, we can write

$$\boldsymbol{\xi} = \boldsymbol{\xi}_{\text{c}} = \begin{bmatrix} \boldsymbol{\xi}_{\text{ind}} \\ \boldsymbol{\xi}_{\text{dep}} \end{bmatrix} = \begin{bmatrix} I \\ -B^{-1}A \end{bmatrix} \boldsymbol{\xi}_{\text{ind}} = T\boldsymbol{\xi}_{\text{ind}} \tag{35}$$

in which

$$T = \begin{bmatrix} I \\ -B^{-1}A \end{bmatrix} \tag{36}$$

represents an $n\times(n-6)$ transformation matrix. Note that choosing the bottom six components of $\boldsymbol{\xi}_{\text{c}}$ as the dependent component was merely to demonstrate the procedure and is a valid choice only if B is nonsingular. Inserting Eq. (35) into Eqs. (31) and multiplying the third of the resulting equations on the left by T^T, we obtain

$$m\dot{\mathbf{V}} + \bar{\Phi}T\dot{\boldsymbol{\eta}}_{\text{ind}} = \left(2\int_D \rho\widetilde{\Phi T\boldsymbol{\eta}_{\text{ind}}}dD + m\tilde{V}\right)\boldsymbol{\omega} + \mathbf{F}$$

$$J\dot{\boldsymbol{\omega}} + \int_D \rho\widetilde{\Phi T\boldsymbol{\xi}_{\text{ind}}}\Phi dDT\dot{\boldsymbol{\eta}}_{\text{ind}} = \left\{\int_D \rho[\widetilde{\Phi T\boldsymbol{\eta}_{\text{ind}}}(\tilde{r} + \widetilde{\Phi T\boldsymbol{\xi}_{\text{ind}}}) + (\tilde{r} + \widetilde{\Phi T\boldsymbol{\xi}_{\text{ind}}})\widetilde{\Phi T\boldsymbol{\eta}_{\text{ind}}}]dD - \tilde{\omega}J\right\}\boldsymbol{\omega} + \tilde{\omega}\int_D \rho(\tilde{r} + \widetilde{\Phi T\boldsymbol{\xi}_{\text{ind}}})^T\Phi dDT\boldsymbol{\eta}_{\text{ind}} + \mathbf{M}$$

$$T^T\bar{\Phi}^T\dot{\mathbf{V}} + T^T\int_D \rho\Phi^T\widetilde{\Phi T\boldsymbol{\xi}}^T_{\text{ind}}dD\dot{\boldsymbol{\omega}} + \bar{\bar{\Phi}}^*\dot{\boldsymbol{\eta}}_{\text{ind}} = -T^T\bar{\Phi}^T\tilde{V}^T\boldsymbol{\omega} - T^T\int_D \rho\Phi^T\tilde{\omega}^2(\mathbf{r} + \Phi T\boldsymbol{\xi}_{\text{ind}})dD - 2T^T\int_D \rho\Phi^T\widetilde{\Phi T\boldsymbol{\eta}}^T_{\text{ind}}dD\boldsymbol{\omega} - C^*\boldsymbol{\eta}_{\text{ind}} - K^*\boldsymbol{\xi}_{\text{ind}} + \mathbf{X}^* \tag{37}$$

in which

$$J = \int_D \rho(\tilde{r} + \widetilde{\Phi T\boldsymbol{\xi}_{\text{ind}}})(\tilde{r} + \widetilde{\Phi T\boldsymbol{\xi}_{\text{ind}}})^T dD$$

$$\bar{\bar{\Phi}}^* = T^T\bar{\bar{\Phi}}T, \quad C^* = T^TCT, \quad K^* = T^TKT, \quad \mathbf{X}^* = T^T\mathbf{X} \tag{38}$$

At this point, we must discuss two important issues, one concerning the number of elastic degrees of freedom of the formulation and the other, a more important one, whether mean axes should be used at all. We recall that the dimension of the independent coordinate vector $\boldsymbol{\xi}_{\text{ind}}$ is $n-6$, so that using the mean axes as a reference frame, there are only $n-6$ elastic degrees of freedom. This is in contrast with the widespread notion that there are fully n elastic degrees of freedom. A paradox arises when mean axes are used and *the flexibility is modeled by a number of shape functions smaller than or equal to six*, as in these cases *the number of elastic degrees of freedom is either negative or zero*, which is a physical impossibility. Moreover, either explicitly or implicitly, the aerodynamic forces depend on the velocities $\mathbf{V}$, $\boldsymbol{\omega}$, and $\boldsymbol{\eta}$, which, according to the formulation leading to Eqs. (31), are in terms of components along body axes embedded in the undeformed body. Consistent with this, $\mathbf{F}$, $\mathbf{M}$, and $\mathbf{X}$ are in terms of the same body axes. Clearly, if the velocities $\mathbf{V}$, $\boldsymbol{\omega}$, and $\boldsymbol{\eta}$ are referred to the mean axes as stated in Nydick and Friedmann [6] and Friedmann et al. [7], then the aerodynamic forces in Eqs. (4) must also be expressed in terms of components along the same mean axes. Yet, there is no indication in Refs. [6,7] that the aerodynamic forces were ever transformed from the original body axes to the mean axes, which raises additional doubts about the validity of the formulation. Hence, if one insists on using the mean axes as a reference frame, then one must fully understand the implications and honor the constraints consistently throughout. Merely invoking the use of mean axes to eliminate terms without enforcing the constraints to both motions and forces makes for an erroneous formulation and results. In view of this, one must question the use mean axes in the first place. Indeed, from Eqs. (37) we conclude that the use of mean axes only results in added complications with no visible benefits.

In the case of helicopters, the equations of motion, Eqs. (23), are not only significantly more complex than those for aircraft but they also depend explicitly on time. Clearly, the use of mean axes for helicopters is hard to comprehend.

Finally, one must wonder why it was necessary to begin with such a complete and rigorous formulation as Eqs. (11) of Ref. [8] and strip it down beyond recognition (see Eqs. (4) of the present paper).

5 Conclusions

The equations of motion for some flexible aerial vehicles, such as missiles, aircraft, and helicopters tend to be very complex. Decades ago, when computers were inadequate, the emphasis was on analytical solutions, which prompted some investigators to resort to undue simplification of the equations of motion, a practice continuing to this date. In this regard, the concept of mean axes seemed quite appealing, as in some cases it held the promise of inertial decoupling of the equations of motion. This approach turned out to be a blind alley for aerial vehicles, as the equations of motion remained coupled through the aerodynamic forces. Ironically, as shown in this paper, proper use of the mean axes can only result in an increase in the complexity of the formulation.

The undue insistence on simplifying the equations of motion through inertial decoupling seems to be a throwback to an age predating the commonplace use of powerful computers. This is particularly true in view of the availability of efficient computer tools capable of implementing solutions to very complex problems. In this regard, we should mention MATHEMATICA for symbolic manipulation and MATLAB for matrix manipulation. Indeed, using MATHEMATICA in conjunction with a formulation in terms of momenta rather than velocities, but entirely equivalent to the formulation in this paper, Meirovitch and Tuzcu [12–14] not only carried out time simulations of the rigid-body and elastic responses of flying aircraft but also implemented feedback controls ensuring the flight stability of maneuvering flexible aircraft.

References

[1] Bisplinghoff, R. L., and Ashley, H., 1962, *Principles of Aeroelasticity*, Wiley, New York, Chap. 9.

[2] Meirovitch, L., 1970, *Methods of Analytical Dynamics*, McGraw-Hill, New York (reprinted in 2003 by Dover).

[3] Meirovitch, L., and Nelson, H. D., 1966, "On the High-Spin Motion of a Satellite Containing Elastic Parts," J. Spacecr. Rockets, **3**(11), pp. 1597–1602.

[4] Canavin, J. R., and Likins, P. W., 1977, "Floating Reference Frames for Flexible Spacecraft," J. Spacecr. Rockets, **14**(12), pp. 724–732.

[5] Dusto, A. R., Brune, G. W., Dornfield, G. M., Mercer, J. E., Pilet, S. C. Rubbert, P. E., Schwanz, R. C., Smutney, P., Tinoco, E. N., and Weber, J. A., 1974, "A Method for Predicting the Stability Characteristics of an Elastic Airplane," Vol. 1-FLEXSTAB Theoretical Description," Report No. NASA CR-114712.

[6] Nydick, I., and Friedmann, P. P., 1999, "Aeroelastic Analysis of a Generic Hypersonic Vehicle," *Proceedings of the CEAS/AIAA/ICASE/NASA Langley International Forum on Aeroelasticity and Structural Dynamics*, NASA/CP-1999–205/36/PT2, Williamsburg, VA, pp. 777–810.

[7] Friedmann, P. P., McNamara, J. J., Thuruthimattam, B. J., and Nydick, I., 2004, "Aeroelastic Analysis of Hypersonic Vehicles," J. Fluids Struct., **19**, pp. 681–712.

[8] Meirovitch, L., 1991, "Hybrid State Equations of Motion for Flexible Bodies in Terms of Quasi-Coordinates," J. Guid. Control Dyn., **14**(5), pp. 1008–1013.

[9] Meirovitch, L., 1997, "A Unified Theory for the Flight Dynamics and Aeroelasticity of Whole Aircraft," *Proceedings of the Eleventh Symposium on Structural Dynamics and Control*, Blacksburg, VA, pp. 461–468.

[10] Cribbs, R. C. Friedmann, P P., and Chiu, T., 2000, "Coupled Helicopter Rotor/ Flexible Fuselage Aeroelastic Model for Control of Structural Response," AIAA J., **38**(10), pp. 1777–1788.

[11] Meirovitch, L., 1997, *Principles and Techniques of Vibrations*, Prentice-Hall, Englewood Cliffs, NJ.

[12] Meirovitch, L., and Tuzcu, I., 2004, "Unified Theory for the Dynamics and Control of Maneuvering Flexible Aircraft," AIAA J., **42**(4), pp. 714–727.

[13] Meirovitch, L., and Tuzcu, I., 2004, "Time Simulations of the Response of Maneuvering Flexible Aircraft," J. Guid. Control Dyn., **27**(5), pp. 814–828.

[14] Meirovitch, L., and Tuzcu, I., 2005 "Control of Flexible Aircraft Executing Time-Dependent Maneuvers," J. Guid. Control Dyn., **28**(6), pp. 1291–1300.

Combined Torsional-Bending-Axial Dynamics of a Twisted Rotating Cantilever Timoshenko Beam With Contact-Impact Loads at the Free End

Sunil K. Sinha
Principal Engineer
GE Aircraft Engines,
M.D.: P-30,
General Electric Company,
1 Neumann Way,
Cincinnati, OH 45215
e-mail: sunil.sinha@ae.ge.com

In this paper, consideration is given to the dynamic response of a rotating cantilever twisted and inclined airfoil blade subjected to contact loads at the free end. Starting with the basic geometrical relations and energy formulation for a rotating Timoshenko beam constrained at the hub in a centrifugal force field, a system of coupled partial differential equations are derived for the combined axial, lateral and twisting motions which includes the transverse shear, rotary inertia, and Coriolis effects, as well. In the mathematical formulation, the torsion of the thin airfoil also considers a very general case of shear center not being coincident with the CG (center of gravity) of the cross section, which allows the equations to be used also for analyzing eccentric tip-rub loading of the blade. Equations are presented in terms of axial load along the longitudinal direction of the beam which enables us to solve the dynamic pulse buckling due to the tip being loaded in the longitudinal as well as transverse directions of the beam column. The Rayleigh–Ritz method is used to convert the set of four coupled-partial differential equations into equivalent classical mass, stiffness, damping, and gyroscopic matrices. Natural frequencies are computed for beams with varying "slenderness ratio" and "aspect ratio" as well as "twist angles." Dynamical equations account for the full coupling effect of the transverse flexural motion of the beam with the torsional and axial motions due to pretwist in the airfoil. Some transient dynamic responses of a rotating beam repeatedly rubbing against the outer casing is shown for a typical airfoil with and without a pretwist.
[DOI: 10.1115/1.2423035]

1 Introduction

Rotating beams, which have importance due to numerous practical usage such as jet engine blades, helicopter rotor blades, airplane propellers, satellite antennas, cutting-tool dynamics, and other turbomachinery applications, have been investigated for a long time. In order to analyze the dynamic characteristics of turbine and compressor blades, it is a common practice to consider it as a rotating radial cantilever beam. At the same time turbine and compressor blade designers have long felt that this characterization ignores some vital geometrical details of a real blade such as lean and twist in the blade; which limits the applicability of such simplified analytical models especially in the area of aerodynamic flutter and rub-induced dynamic instabilities in the blade. One such aspect with direct applications to turbine and compressor blades is the vibration of pretwisted beams, which is commonly referred as "twist-bend coupling characteristics of airfoils." During a typical rub-induced vibration event, the blade-tip moving with a tangential velocity of about 400–500 m/s makes a sudden glancing contact (impacting at a very shallow incidence angle) on the casing inner surface; which becomes the excitation mechanism for initiating free vibration in the blade. This process is repeated hundreds of times usually with one rub event every revolution. The typical radial interference between the blade tip and the casing inner surface responsible for generating the periodic contact rub load usually does not exceed 0.10–0.15 mm. After the earlier works [1–4] done about 30 years ago on free vibration characteristics of pretwisted beams simulating the airfoil, there is a considerable current interest in applying the methods of nonlinear vibration to the dynamic stability of asymmetric airfoils cross sections, especially rotating blades with aerodynamic excitations [5,6]. Dynamic stability of cantilever beams with varying levels of complexity and different types of loading conditions has been investigated by Chen and Peng [7], Hodges [8], Sinha [9], Chen and Ho [10], etc.

After the importance of transverse shear and rotary inertia in the beam formulation was shown by Timoshenko [11], many different aspects of his beam theory have been studied by several authors over the past 40 years. Leissa and Jacob [12] were the first ones to investigate the free vibration characteristics of cantilevered twisted beams and plates as a three-dimensional vibration problem. Rosen [13] has presented a comprehensive review of structural and dynamic aspects of pretwisted beams. Lin and his coworkers [14] have performed dynamic analysis of nonuniform pretwisted Timoshenko beam with elastic boundary conditions. Petrov and Geraldin [15] have developed the finite-element theory for a curved and twisted beams based upon a geometrically nonlinear formulation. Among the newer contributions, Tang and Yu [16] have presented a generalized variational principle on the nonlinear theory of a pretwisted curved beam. Other recent contributions in this field [17–22] primarily deal with the free vibration characteristics of the twisted rotating beams; which also include the effect of transverse shear and rotary inertia. Different parameters of dynamic stability of twisted rotating beams under external axial loads have been investigated by Chen and Keer [23], Lee [24], Liao and Huang [25], and Sakar and Sabuncu [26]. Yang and Tsao [27] studied the dynamic stability of pretwisted blade due to changing rotational speed. Temel [28,29] was the first one to analyze the transient response of a curved beam in the form of a helix and subjected to time-dependent loads. Turhan and Bulut [30]

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received February 15, 2005; final manuscript received May 31, 2006. Review conducted by Marc P. Mignolet.

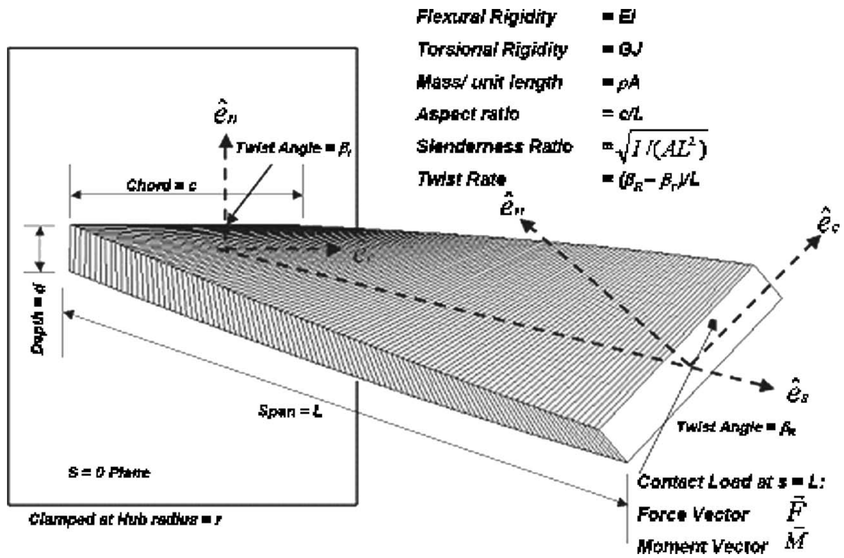


Fig. 1 A pretwisted Timoshenko beam and the local coordinate system

focused on the dynamic stability of a rotating blade due to fluctuations in the speed of the shaft. In the current work, our main focus is on developing the governing dynamical equations to analyze the effect of rub-induced contact-impact forces at the free end of the rotating blade modeled as a cantilever Timoshenko beam with a pretwist (see Fig. 1).

Starting from the basic deformation and velocity equations along with the rotary inertia and gyroscopic effect terms, a complete set of coupled dynamic equations has been derived for this problem. The eigenvalue problem of these equations in a matrix form is solved to determine the fundamental natural frequencies for a combination of varying geometrical parameters to characterize the Timoshenko beam. We have also solved the corresponding transient dynamics problem due to time-dependent contact-impact loading at the free end for twisted and untwisted blades.

2 Rotating Cantilever Timoshenko Beam Formulation for a Pretwisted Blade With Lean

For mathematical derivation, we consider that elastically deformable blades of outer radius "R" with the stagger angle "β_r," which the blade root-chord makes with the engine axis, are mounted on a rigid disk of hub radius "r" (see Fig. 2(*a*) and the Nomenclature for a full list of notations). We introduce two different local coordinate frames of reference attached to the rotating blade called "axial-tangential-radial" with unit vectors as $(\hat{e}_a, \hat{e}_t, \hat{e}_r)$ and, "chord-normal-span" with unit vectors as $(\hat{e}_c, \hat{e}_n, \hat{e}_s)$, respectively (see Fig. 2(*b*)) such that the longitudinal axis of the equivalent Timoshenko beam passes through the center of gravity (CG) of the beam cross section. The blade twist angle β is defined such that the airfoil center of curvature in a fan or compressor blade is towards the direction of rotation or spin-velocity Ω whereas in a turbine blade it is in the opposite direction of Ω. It should be noted that in general due to lean in the blade, the blade longitudinal axis in the span direction may not necessarily coincide with the local radial direction. Thus, the effect of the sweep angle α in a blade with a lean about the local radial direction is as follows:

$\alpha > 0$: forward-swept blade

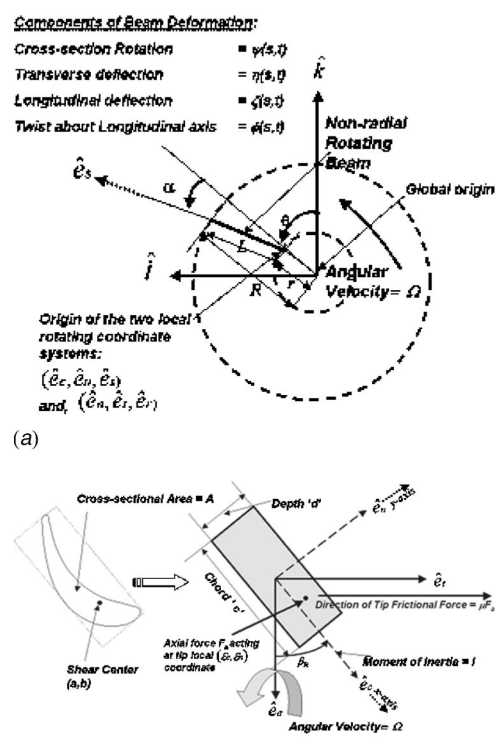
$\alpha = 0$: radial blade

$\alpha < 0$: backward-swept blade

In actual applications, the typical magnitude of the sweep-angle α is relatively small which ranges from −15 to about 15 deg. For such small values of α, one can assume that the normal to any beam cross-section surface makes a constant angle with the local radial direction $(\hat{e}_r)$ passing through the centroid of that cross section. In this analysis we further assume that due to twist in the beam, the "stagger angle $\beta(s)$" changes linearly as we move along the blade longitudinal axis from the root to the tip, such that

$$\beta(s) = \beta_r + s\beta' \tag{1}$$

and the rate of twist



(*a*)

(*b*)

Fig. 2 (*a*) Schematic representation of an inclined rotating beam with respect to the fixed global frame of reference as viewed along the spin axis. (*b*) Airfoil cross section and its equivalent Timoshenko beam representation as viewed from the free end of the blade.

$$\beta' = \frac{\beta_R - \beta_r}{L} = \frac{(\beta_R - \beta_r)\cos\alpha}{(R - r)} \tag{2}$$

The individual blades behave like a cantilever beam column of span length "L" and are subjected to a centrifugal force field F_{cf} generated due to the rotor spin velocity "Ω," which for the derivation purposes can be treated as an external force on the system. Thus, the free end of the cantilever beam at $s=L$ is subjected to a generalized external force vector $\mathcal{F}$ and moment vector $\mathcal{M}$ such that

$$\begin{aligned}\mathcal{F} &= F_C\hat{e}_c + F_N\hat{e}_n + F_S\hat{e}_s \\ \mathcal{M} &= M_C\hat{e}_c + M_N\hat{e}_n + M_S\hat{e}_s\end{aligned} \tag{3}$$

The span direction component "F_S (+ sign: tension and − sign: compression)" of the external blade tip rub load vector $\mathcal{F}$ acting along the longitudinal axis of the beam, henceforth is represented by F_a. The tip rub force F_a is a dynamic contact load, which is nonzero only during the tip travel through the rub zone on the stator and will always have a (−) sign due to the contact load being compressive in nature. The chord and normal (thickness direction) components F_C and F_N are generated when friction at the contact surfaces is also considered in the analysis. In the most general case, the tip-rub force F_a may be acting eccentrically at a point with its coordinate location as $(\varepsilon_c, \varepsilon_n)$ with respect to the CG of the beam cross section at the tip. For thick blades, the value of ε_n can be in the range of $[(-d/2) \geqslant \varepsilon_n \geqslant (d/2)]$ depending upon whether the blade is rubbing at the concave side or the convex side of the airfoil. For thin blades, $\varepsilon_n \approx 0$, and in the extreme case of tip rub at the edge of the beam cross section, we will have, $\varepsilon_c = \pm(c/2)$. In addition, the blade is deformed in bending by applying bending moment about its chord in such a way that a typical cross section of the deformed blade produces the cross-section rotation "ψ," the lateral deflection "η" at the neutral axis, the axial deflection "ζ" and the angle of twist as "ϕ." It is assumed that all four components of deformation are functions of spatial coordinate "s," measured along the beam axis and the temporal parameter time "t." It is also assumed that the minor principal moment of inertia of the blade cross-section "I" coincides with the chord direction so that under pure bending moment the blade lateral deflection $\eta(s,t)$ takes place in the direction normal to the chord with the neutral surface passing through the radial-chord plane.

For the analytical derivation, we will use the usual notations such as, elastic Young's modulus "E," Poisson's ratio "ν," shear modulus "$G=E/2(1+\nu)$," material mass density "ρ," and cross-sectional area "A." With respect to the stationary global cartesian unit vectors $(\hat{i},\hat{j},\hat{k})$, the local unit vectors attached to the beam $(\hat{e}_c,\hat{e}_n,\hat{e}_s)$, rotating at a constant angular velocity Ω such that $\theta=\Omega t$, are related to each other as

$$\begin{Bmatrix}\hat{e}_c \\ \hat{e}_n \\ \hat{e}_s\end{Bmatrix} = \begin{bmatrix} -\sin\beta\cos(\theta+\alpha) & \cos\beta & \sin\beta\sin(\theta+\alpha) \\ -\cos\beta\cos(\theta+\alpha) & -\sin\beta & \cos\beta\sin(\theta+\alpha) \\ \sin(\theta+\alpha) & 0 & \cos(\theta+\alpha)\end{bmatrix}\begin{Bmatrix}\hat{i} \\ \hat{j} \\ \hat{k}\end{Bmatrix} \tag{4}$$

It is assumed that any warping of the airfoil cross section caused by the torque varying along the span of the beam will be negligibly small. A typical airfoil cross section is not symmetrical about any of the principal axes and as such, in general, its "shear center" may not necessarily coincide with the centroid (CG) of the beam cross section. As a result, under the combined twisting and bending deformation of the airfoil, it is assumed that the shear center in the local coordinate system (chord-normal-span) is located at (a,b,s) in such a way that its position vector is described as

$$[a]\hat{e}_c + [b]\hat{e}_n + [s]\hat{e}_s \tag{5}$$

Then, with (a,b) as the shear center of the cross section and J_0 as the centroidal polar moment of inertia of the cross section, the effective polar moment of inertia for the twist motion can be written as

$$J_0 + A(a^2 + b^2) \tag{6}$$

Here, we neglect the warping of the beam cross section and the slope $\eta_{,s}(s,t)$ of the transverse deformation of the beam is the sum of the rotation $\psi(s,t)$ and the rotation of the cross section due to shear force $Q(s,t)$ expressed as $-(Q/\kappa AG)$, i.e.

$$\eta_{,s}(s,t) = \psi(s,t) - \frac{Q(s,t)}{\kappa AG} \tag{7}$$

The angular velocity vector of the beam column due to spin velocity Ω is

$$[\Omega\cos\beta]\hat{e}_c + [-\Omega\sin\beta]\hat{e}_n + [0]\hat{e}_s \tag{8}$$

Under the small rotation assumptions, the rotation of the beam cross section after the deformation can be expressed as a rotation vector $\mathfrak{R}$ such that

$$\mathfrak{R} = [\psi]\hat{e}_c + [\phi]\hat{e}_s \tag{9}$$

Due to pretwist in the beam about the span axis (s-direction) with the twist rate of β', the derivative of the cross-section rotation vector term $\mathfrak{R}$ is derived using the chain rule as

$$\frac{d\mathfrak{R}}{ds} = \left[\frac{\partial\psi}{\partial s}\hat{e}_c + \psi\frac{\partial\hat{e}_c}{\partial s}\right] + \frac{\partial\phi}{\partial s}\hat{e}_s$$

i.e.

$$\frac{d\mathfrak{R}}{ds} = [\psi_{,s}]\hat{e}_c + [\beta'\psi]\hat{e}_n + [\phi_{,s}]\hat{e}_s \tag{10}$$

In the above equation, $\psi_{,s}$ and $\beta'\psi$ represent the changes in the curvature of the beam about the chord and normal axes of the cross section due to bending about the two principal directions. In the derivative of the rotation vector $\mathfrak{R}$ the contribution of the term containing unit vector component $\hat{e}_n$ is only due to pretwist in the beam. For example, in an untwisted beam with $\beta'=0$, a bending moment about $\hat{e}_c$ will not produce any curvature change about $\hat{e}_n$. The deformation vector of any point on the Timoshenko beam located at (x,y,s) in the local chord-normal-span coordinate system caused by the four components of deformation (ψ,η,ζ,ϕ) is obtained as

$$[-(y-b)\phi]\hat{e}_c + [\eta + (x-a)\phi]\hat{e}_n + [\zeta - y\psi]\hat{e}_s \tag{11}$$

The position vector in the local chord-normal-span coordinate system after deformation of any typical point at (x,y,s) can be written as

$$[x - y\phi + b\phi]\hat{e}_c + [y + (\eta - a\phi) + x\phi]\hat{e}_n + [s + \zeta - y\psi]\hat{e}_s \tag{12}$$

The time-dependent position vector of the rotating beam clamping point at the hub radius$=r$ in the global fixed frame of reference is $(r\sin\theta\hat{i} + r\cos\theta\hat{k})$. After some lengthy algebraic manipulation, the position vector of the clamping point in the local $(\hat{e}_c,\hat{e}_n,\hat{e}_s)$ system is expressed as

$$[r\sin\beta\sin\alpha]\hat{e}_c + [r\cos\beta\sin\alpha]\hat{e}_n + [r\cos\alpha]\hat{e}_s \tag{13}$$

Thus, the corresponding global position vector in the chord-normal-span coordinate system after deformation of a typical point at (x,y,s) on the beam can be written as

$$\begin{aligned}&[r\sin\beta\sin\alpha + x - y\phi + b\phi]\hat{e}_c + [r\cos\beta\sin\alpha + y + (\eta - a\phi) \\ &\quad + x\phi]\hat{e}_n + [r\cos\alpha + s + \zeta - y\psi]\hat{e}_s\end{aligned} \tag{14}$$

The time derivatives of the unit vectors $(\hat{e}_c,\hat{e}_n,\hat{e}_s)$ are obtained as

$$\begin{aligned}\dot{\hat{e}}_c &= \Omega \sin \beta(s)\hat{e}_s \\ \dot{\hat{e}}_n &= \Omega \cos \beta(s)\hat{e}_s \\ \dot{\hat{e}}_s &= -\Omega \sin \beta(s)\hat{e}_c - \Omega \cos \beta(s)\hat{e}_n\end{aligned} \tag{15}$$

Hence, the corresponding velocity vector $\boldsymbol{V}$ in a fixed global frame of reference, for any typical point on a rotating beam located at (x,y,s) with respect to the local $(\hat{e}_c,\hat{e}_n,\hat{e}_s)$ system, is derived as

$$\begin{aligned}\boldsymbol{V} = &\left\{ \begin{bmatrix} \zeta_{,t} - y\psi_{,t} + (r\cos\beta\sin\alpha + y + \eta - a\phi + x\phi)\Omega\cos\beta \\ + (r\sin\beta\sin\alpha + x - y\phi + b\phi)\Omega\sin\beta \end{bmatrix}\sin\theta + \begin{bmatrix} (s + r\cos\alpha + \zeta - y\psi)\Omega - \eta_{,t}\cos\beta \\ -(x\cos\beta - a\cos\beta - y\sin\beta + b\sin\beta)\phi_{,t} \end{bmatrix}\cos\theta \right\}.\hat{i} \\ &- [(x\sin\beta - a\sin\beta + y\cos\beta - b\cos\beta)\phi_{,t} + \eta_{,t}\sin\beta].\hat{j} \\ &+ \left\{ \begin{bmatrix} \zeta_{,t} - y\psi_{,t} + (r\cos\beta\sin\alpha + y + \eta - a\phi + x\phi)\Omega\cos\beta \\ + (r\sin\beta\sin\alpha + x - y\phi + b\phi)\Omega\sin\beta \end{bmatrix}\cos\theta - \begin{bmatrix} (s + r\cos\alpha + \zeta - y\psi)\Omega - \eta_{,t}\cos\beta \\ -(x\cos\beta - a\cos\beta - y\sin\beta + b\sin\beta)\phi_{,t} \end{bmatrix}\sin\theta \right\}.\hat{k}\end{aligned} \tag{16}$$

In a typical twisted beam formulation, it is a common practice to characterize the bending mode deformation about the two principal axes of the beam cross section (see the Appendix for details) as a coupled deflection in $(x-x)$ and $(y-y)$ directions, separately. In the present approach, we have introduced twist ϕ of the beam cross section about the longitudinal axis as an independent degree-of-freedom. Recalling that in a typical beam formulation inherent assumption of $I_{yy} \gg I_{xx}$ along with (a,b) being the shear center of a nonsymmetric cross section and J_0 its polar moment of inertia, we have $I_{xx}+I_{yy}=J_0+A(a^2+b^2)$ such that for shear center coincident with CG

$$I_{xx} = \iint_{\text{Area}} y^2 dA = I, \quad I_{yy} = \iint_{\text{Area}} x^2 dA \approx \infty \quad \text{and} \qquad J_0 = \iint_{\text{Area}} (x^2+y^2)dA \tag{17}$$

The above relationship assumes that in an equivalent symmetrical cross section under combined twist and bending, the shear center would coincide with its centroid such that, $a=0$ and $b=0$. It is obvious that when the shear center is not coincident with the CG of the cross section, the contact forces will always generate a moment at the free end. We will recall that with the shear modulus $G=E/2(1+\nu)$, the torsional rigidity for a thin cross section can be written as "GJ" in which according to Timoshenko and Goodier [31] for rectangular cross sections, the torsion constant $J=(1/3)cd^3$. Since, the flexural bending in a pretwisted beam would inherently result in a twist-bend-coupling caused by the components of the beam deformation in the two principal directions of the cross section, it is convenient to represent the flexural rigidity term "EI_{yy}" as a function of its torsional rigidity term GJ. In addition, in such beams with thin cross sections the effect of Poisson's ratio ν is negligible and as such the flexural rigidity about the major principal axis EI_{yy} in terms of its torsional rigidity GJ can be approximated as

$$EI_{yy} \approx (2GJ - EI) \tag{18}$$

Thus, combining the kinetic energy "T" and the potential energy "U" due to bending, transverse shear, twisting and centrifugal loads "F_{cf}" as well as the axial force due to contact "F_a" yields the simplified form of Lagrangian "Λ" for the rotating Timoshenko's beam column as

$$\Lambda = T - U = \frac{1}{2}\int_0^L \left[\begin{aligned} &\rho A \left(\begin{aligned} &\begin{bmatrix} \zeta_{,t} - y\psi_{,t} + (r\cos\beta_r\sin\alpha + y + \eta - a\phi + x\phi)\Omega\cos\beta \\ + (r\sin\beta_r\sin\alpha + x - y\phi + b\phi)\Omega\sin\beta \end{bmatrix}^2 \\ &+ \left\{ (s + r\cos\alpha + \zeta - y\psi)\Omega - \eta_{,t}\cos\beta - \begin{bmatrix} (x-a)\cos\beta \\ -(y-b)\sin\beta \end{bmatrix}\phi_{,t} \right\}^2 \\ &\{[(x-a)\sin\beta + (y-b)\cos\beta]\phi_{,t} + \eta_{,t}\sin\beta\}^2 \end{aligned} \right) \\ &- \begin{bmatrix} EI(\psi_{,s})^2 + (2GJ - EI)(\beta'\psi)^2 + GJ(\phi_{,s})^2 + AE(\zeta_{,s})^2 \\ + \kappa AG(\eta_{,s} - \psi)^2 + \kappa AG(\beta'\eta)^2 + 2\kappa AG\beta'(y-b)\phi(\eta_{,s} - \psi) \\ + 2(2GJ - EI)\beta'\phi_{,s}[\psi + (y-b)\psi_{,s}] \end{bmatrix} \\ &- F_a\{[(\eta + (x-a)\phi)_{,s}]^2 + [\beta'\eta]^2 + [(y-b)\phi_{,s}]^2\} \\ &- F_{cf}\cos\alpha\{[(\eta + (x-a)\phi)_{,s}]^2 + [(y-b)\phi_{,s}]^2\} + 2\mu F_a\cos\beta_R\psi\zeta_{,s} \\ &+ 2F_a[(a - \varepsilon_c)\cos\beta'(s-L) + (b - \varepsilon_n)\sin\beta'(s-L)]\psi\phi_{,s} \\ &- 2F_a[(b - \varepsilon_n)\cos\beta'(s-L) - (a - \varepsilon_c)\sin\beta'(s-L)]\beta'(\eta - b)\phi_{,s} \end{aligned} \right] ds \tag{19}$$

The Lagrangian equation is

$$\frac{d}{dt}\left(\frac{\partial \Lambda}{\partial x_{i,t}}\right) - \frac{\partial \Lambda}{\partial x_i} + \frac{d}{ds}\left(\frac{\partial \Lambda}{\partial x_{i,s}}\right) = R_i \tag{20}$$

Using the Lagrangian Λ in the Lagrange's equation yields the following 4 coupled partial differential equations of motion for ψ, η, ζ, and ϕ, respectively, in a local frame of reference attached to the rotating beam-column with angular velocity Ω as

(a) Bending moment balance due to rotation of the beam cross section about its minor principal axis (chord)

$$\begin{aligned} &-EI\psi_{,ss} - \kappa AG(\eta_{,s} - \psi) + (2GJ - EI)\beta'(\beta'\psi + \phi_{,s} + b\phi_{,ss}) + \kappa AG\beta' b\phi \\ &- F_a[(a-\varepsilon_c)\cos\beta'(s-L) + (b-\varepsilon_n)\sin\beta'(s-L)]\phi_{,s} - \mu F_a \cos\beta_R \zeta_{,s} \\ &- \rho I\Omega^2\psi + \rho I\dot{\Omega}\sin\beta\phi + \rho I\psi_{,tt} + 2\rho I\Omega\sin\beta\phi_{,t} = T_c(s,t) \end{aligned} \tag{21a}$$

(b) Shear force balance in the normal direction through the cross section of the beam

$$\begin{aligned} &-\kappa AG[\eta_{,ss} - \psi_{,s}] + \kappa AG\beta' b\phi_{,s} + \kappa AG(\beta')^2\eta - F_a[\eta_{,ss} - (\beta')^2\eta - a\phi_{,ss}] \\ &+ F_a[(b-\varepsilon_n)\cos\beta'(s-L) - (a-\varepsilon_c)\sin\beta'(s-L)]\beta'\phi_{,s} - F_{\mathrm{cf}}\cos\alpha[\eta_{,ss} - a\phi_{,ss}] \\ &- \rho A\dot{\Omega}\cos\beta\zeta - \rho A\Omega^2\cos\beta[\cos\beta(\eta - a\phi) + b\sin\beta\phi] \\ &- 2\rho A\Omega\cos\beta\zeta_{,t} + \rho A\eta_{,tt} - \rho Aa\phi_{,tt} = Q_n(s,t) \end{aligned} \tag{21b}$$

(c) Axial force balance due to membrane stretching along the neutral surface of the beam

$$\begin{aligned} &-EA\zeta_{,ss} - \rho A\Omega^2\zeta + \rho A\dot{\Omega}[\cos\beta(\eta - a\phi) + b\sin\beta\phi] + \mu F_a\cos\beta_R\psi_{,s} \\ &+ \rho A[\zeta_{,tt} + 2\Omega\cos\beta\eta_{,t} - 2\Omega(a\cos\beta - b\sin\beta)\phi_{,t}] = Q_s(s,t) \end{aligned} \tag{21c}$$

(d) Torque balance due to twist in the beam (neglecting the warping of the cross section)

$$\begin{aligned} &-GJ\phi_{,ss} + \rho I\dot{\Omega}\sin\beta\psi + \rho I\Omega^2\sin^2\beta\phi - \kappa AG\beta' b(\eta_{,s} - \psi) \\ &-(2GJ - EI)\beta'(\psi_{,s} - b\psi_{,ss}) - [F_a + F_{\mathrm{cf}}\cos\alpha][(J_0/A)\phi_{,ss} - a\eta_{,ss}] \\ &+ F_a[(a-\varepsilon_c)\cos\beta'(s-L) + (b-\varepsilon_n)\sin\beta'(s-L)][(\beta')^2\eta + \psi_{,s}] \\ &- F_a[(b-\varepsilon_n)\cos\beta'(s-L) - (a-\varepsilon_c)\sin\beta'(s-L)]\beta'(\eta_{,s} - \psi) \\ &+ \rho J_0\phi_{,tt} - \rho Aa\eta_{,tt} + \rho A(a^2+b^2)\phi_{,tt} - 2\rho I\Omega\sin\beta\psi_{,t} + 2\rho A\Omega(a\cos\beta - b\sin\beta)\zeta_{,t} \\ &= T_s(s,t) \end{aligned} \tag{21d}$$

In the above equations, $Q_n(s,t)$ and $Q_s(s,t)$ account for the distributed lateral loads on the beam column in the thickness (normal) and in the longitudinal (span) directions, respectively. Similarly, $T_n(s,t)$ and $T_s(s,t)$ account for the distributed bending and twist moments on the beam column about the neutral (chord) axis and longitudinal (span) directions, respectively. These distributed external force and moments are caused due to centrifugal loads, nonconstant spin-velocity and gas loads due to fluid flow over the surface of the blade. For example, due to radial lean in the blade by an angle α the centrifugal load acting at the CG of the beam cross section generates distributed transverse forces and twist moments given by

$$Q_n(s,t) = \rho A\Omega^2(s + r\cos\alpha)\sin\alpha\cos(\beta_r + \beta' s) \tag{22}$$

$$T_s(s,t) = \rho A\Omega^2(s + r\cos\alpha)\sin\alpha[b\cos(\beta_r + \beta' s) - a\sin(\beta_r + \beta' s)] \tag{23}$$

Obviously, in a free-vibration problem, all nonhomogeneous terms on the right-hand side of the above set of equations are set to zero. The presence of the axial-force term F_a acting on the free end of the beam in these equations contributes to the lateral as well as torsional buckling of the Timoshenko beam. In the above equations, it should be noted that in a rotating beam; its lateral motion $\eta(s,t)$ is coupled with the longitudinal motion $\zeta(s,t)$ and its cross section rotation $\psi(s,t)$ is coupled with the twist in the beam $\phi(s,t)$. These couplings are due to Coriolis effects in the dynamical system, which introduce velocity-dependent skew-symmetric terms in the equations of motion. In a dynamics problem, F_a would become a function of time t and can be expressed as $F_a=F(t)$. If $F(t)$ is an oscillating force with a pulse frequency of f_p-Hz, it develops a parametric excitation in the system which in a sinusoidal form is written as

$$F_a = F(t) = F_{\max}\cos(2\pi f_p t) \tag{24}$$

The set of four partial differential equations outlined in Eqs. (21a)–(21d) describe the fully coupled dynamical characteristics of a twisted rotating cantilever Timoshenko beam with axial loading at the free end, which also includes the effects of nonconstant rotational speed as well as the Coriolis forces. This is the first attempt in any published literature to formulate the complex set of equations in its entirety. The simpler forms of these equations used by other researchers can easily be derived by setting certain parameters equal to zero, such as by making $\dot{\Omega}=0$, these equations represent the dynamics of a beam rotating at a constant speed. Previous derivations for the cantilever airfoil vibration in coupled torsional-bending mode reported in the literature [1–4] are simplified using Euler–Bernoulli beam formulation with $\dot{\Omega}=0$, $a=0$, $b=0$, and $F_a=0$. Furthermore, by setting the sweep or the radial lean angle $\alpha=0$, one obtains the equations for a radial rotating beam. Similarly, by setting the axial loading term $F_a=0$, the corresponding equations for the free vibrations of a beam are obtained. The contributions of axial motion can be disregarded by dropping the terms containing ζ due to additional degree-of-freedom in the longitudinal direction of the beam. In addition, by setting the twist parameter $\beta'=0$, one can simplify the equations similar to one used by Lin [19] for a beam with a constant stagger angle. Similarly, by setting the hub radius term $r=0$, one can derive the equations similar to one used by Oguamanam and Heppler [18]. These equations can be further simplified to represent the coupled flexural and torsional vibrations of classical Euler–Bernoulli's beam as shown by Timoshenko et al. [11]. It has been verified that the coupled flexural and torsional elastic buckling equation due to axial forces and end moments also reduce to the form similar to one derived by Timoshenko and Gere [32]. On replacing the torsional rigidity term GJ by the relations expressed in Eq. (18) one obtains a set of equations similar to that used by Banerjee [21].

3 Boundary Conditions and External Forces

The geometric boundary conditions for the Timoshenko's beam under consideration with four components of deformation $(\psi, \eta, \zeta, \phi)$ are as follows:

$$\psi(0,t)=0, \quad \eta(0,t)=0, \quad \zeta(0,t)=0 \quad \phi(0,t)=0 \tag{25}$$

The corresponding four natural boundary conditions at the free end of the cantilever beam for $(s=L)$ are expressed in terms of the contact force vector $\mathcal{F}$ and moment vector $\mathcal{M}$ components (see Eq. (3)) as

$$F_C=0, \quad M_N=0 \tag{26}$$

$$(\text{bending moment})_{\text{at } s=L}=M_C=EI\psi_{,s}|_{s=L}=F_a[\varepsilon_n+\varepsilon_c\sin(\beta' L)] \tag{27}$$

$$(\text{shear force})_{\text{at } s=L}=F_N=-\kappa AG[\eta_{,s}-\psi]_{s=L}=-\mu F_a\cos\beta_R \tag{28}$$

$$(\text{axial force})_{\text{at } s=L}=F_S=EA\zeta_{,s}|_{s=L}=F_a \tag{29}$$

$$(\text{torque})_{\text{at } s=L}=M_S=GJ\phi_{,s}|_{s=L} = \mu F_a[(\varepsilon_c-a)\cos\beta_R-(\varepsilon_n-b)\sin\beta_R] \tag{30}$$

In a contact-dynamics problem, the tip load F_a along the longitudinal axis and its point of application on the beam cross section $(\varepsilon_c, \varepsilon_n)$ will be time dependent. If the outer case radial tip-clearance γ and its radial stiffness K_{case} need to be included in the analysis then

$$F_a=0 \quad \text{for } \{\cos\alpha\zeta+\cos\beta_R\sin\alpha[\eta+(\varepsilon_c-a)\phi]\}|_{s=L}<\gamma \tag{31}$$

$$\text{and} \quad F_a=-K_{\text{case}}\{\cos\alpha\zeta+\cos\beta_R\sin\alpha[\eta+(\varepsilon_c-a)\phi]-\gamma\}|_{s=L}$$
$$\text{for} \quad \{\cos\alpha\zeta+\cos\beta_R\sin\alpha[\eta+(\varepsilon_c-a)\phi]\}|_{s=L}\geqslant\gamma \tag{32}$$

In order to account for the torsional deformation in the blade in Eq. (31), the positive sign on the chord is used, if the tip leading edge of the blade is rubbing and a negative sign is used if the tip trailing edge is in contact. It should be obvious that for the contact load at the mid-point of the tip cross section, the contribution from the twist parameter ϕ would be zero. The external tip forces $\mathcal{F}$ and moments $\mathcal{M}$ at the free-end of the beam act like point loads at $s=L$, and as such mathematically with the use of Dirac's Delta function $\delta(s-L)$ can be treated like a continuous or distributed external force, which are written as

$$T_c(s,t)=F_a[\varepsilon_n+\varepsilon_c\sin(\beta' L)]\delta(s-L) \tag{33a}$$

$$Q_n(s,t)=-\mu F_a\cos\beta_R\delta(s-L) \tag{33b}$$

$$Q_s(s,t)=F_a\delta(s-L) \tag{33c}$$

$$T_s(s,t)=\mu F_a[(\varepsilon_c-a)\cos\beta_R-(\varepsilon_n-b)\sin\beta_R]\delta(s-L) \tag{33d}$$

4 Rayleigh–Ritz Method and Ordinary Differential Equations of Motion in Matrix Form

One can use several different methods such Galerkin's or other weighted-residual techniques to convert the set of partial differential Eqs. (21a)–(21d) in a set of ordinary differential equations. Each technique imposes certain necessary boundary condition requirements on the approximating functions. Here, we have employed the classical Rayleigh–Ritz method for this purpose, which requires that as a necessary condition, the approximating function must satisfy the geometric constraints arbitrarily but the force-dependent natural boundary conditions may be relaxed. In the ideal situation, they may satisfy all the geometric as well as force boundary conditions of the present problem, however, it is not necessary in general. If the approximating functions do not satisfy all the force boundary conditions as well, then the integrated sum of the unbalanced weighted-residual force and moment terms must be set to zero at the free end. Thus, under these conditions, the solution of the above set of equations can be assumed as

$$\phi(s,t)=\sum_{j=0}^{\infty}U_j(s)W_j(t)=\sum_{j=1}^{\infty}[\sin\varphi_j s]W_j(t) \tag{34}$$

$$\psi(s,t)=\sum_{j=0}^{\infty}U_j(s)X_j(t)=\sum_{j=1}^{\infty}[\sin\varphi_j s]X_j(t) \tag{35}$$

$$\eta(s,t)=\sum_{j=0}^{\infty}V_j(s)Y_j(t)=\sum_{j=1}^{\infty}\frac{[1-\cos\varphi_j s]}{\varphi_j}Y_j(t) \tag{36}$$

$$\zeta(s,t)=\sum_{j=0}^{\infty}S_j(s)Z_j(t)=\sum_{j=1}^{\infty}\left[\frac{\sin\varphi_j s}{\varphi_j}\right]Z_j(t) \tag{37}$$

where

$$\varphi_j=\frac{(2j-1)\pi}{2L}$$

Hence, in order to apply the Rayleigh–Ritz's method, we substitute the assumed deflection shape functions in such a way that the shape function terms have proper dimensions of either length or slope (radians) as necessary. In the above sets of equations, the spatial derivative terms can be written as a function of a set of differential operators. In addition, the discretized form of Eqs. (21a)–(21d) is complete only when all the terms in the infinite series for the displacement functions $S_j(s)$, $U_j(s)$, and $V_j(s)$ are considered, however, in a numerical technique they must be truncated after a certain number of terms in the sequence. On applying the Rayleigh–Ritz's method with the assumed displacement functions, one obtains a set of ordinary differential equations in terms of time-dependent variables. The complete equations with the homogeneous as well as external force terms of these equations in a matrix form can be written as

$$\boldsymbol{M}\{\ddot{f}(t)\}+\boldsymbol{C}\{\dot{f}(t)\}+\boldsymbol{K}\{f(t)\}=\{P(t)\} \tag{38}$$

Here $\boldsymbol{M}$ is the coefficient matrix for the acceleration-dependent force terms generally known as inertia or the mass matrix, $\boldsymbol{C}$ is the coefficient matrix for the velocity-dependent force terms which can be due to damping or due to gyroscopic effects in the dynamical system, and $\boldsymbol{K}$ is the coefficient matrix for the displacement-dependent force terms generally known as the stiffness matrix. It can be seen that in Eq. (38) the terms containing generalized coordinates in the column-vector $\{f(t)\}$ are X, Y, Z, and $W's$, which are dimensionless. Suppose, we consider "N" number of terms for each of the four basic deformation trial functions outlined in Eqs. (33)–(36), then for brevity we can introduce following notations in lieu of the generalized coordinates $\{f(t)\}$

$$\begin{bmatrix} \boldsymbol{M} \end{bmatrix}\begin{Bmatrix} \{\ddot{X}\}_N \\ \{\ddot{Y}\}_N \\ \{\ddot{Z}\}_N \\ \{\ddot{W}\}_N \end{Bmatrix} + \begin{bmatrix} \boldsymbol{C} \end{bmatrix}\begin{Bmatrix} \{\dot{X}\}_N \\ \{\dot{Y}\}_N \\ \{\dot{Z}\}_N \\ \{\dot{W}\}_N \end{Bmatrix} + \begin{bmatrix} \boldsymbol{K} \end{bmatrix}\begin{Bmatrix} \{X\}_N \\ \{Y\}_N \\ \{Z\}_N \\ \{W\}_N \end{Bmatrix} = \begin{Bmatrix} P(t) \end{Bmatrix}_{4N} \tag{39}$$

It is obvious that with the N number of terms used to represent each of the four displacement functions viz. $(\psi, \eta, \zeta, \phi)$, the total number of degrees of freedom in the numerical scheme will be "$4N$." In the above equation, the right-hand side column-vector $\{P(t)\}$ representing all the external forces due to contact load F_a acting along the beam-axis eccentrically at $(\varepsilon_c, \varepsilon_n)$ in the chord-normal-plane at the free end of the cantilever beam column can be expressed as

$$\{P(t)\} = \begin{pmatrix} \{P^{\psi}(t)\}_N \\ \{P^{\eta}(t)\}_N \\ \{P^{\zeta}(t)\}_N \\ \{P^{\phi}(t)\}_N \end{pmatrix} = \begin{pmatrix} \{F_a[\varepsilon_n + \varepsilon_c \sin(\beta' L)]U(L)\} \\ \{-\mu F_a \cos \beta_R V(L)\} \\ \left[\rho A \Omega^2 \int_0^L (s + r\cos\alpha) S ds + F_a S(L)\right]_N \\ \{\mu F_a[(\varepsilon_c - a)\cos \beta_R - (\varepsilon_n - b)\sin \beta_R]U(L)\} \end{pmatrix} \tag{40}$$

Furthermore, the matrix terms used in Eq. (38) can be broken into following separate matrices

$$\boldsymbol{M}\{\ddot{f}(t)\} + [\boldsymbol{C}_D + \boldsymbol{C}_G]\{\dot{f}(t)\} + [\boldsymbol{K}_S + \boldsymbol{K}_F + \boldsymbol{K}_\Omega]\{f(t)\} = \{P(t)\} \tag{41}$$

where $\boldsymbol{M}$ is the mass matrix (symmetric), a function of density ρ and I, $\boldsymbol{K}_S$ is the elastic stiffness matrix (symmetric), a function of E, G, I, etc., $\boldsymbol{K}_\Omega$ is the centrifugal stress-related stiffness-matrix as a result of spin-velocity Ω (symmetric), $\boldsymbol{K}_F$ is the in-plane force-dependent circulatory matrix due to contact force F_a acting along the longitudinal axis at $(\varepsilon_c, \varepsilon_n, L)$ of the beam-column (nonsymmetric), $\boldsymbol{C}_D$ is the damping matrix due to the material internal damping $=(2\chi/\Omega)\boldsymbol{K}_s$, $\boldsymbol{C}_G(\Omega)$ is the gyroscopic matrix (skew symmetric), causes coupling of axial and lateral motions in the beam, and $\{P(t)\}$ is the column vector containing external forces on the dynamical system.

Thus, the individual nonzero terms in the equivalent mass $\boldsymbol{M}$, damping $\boldsymbol{C}$ and stiffness $\boldsymbol{K}$ matrices are as follows:

$$[Ki,j] = -EI\int_0^L U_i U_j'' ds + \kappa AG \int_0^L U_i U_j ds + (2GJ - EI)(\beta')^2 \int_0^L U_i U_j ds - \rho I \Omega^2 \int_0^L U_i U_j ds \tag{42}$$

$$[M_{i,j}] = \rho I \int_0^L U_i U_j ds \tag{43}$$

$$[K_{i,j+N}] = -\kappa AG \int_0^L U_i V_j' ds \tag{44}$$

$$[K_{i,j+2N}] = -\mu F_a \cos \beta_R \left[\int_0^L U_i S_j' ds\right] \tag{45}$$

$$\begin{aligned}[K_{i,j+3N}] = {} & (2GJ - EI)\beta'\left[\int_0^L U_i U_j' ds - U_i(L)U_j(L)\right] \\ & + \kappa AG\beta' b \int_0^L U_i U_j ds + (2GJ - EI)\beta' b\left[\int_0^L U_i U_j'' ds\right] \\ & + \rho I \dot{\Omega}\int_0^L \sin \beta U_i U_j ds - F_a\left[\int_0^L [(a - \varepsilon_c)\cos \beta'(s - L) \right. \\ & \left. + (b - \varepsilon_n)\sin \beta'(s - L)]U_i U_j' ds - (a - \varepsilon_c)U_i(L)U_j(L)\right] \end{aligned} \tag{46}$$

$$[C_{i,j+3N}] = 2\rho I \Omega \int_0^L \sin \beta U_i U_j ds \tag{47}$$

$$[K_{i+N,j}] = \kappa AG\left[\int_0^L V_i U_j' ds - V_i(L)U_j(L)\right] \tag{48}$$

$$\begin{aligned}[K_{i+N,j+N}] = {} & -\kappa AG\left[\int_0^L V_i V_j'' ds - V_i(L)V_j'(L)\right] \\ & - \rho A \Omega^2 \int_0^L \cos^2 \beta V_i V_j ds \\ & - \frac{\rho A \Omega^2}{2}\cos\alpha \int_0^L (R^2 - s^2\cos^2\alpha - r^2 \\ & - 2sr\cos\alpha)V_i V_j'' ds + (\beta')^2 \kappa AG \int_0^L V_i V_j ds \\ & - F_a\left[\int_0^L V_i V_j'' ds - V_i(L)V_j'(L) - (\beta')^2 \int_0^L V_i V_j ds\right] \end{aligned} \tag{49}$$

$$[M_{i+N,j+N}] = \rho A \int_0^L V_i V_j ds \tag{50}$$

$$[K_{i+N,j+2N}] = -\rho A \dot{\Omega}\int_0^L \cos \beta V_i S_j ds \tag{51}$$

$$[C_{i+N,j+2N}] = -2\rho A \Omega \int_0^L \cos \beta V_i S_j ds \tag{52}$$

$$[K_{i+N,j+3N}] = (\kappa AG\beta' b)\left[\int_0^L V_i U_j' ds - V_i(L)U_j(L)\right] - \rho A\Omega^2 \int_0^L \cos\beta(b\sin\beta - a\cos\beta)V_i U_j ds + a\cos\alpha\frac{\rho A\Omega^2}{2}\int_0^L (R^2 - s^2\cos^2\alpha - r^2 - 2sr\cos\alpha)V_i U_j'' ds - F_a\left\{\begin{array}{l} -a\left[\int_0^L V_i U_j'' ds - V_i(L)U_j'(L)\right] \\ -\beta'\int_0^L [(b-\varepsilon_n)\cos\beta'(s-L) - (a-\varepsilon_c)\sin\beta'(s-L)]V_i U_j' ds + \beta'(b-\varepsilon_n)V_i(L)U_j(L)\end{array}\right\} \quad (53)$$

$$[M_{i+N,j+3N}] = -\rho Aa\int_0^L V_i U_j ds \quad (54)$$

$$[K_{i+2N,j}] = \mu F_a \cos\beta_R\left[\int_0^L S_i U_j' ds\right] \quad (55)$$

$$[K_{i+2N,j+N}] = \rho A\dot{\Omega}\int_0^L \cos\beta S_i V_j ds \quad (56)$$

$$[C_{i+2N,j+N}] = 2\rho A\Omega\int_0^L \cos\beta S_i V_j ds \quad (57)$$

$$[K_{i+2N,j+2N}] = -EA\left[\int_0^L S_i S_j'' ds - S_i(L)S_j'(L)\right] - \rho A\Omega^2\int_0^L S_i S_j ds \quad (58)$$

$$[M_{i+2N,j+2N}] = \rho A\int_0^L S_i S_j ds \quad (59)$$

$$[K_{i+2N,j+3N}] = -\rho A\dot{\Omega}\int_0^L (a\cos\beta - b\sin\beta)S_i U_j ds \quad (60)$$

$$[C_{i+2N,j+3N}] = -2\rho A\Omega\int_0^L (a\cos\beta - b\sin\beta)S_i U_j ds \quad (61)$$

$$[K_{i+3N,j}] = -(2GJ - EI)\beta'\left[\int_0^L U_i U_j' ds\right] + (2GJ - EI)\beta' b\left[\int_0^L U_i U_j'' ds\right] + \kappa AG\beta' b\int_0^L U_i U_j ds + \rho I\dot{\Omega}\int_0^L \sin\beta U_i U_j ds - F_a\left\{\begin{array}{l} -\int_0^L [(a-\varepsilon_c)\cos\beta'(s-L) + (b-\varepsilon_n)\sin\beta'(s-L)]U_i U_j' ds + (a-\varepsilon_c)U_i(L)U_j(L) \\ +\beta'\int_0^L [(b-\varepsilon_n)\cos\beta'(s-L) - (a-\varepsilon_c)\sin\beta'(s-L)]U_i U_j ds\end{array}\right\} \quad (62)$$

$$[C_{i+3N,j}] = -2\rho I\Omega\int_0^L \sin\beta U_i U_j ds \quad (63)$$

$$[K_{i+3N,j+N}] = -\kappa AG\beta' b\int_0^L U_i V_j' ds + \cos\alpha\frac{\rho aA\Omega^2}{2}\int_0^L (R^2 - s^2\cos^2\alpha - r^2 - 2sr\cos\alpha)U_i V_j'' ds - F_a\left\{\begin{array}{l} \beta'\int_0^L [(b-\varepsilon_n)\cos\beta'(s-L) - (a-\varepsilon_c)\sin\beta'(s-L)]U_i V_j' ds - \beta'(b-\varepsilon_n)U_i(L)V_j(L) \\ -(\beta')^2\int_0^L [(a-\varepsilon_c)\cos\beta'(s-L) + (b-\varepsilon_n)\sin\beta'(s-L)]U_i V_j ds - a\int_0^L U_i V_j'' ds\end{array}\right\} \quad (64)$$

$$[M_{i+3N,j+N}] = -\rho Aa\int_0^L U_i V_j ds \quad (65)$$

$$[C_{i+3N,j+2N}] = 2\rho A\Omega\int_0^L (a\cos\beta - b\sin\beta)U_i S_j ds \quad (66)$$

$$[K_{i+3N,j+3N}] = -GJ\int_0^L U_i U_j'' ds + \rho I\Omega^2\int_0^L \sin^2\beta U_i U_j ds - \cos\alpha\frac{J_0\rho\Omega^2}{2}\int_0^L (R^2 - s^2\cos^2\alpha - r^2 - 2sr\cos\alpha)U_i U_j'' ds - F_a(J_0/A)\int_0^L U_i U_j'' ds \quad (67)$$

$$[M_{i+3N,j+3N}] = \rho J_0 \int_0^L U_i U_j ds + \rho A(a^2+b^2) \int_0^L U_i U_j ds \quad (68)$$

All the terms of the matrices $\boldsymbol{M}$, $\boldsymbol{C}$, and $\boldsymbol{K}$ for a given beam dimension and the assumed displacement functions outlined in Eq. (38) can easily be determined by routine numerical integration method such as Simpson's rule. It should be noted that due to pretwist in the beam, we have kept all sine and cosine functions of the twist angle β inside the integration sign. During the spanwise integration of terms in the matrices, we determine the local value of $\beta(s)$ by the relationship described in Eq. (1) as

$$\beta(s) = \beta_r + s\beta' \quad (69)$$

Additionally, it can be seen that the rate of angular acceleration $\dot{\Omega}$ enters into the equations as the stiffness term. The above nonzero terms in the velocity-dependent coefficient matrix $[C_{i,j}]$ are all due to spin velocity Ω and represent the gyroscopic effect in the system by being skew symmetric in nature, which in the governing equations are shown as $\boldsymbol{C}_G$. The material internal damping can be taken into account as a function of the nondimensional factor χ of the critical damping of the beam material and the spin angular velocity Ω. In this situation, the typical terms in the damping matrix $[C_{i,j}]_D$ due to the material internal damping are computed as functions of stiffness matrix terms $[K_{i,j}]_S$ containing material parameters Young's modulus E and shear modulus G as well as shear coefficient κ of the beam, which are written as

$$[C_{i,j}]_D = \frac{2\chi}{\Omega}[K_{i,j}]_S \quad (70)$$

5 Sample Results of Fundamental Frequencies

The corresponding $\boldsymbol{M}$ and $\boldsymbol{K}$ matrices have been used for determining the nondimensional natural frequency term ξ by solving the following eigenvalue problem:

$$\boldsymbol{M}\{\ddot{f}(t)\} + [\boldsymbol{K}_S + \boldsymbol{K}_\Omega]\{f(t)\} = \{0\} \quad (71)$$

where nondimensional frequency parameter ξ is defined such that

$$\text{Natural frequency } \omega_N = \frac{\xi}{L^2}\sqrt{\frac{EI}{\rho A}} \text{ rad/s} \quad (72)$$

The eigenvalue solution of Eq. (71) yields natural frequencies both for rotating as well as nonrotating $(\Omega=0, \boldsymbol{K}_\Omega=0)$ conditions of the beam. It should be recognized that there is a scarcity of published data with all the parameters considered in the current analytical model, such as a blade rotating with an angular velocity Ω, radial lean α, initial twist $(\beta_R-\beta_r)$, coefficient of friction μ, longitudinal load F_a and its eccentricity $(\varepsilon_c, \varepsilon_n)$, etc. Thus, in order to demonstrate the accuracy of the present method, we have compared the finite-element results and other limited amount of published data with the natural frequency values yielded by the much simplified versions of the current model.

5.1 Current Model Validation With Finite-Element Results. In an attempt to validate the present model for its frequency response, the analytically predicted frequencies are compared with the finite-element results for a typical low-pressure compressor blade with the following parameters:

L	= span length of the airfoil (beam)	=15.8 cm
c/L	=aspect ratio	=0.43
M	= mass of the airfoil (ρAL)	=120 gm
α	= angle of lean with respect to radius	=0
I	= moment of inertia of the cross section (airfoil)	=0.04213 cm^4
E	= elastic Young's modulus of the beam material	=117 G Pa
ρ	= beam mass material density	=4.466 gm/cm^3
$(\beta_R-\beta_r)$	= total twist in the blade	=−25°
Ω	= angular velocity of rotation (40 Hz)	=251.327 rad/s

Here, the finite-element (FE) model (shown in Fig. 3) was analyzed using a commonly used commercial code called ANSYS. In the FE model in order to ensure that the shear deformation effect is included, we have used three brick elements through the airfoil thickness.

For this particular blade, the first five vibrational mode finite-element computed frequencies have been compared with those determined by the current model and are shown in Table 1 for the stationary condition and blades rotating at 2400 rpm, respectively. As one can see that the correlation for the first five modes with the FE model is very good with the maximum error limited to 3.32%.

5.2 Comparison of Current Model With Other Published Data. We have also verified the current analytical model with the results reported in published literature by other researchers as well. Using the pretwisted Timoshenko beam finite-element approach Yardimoglu and Yildirim [20] have computed the frequencies for the first four modes for a blade with the span-length L =15.24 cm, chord c=2.54 cm, depth d=0.17272 cm, and a total twist of $(\beta_R-\beta_r)$=45 deg. The material properties for this blade such as, Young's modulus of elasticity and the mass density are E=206.85 G Pa and ρ=7.8576 gm/cm^3, respectively. The frequencies reported by Yardimoglu and Yildirim for this blade and

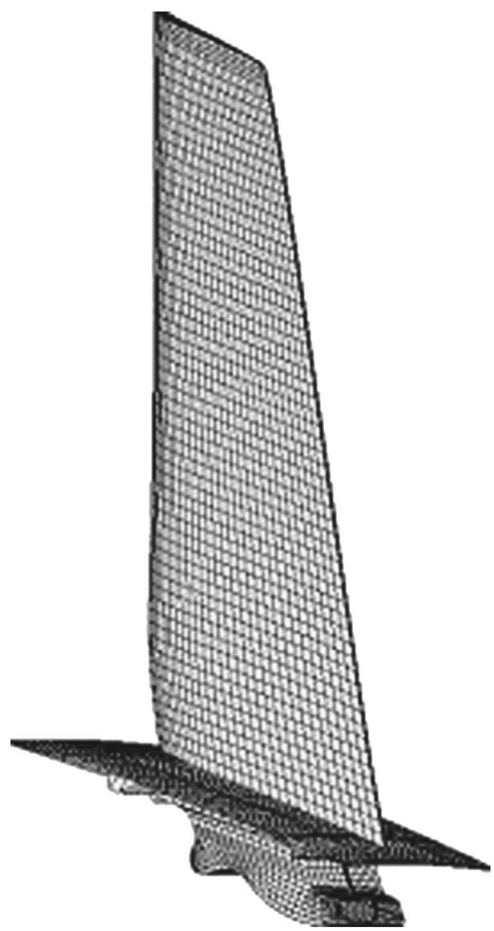


Fig. 3 Finite-element model of the blade (L=15.8 cm) with a total twist of $(\beta_R-\beta_r)$=−25 deg

Table 1 Comparison of natural frequency results computed using present analytical method versus finite-element results ($L=15.8$ cm, $a=0$, $(\beta_R-\beta_r)=-25$ deg, $c/L=0.430$)

Vibrational mode		Current analytical model		FE model with eight-noded brick elements	
Mode number	Mode shape	Nondimensional ξ	Frequency (Hz)	Frequency (Hz)	% difference
Stationary frequency (0 rpm)					
1	First flexural	3.6533	187.14	182	2.82
2	First torsion	13.8916	711.62	715	−0.47
3	Second flexural	22.8508	1170.57	1154	1.44
4	Second torsion	46.5902	2386.65	2310	3.32
5	Third flexural	61.2276	3136.47	3226	−2.78
Rotating frequency (2400 rpm)					
1	First flexural	4.1945	214.87	213	0.88
2	First torsion	13.9608	715.16	735	−2.70
3	Second flexural	23.3270	1194.96	1182	1.10
4	Second torsion	46.6035	2387.33	2350	1.59
5	Third flexural	61.7840	3164.97	3263	−3.00

the corresponding results computed by the current analytical model are shown in Table 2. It can be seen that their twisted beam FE model fails to capture the first torsional mode frequency completely.

Similarly, using the pretwisted Timoshenko beam bending equation in two directions, Banerjee [21] reports the frequencies only for the first three modes of vibration in his paper. In the sample problem to validate the results of his twisted Timoshenko beam model, Banerjee has computed these frequencies for a blade with the span-length $L=304.8$ cm, cross-sectional area $A=127.667$ cm^2, aspect ratio $(c/L)=0.667$, flexural rigidity $EI=14.3485\times10^{10}$ N cm^2, and a total twist of $(\beta_R-\beta_r)=40$ deg. The material properties for this blade such as, Young's modulus of elasticity and the mass density are $E=70$ G Pa and $\rho=2.7$ gm/cm^3, respectively. The frequencies reported by Banerjee [21] for this blade and the corresponding results computed by the current analytical model are shown in Table 2. Again, the analytical results from the current model for this particular case are in very good agreement with the maximum error limited to 3.6%.

5.3 Sample Numerical Results. For the presentation of numerical results from the current analytical model, the natural frequencies of the beam are computed in terms of nondimensional frequency parameter ξ for a very wide range of varying input parameters such as aspect ratio (c/L), total twist angle $(\beta_R-\beta_r)$, slenderness ratio $\bar{d}$, etc. These nondimensional results are shown in Figs. 4 and 5, for the untwisted and twisted beams, respectively. Figure 4 illustrates the drop in the values of nondimensional frequency parameter ξ with the increase in the aspect ratio from 0.1 (long beam) to 1.0 (square plate) for a typical value of total twist angle equal to zero (untwisted or flat beam). These values for the first two modes (first and second Flexural modes) match very well with the previous cantilever flat plate results reported by Harris and Crede [33], shown here with dotted line. The two torsion modes are also in reasonable agreement with the published data. Due to inherent limitation of the current beam model, the two-stripe vibrational mode for the flat plate shown by Harris and Crede is not picked up by the eigenvalue solution.

Figure 5 illustrates the effect of aspect ratio (c/L) variation on the changes in the values of the nondimensional frequency parameter ξ for a particular case of total twist angle of $(\beta_R-\beta_r)=45$ deg. In order to compare the results on side-by-side basis with those shown in Fig. 4, the values of the nondimensional frequency parameter ξ has been plotted for an identical range of increasing aspect ratio of (c/L) as the untwisted beam.

Table 2 Comparison of natural frequency results computed using present analytical method versus other published results for nonrotating ($\Omega=0$) twisted Timoshenko beam (Yardimoglu and Yildirim (see Ref. [20]), Banerjee (see Ref. [22]))

Vibrational mode		Current analytical model		Results from published literature	
Mode number	Mode shape	Nondimensional ξ	Frequency (Hz)	Frequency (Hz)	% difference
Yardimoglu and Yildirim[a] ($L=15.24$ cm, $\alpha=0$, $(\beta_R-\beta_r)=45$ deg, $c/L=0.167$)					
1	First flexural	3.6206	62.61	61.8	1.31
2	Second flexural	17.8910	309.38	304.8	1.50
3	First torsion	44.5846	770.98	Not shown	–
4	Twist-bend combination	55.0548	952.31	944.5	0.83
5	Third flexural	69.3112	1198.91	1193.0	0.50
Banerjee[b] ($L=304.8$ cm, $\alpha=0$, $(\beta_R-\beta_r)=40$ deg, $c/L=0.667$)					
1	First flexural	3.8123	39.42	38.05	3.60
2	First torsion	11.5461	119.39	122.51	−2.55
3	Second flexural	22.7070	234.79	226.87	3.49

[a]See Ref. [20].
[b]See Ref. [21].

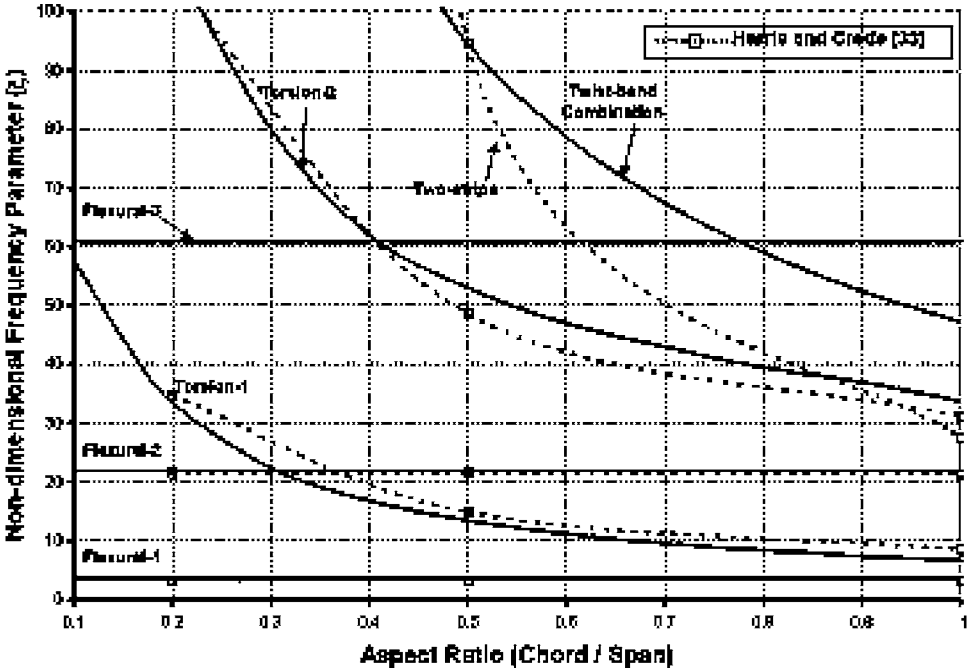


Fig. 4 Change in cantilever beam frequencies with no-twist as a function of aspect ratio (slenderness ratio=0.01)

The changes in the cantilever beam frequencies as a function of total twist angle for both stationary and rotating conditions are shown in Figs. 6–9 for aspect ratios=0.125, 0.25, 0.5, and 0.667, respectively. In these figures, we have plotted the value of nondimensional frequency parameter ξ for the first six modes. However, as the twist angle β or the aspect ratio (c/L) changes, some of the mode shapes also change as these lines cross each other. Due to these mode-crossing conditions, for example mode 2 may be first torsional mode for one twist angle, but it may become second flexural mode for some other twist angle or aspect ratio. It is observed that there is a small increase in the computed frequencies for the flexural modes as the angle of twist increases, however, the fundamental frequencies for the modes associated with the torsion about the span direction decreases rapidly with the increasing twist. In addition, the presence of centrifugal force field always tends to increase the frequencies with respect to its values in stationary condition due to stress stiffening. This trend is similar to the one observed by Hu and his co-workers [22].

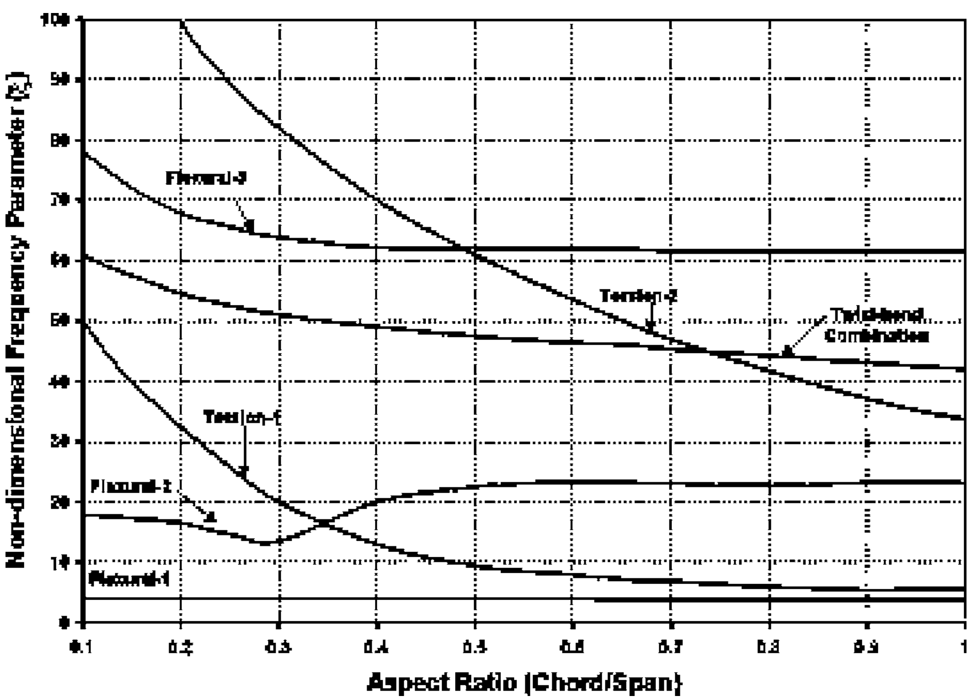


Fig. 5 Change in cantilever beam frequencies twisted at 45 deg as a function of aspect ratio (slenderness ratio=0.01)

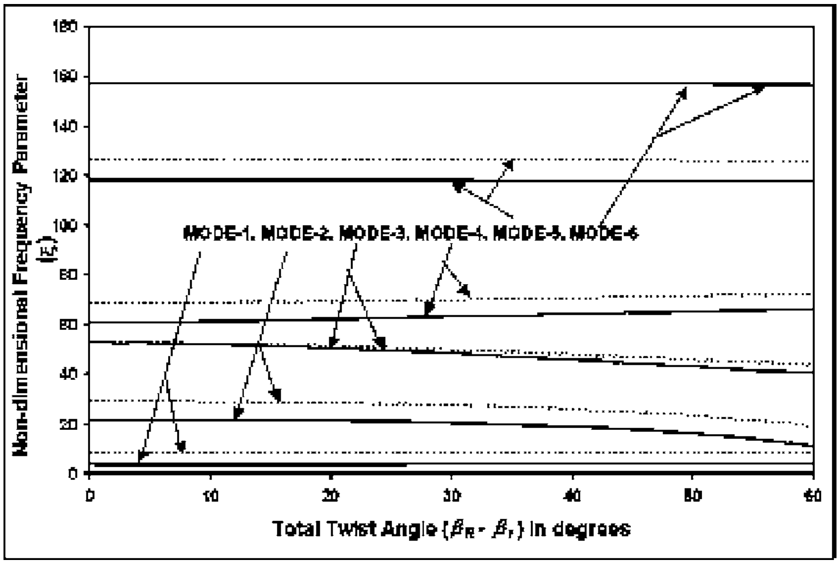


Fig. 6 Change in the twisted cantilever beam frequencies with aspect ratio (chord/span)=0.125 as a function of the total twist angle (slenderness ratio=0.01, angular velocity Ω=0.0 (stationary)—, angular velocity Ω=300 rad/s (rotating)- - -)

It should be noted that the present derivation for a pretwisted beam is completely different than other researchers' formulation [20,21,23] of using product moment of inertia terms such as, I_{yy}, I_{xy}, I_{yx}, etc. instead of torsional constant term J and shear center term (a,b). For a general asymmetric cross section such as a typical airfoil or hollow blades and turbine blades with cooling holes, the numerical computation of terms like I_{xy}, I_{yx} about the CG of the cross section is extremely cumbersome. The current approach is very convenient to analyze the dynamics of pretwisted asymmetric thin cross section such as typical airfoil blades. In addition, all the earlier researchers' derivation disregards the coupling effect of axial motion of the beam. Until now, all the axial-bending coupling investigations have been limited to axial force being treated as a buckling load on a column [23–26] rather axial motion being considered as a separate degree of freedom. The natural frequencies associated with the axial mode of vibrations (ζ degree of freedom) for the rotating Timoshenko beam have been determined and reported by the author in his previous work [9].

6 Transient Analysis Results With Contact Impact at the Free End

The transient motions caused by the periodic tip-impact load along the longitudinal axis of the beam column initiates the high-frequency axial mode of vibrations, which interacts with the low-

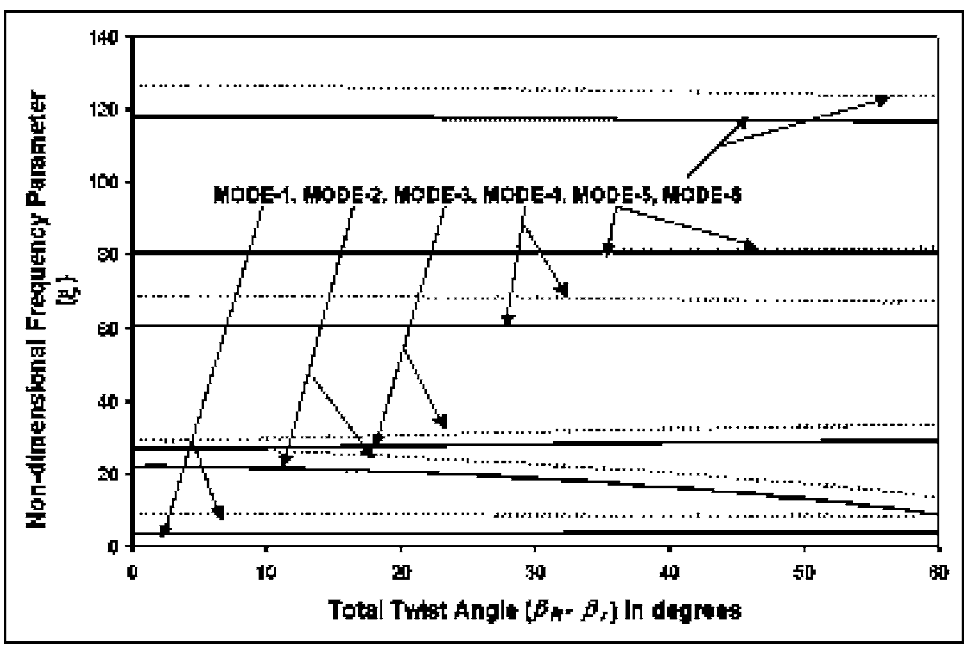


Fig. 7 Change in the twisted cantilever beam frequencies with aspect ratio (chord/span)=0.25 as a function of the total twist angle (slenderness ratio=0.01, angular velocity Ω=0.0 (stationary)—, angular velocity Ω=300 rad/s (rotating)- - -)

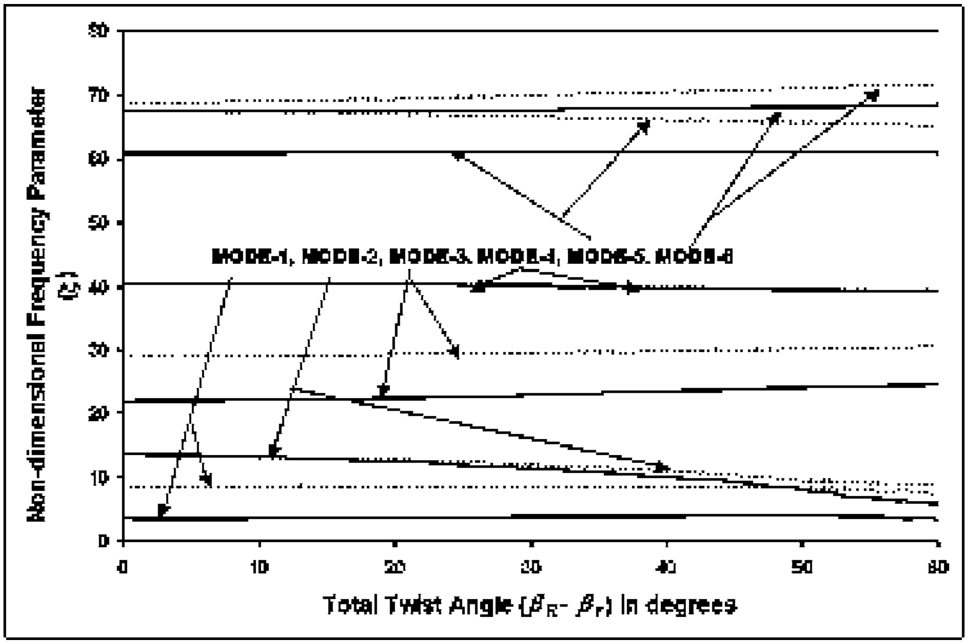


Fig. 8 Change in the twisted cantilever beam frequencies with aspect ratio (chord/span)=0.5 as a function of the total twist angle (slenderness ratio=0.01, angular velocity Ω=0.0 (stationary)—, angular velocity Ω=300 rad/s (rotating)- - -)

frequency flexural-bending mode oscillations of the beam and changes its dynamic response considerably during spin. The dynamic coupling of axial motion with the lateral deflection in a spinning beam introduces the Coriolis forces, which have very significant effect on the rub-induced vibration in a rotating machinery.

6.1 Analytically Predicted Transient Response Versus Strain Gage Data. In order to establish the accuracy of the current analytical model for its time-domain results, the numerically computed transient dynamic response of a typical high-pressure compressor blade is compared with the measured strain-gage data from a rig test [34]. In this especially developed experimental rig, it was observed that during a controlled periodic rub scenario each rub-impact produced a somewhat different transient dynamic characteristics than the one preceding rub event until the rub-induced vibration of the blade reached to a limit-cycle response under repeated rubs. For this part of the analysis using the sixth order Runge–Kutta scheme, a direct-time integration of the equations of motion outlined in Eq. (39) is performed. The aspect ratio (c/L) of the test blade is 0.659 with the span length L equal to 4 cm, and it is rotating at 16,500 rpm with the tip tangential velocity of 400 m/s. The radial lean angle is α=0 for this blade and the stagger angle is $\beta_r = \beta_R = -45$ deg. The bade tip rubs against a 72 deg circumferential rub zone with the contact-impact tip load-

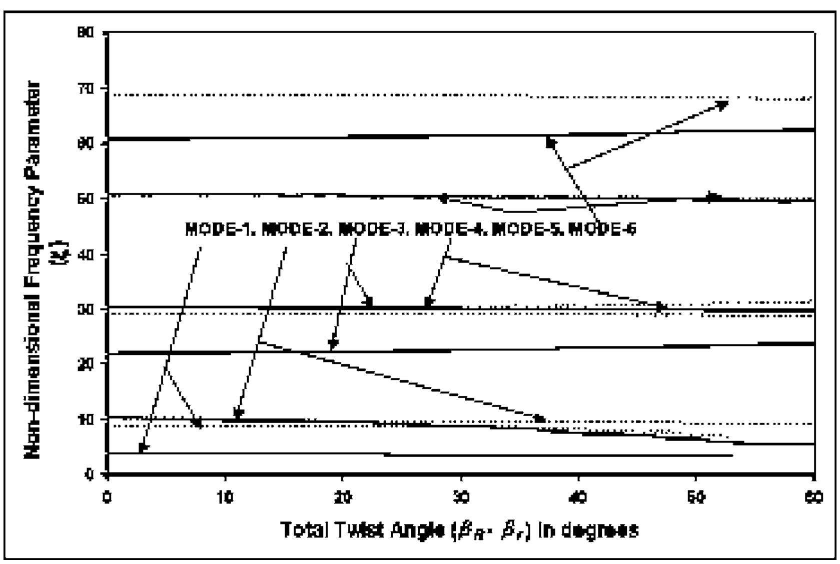


Fig. 9 Change in the twisted cantilever beam frequencies with aspect ratio (chord/span)=0.667 as a function of the total twist angle (slenderness ratio=0.01, angular velocity Ω=0.0 (stationary)——, angular velocity Ω=300 rad/s (rotating) – – –)

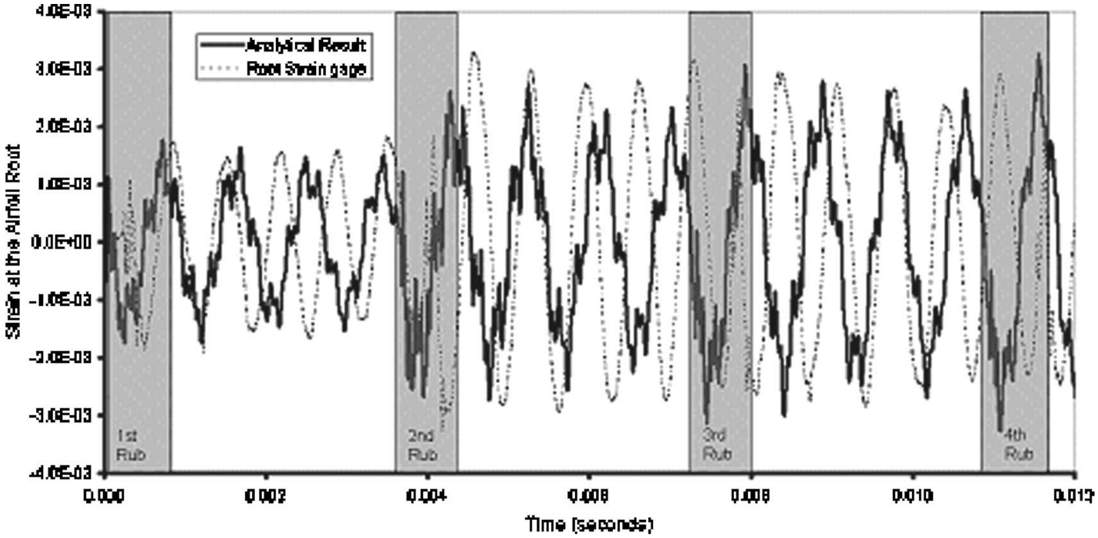


Fig. 10 Comparison of measured airfoil root strain gage data (……) **versus results from the present analytical model (—) during repeated radial incursion of 0.1 mm at the blade tip with 72 deg circumferential rub zone for the first four rubs**

ing applied at the rate of one pulse per revolution with 0.1 mm radial interference. In the rig test, the 72 deg forced rub zone is created by inserting a partial sector of a circumferential shoe in the path of the moving blade tip. The transient analysis has been carried out for four repeated impacts. The measured dynamic data from a spanwise strain gage at the fillet of the airfoil root is compared with the numerically computed strain time history near the clamped end of the corresponding Timoshenko beam model using the current analytical technique. From the two sets ofplotted data illustrated in Fig. 10, it can be seen that the transient analytical results predicts the dynamic characteristics and the resulting strain time history in the rubbing blade very well.

Numerically computed response shows highly nonlinear behavior of the airfoil root strains. However, in terms of frequency response, the analytical model responds at a slightly lower frequency than the test data. This can be attributed to the nonlinearity in the boundary conditions at the tip during the actual rub event. In the rig test the tip is partially constrained during the rub, whereas in the analytical model it is considered free with rub-related forces as external loads on the system. In addition, both the analytical and the test data illustrate as to how the magnitude of the response builds up after the first rub, until it stabilizes after about third rub. On this plot, the dynamic response of the blade as it passes through the 72 deg circumferential rub zone, is shown by rectangular shaded areas with legends as the first, second, third, and fourth rubs. As the spinning blade tip comes out of the forced-rub zone, the extensional wave in the blade travels up-and-down its longitudinal axis with very high velocity, giving rise to large Coriolis forces, which are oscillatory in nature. Mathematically, this Coriolis force is a distributed load; which is represented by the term such as $-2\rho A\Omega \cos \beta\zeta_{,t}$ shown in Eq. (21b), and its value at the blade root is computed as the integrated sum over the span length of the blade given by

$$\text{Coriolis force at the airfoil root} = -2\Omega\rho A\int_0^L \cos(\beta_r + \beta' s)\zeta_{,t} ds \tag{73}$$

The transient characteristics of this Coriolis force is shown in Fig. 11, the magnitude of which at the airfoil root could be as high as 3200 G. For this test blade, whereas the first-flex bending mode frequency is 1600 Hz, the longitudinal wave frequency is about 33,400 Hz. These longitudinal stress waves, frequency of which is more than 20 times higher than the flexural bending waves, can generate intense heat at the mating surface of the blade root (dovetail) with the disk, which has been observed to result into local welding and fretting of the mating surfaces as well as severe bearing damage even after short-duration heavy rubs. It should be noted that the decay in the magnitude of the Coriolis force, outside the imposed rub zone during the free vibration of the blade, is not due to damping in the system rather it is caused by the transfer of kinetic energy associated with the longitudinal motion into the lowest vibrational mode frequency, which invariably corresponds to its first flexural bending mode motion. The transient vibratory dynamic stresses in a rubbing airfoil is due to the interaction of longitudinal motion (hyperbolic wave) with the lateral motion (dispersive wave) of the beam, which in the case of rub-induced dynamic instability results into fatigue-type damage to the blade. Depending upon the eccentricity $(\varepsilon_c, \varepsilon_n)$ of the rub location at the blade-tip cross section, these rub-related damages can range from local tip curl due to plasticity to complete separation of the airfoil at the blade root.

6.2 Effect of Pre-twist on Transient Response During Tip-Rub. In this section, we will apply the current analytical model to investigate the transient response of a twisted blade as opposed to a similar blade but without any pretwist. For this investigation, we will use the blade parameters same as outlined in Sec. 6.1 with two different values of the total twist, that is $(\beta_R - \beta_r) = 0$ deg and $(\beta_R - \beta_r) = -45$ deg. Here, we have compared the dynamic responses of these two blades, subjected to the same contact-impact loads during a typical rub event of one pulse per revolution, in terms of their nondimensional lateral tip-deflection (η/L). The main difference of external load application between the results shown in the previous section to the current section is that the previous section results were generated for the rig-test conditions with displacement-controlled radial incursion of 0.1 mm, whereas in this section the results are for outer casing-imposed radial force of $F_{\max} = 0.1$ times of the Euler critical buckling load applied at the free end of the beam. The numerical technique to implement the controlled radial incursion or, longitudinal pulse of controlled magnitude in a transient simulation has been discussed in detail in the author's previous work [9] on the related topic. The respective results of the lateral tip-deflection (η/L) are plotted in Fig. 12.

These time-history plots up to 2.5 m s clearly show that for a

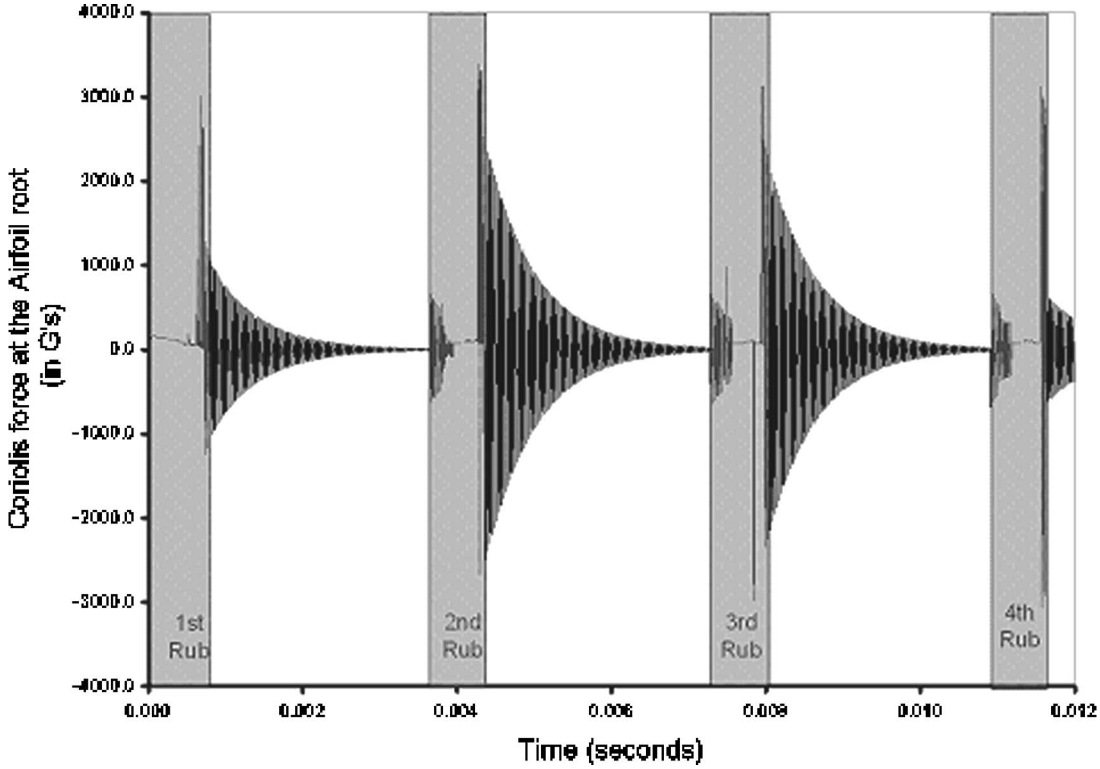


Fig. 11 Analytically computed Coriolis forces at the blade root during repeated radial incursion of 0.1 mm at the blade tip with 72 deg circumferential rub zone for the first four rubs

periodic contact-impact load of the same magnitude, the response of an untwisted blade is monotonically increasing, whereas for blade twisted at 45 deg the dynamic response shows a beating pattern. The beating pattern of a quasi-periodic nature indicates that this dynamic system is responding simultaneously at two different frequencies, which are very close to each other. It is observed that for both twisted and untwisted blades, the dynamic response until the fourth contact-impact pulse is almost identical. In addition, after each pulse loading the twisted blade responds at a slightly higher frequency than an untwisted blade.

7 Concluding Remark

The present analytical model captures the full dynamics of pretwisted cantilever Timoshenko beam with combined torsional-

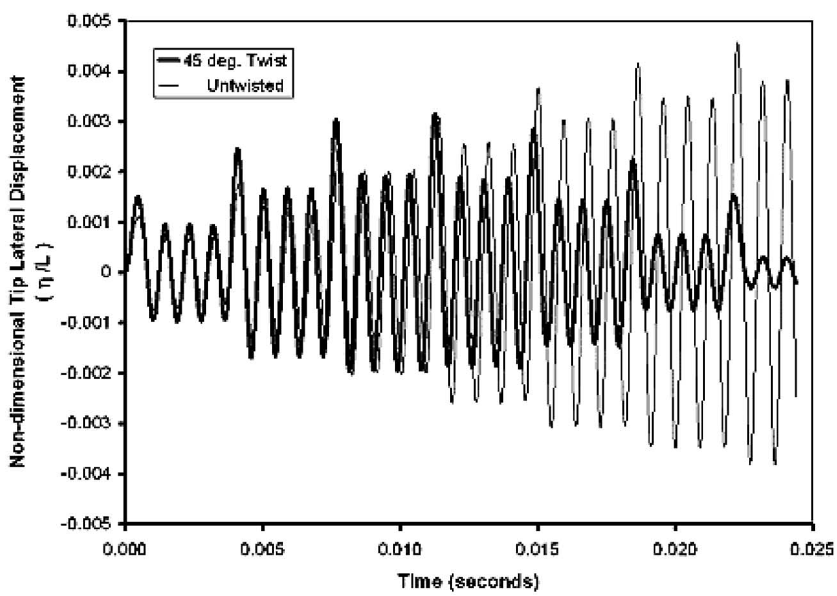


Fig. 12 Comparison of analytically computed transient lateral tip displacements for an untwisted beam (—) versus a 45 deg twisted beam (.....) during repeated rubs (one pulse per revolution) at the blade tip with a periodic contact force of magnitude F_{max}=0.1× (Euler critical buckling load)

bending-axial motion subjected to contact load F_a including Coulomb friction μ at the tip. The axial force F_a accounts for dynamic buckling effect in the event of contact-impact load at the free end. In the contact-impact scenario, the axial force F_a is a transient load represented by time-dependent function $F(t)$. As shown in the author's previous work [9] on periodic tip-pulse-loading, the general wave form of the dynamic force $F(t)$ with a frequency of $fp-Hz$ along the longitudinal axis of the Timoshenko beam can have many different time-dependent distributions, such as half-sine wave, triangular pulse, rectangular pulse, full-cosine wave with an offset, sawtooth profile, etc. The dynamic characteristics of the twisted beam are expressed by a set of four partial differential equations. These equations contain not only terms due to displacement-dependent forces rather they also include important, but very rarely derived velocity-dependent forces as well. By introducing four assumed displacement functions, the terms containing spatial coordinates s are eliminated from the equations by using Rayleigh–Ritz technique. We have formulated every term including forces due to Coriolis effect in the form of conventional $\boldsymbol{M}$, $\boldsymbol{C}$, and $\boldsymbol{K}$ matrices. The main limitation of the current beam model is its inability to obtain the classical two-stripe mode of the rotating blade. In the airfoil blade dynamics, it is well known that two-stripe mode is an important mode of vibration to be concerned, especially for short airfoils with $(c/L)\rightarrow 1$. In order to predict the two-stripe mode correctly, one must consider the coupled beam bending formulation in two planes. It is worth noting that the Timoshenko beam model developed here can be easily expanded to the more general case of coupled bending deformation in two principal planes ($x-x$ and $y-y$) as outlined in the Appendix. We would like to point out that in the coupled two-plane-bending formulation of the beam with contact-impact loading, the number of independent degrees-of-freedom in the dynamical system suddenly jumps from 4 to 6; which makes it more challenging to solve due to added complexity.

The equations of motion given in Eq. (39) are also integrated by the Runge–Kutta method to obtain the transient dynamic response in the time domain under different types of contact-impact loading at the blade tip. For accurate computing of the beam dynamic deformations associated with very high strain rates and other nonlinearities, the direct integration of equations of motion is a much preferred technique, which is also used for determining the rub-induced dynamic instability of rotating twisted blades. Using the current analytical model, we are able to predict the transient response of a rotating blade subjected to repeated rub impacts; which depending upon the contribution of various parameters such as rotational speed Ω, coefficient of friction μ, longitudinal contact load at the blade-tip F_a, load eccentricity $(\varepsilon_c, \varepsilon_n)$, etc., can make a typical rub either unstable by showing a growth in the amplitude of lateral oscillations or turn into a stable rub as a limit-cycle response.

Nomenclature

(a,b) = x-coordinate and y-coordinate of the shear center, respectively
A = cross-sectional area of the beam or blade
c = chord length of the blade airfoil cross section
$\boldsymbol{C}$ = general coefficient matrix for velocity-dependent forces
$\boldsymbol{C}_{\mathrm{D}}$ = damping matrix (symmetric)
$\boldsymbol{C}_{\mathrm{G}}$ = gyroscopic matrix (skew symmetric, causes forward and backward frequency shift in the blade)
$[C_{i,j}]$ = typical ith row and jth column term in the velocity-dependent matrix
d = depth of the blade airfoil cross section
$\bar{d}$ = slenderness ratio of the blade airfoil cross section $=\sqrt{I/AL^2}$
$(\hat{e}_a, \hat{e}_t, \hat{e}_r)$ = unit vectors in the local axial-tangential-radial system
$(\hat{e}_c, \hat{e}_n, \hat{e}_s)$ = unit vectors in the local chord-normal-span system
E = Young's modulus of elasticity of the blade or beam material
EI = flexural rigidity of the blade cross section about local x-axis (minor principal direction)
EI_{yy} = flexural rigidity of the blade cross section about local y-axis (major principal direction)
f_p = pulse frequency of the blade tip contact-load in Hz
$\boldsymbol{\mathcal{F}}$ = generalized external force vector at the blade tip
F_a = axial load on the blade (along the span direction of the blade)
F_{cf} = centrifugal force at the blade airfoil CG
F_C, F_N, F_S = components of external force vector at blade-tip $\boldsymbol{\mathcal{F}}$ due to contact
$\{f(t)\}$ = column vector containing generalized time-dependent displacement coordinates of the dynamical system
$F(t)$ = time-dependent axial load on the blade due to contact impact (along the span direction of the blade)
G = shear modulus of the blade or beam material
GJ = torsional rigidity of the thin blade cross section
I_{xx}, I_{yy} = principal moment of inertias of the blade cross section
J_o = polar moment of inertia of the blade cross section $=(I_{xx}+I_{yy})$
K_{case} = radial stiffness of the outer case filler material during rub
$\boldsymbol{K}$ = general coefficient matrix for displacement-dependent forces
$\boldsymbol{K}_F$ = in-plane force-dependent circulatory matrix due to contact force F_a
$\boldsymbol{K}_S$ = elastic stiffness matrix (symmetric)
$\boldsymbol{K}_\Omega$ = stress-stiffening or softening matrix due to spin velocity Ω
$[K_{i,j}]$ = a typical ith row and jth column term in the stiffness matrix
L = span length of the cantilever blade
$\boldsymbol{M}$ = general coefficient matrix for acceleration-dependent forces (symmetric)
$[M_{i,j}]$ = typical ith row and jth column term in the mass matrix
M = mass of the airfoil or, cantilever beam (ρAL)
$\boldsymbol{\mathcal{M}}$ = generalized external moment vector at the blade tip
M_C, M_N, M_S = components of external moment vector at blade-tip $\boldsymbol{\mathcal{M}}$ due to contact
$\{P(t)\}$ = column vector containing external forces on the dynamical system (components of this vector: $P^\psi, P^\eta, P^\zeta, P^\phi$)
$Q(s,t)$ = shear force at span location s and time t
Q_n, Q_s = distributed lateral loads on the beam in the transverse and longitudinal directions (per unit length)
R = blade tip radius

r = blade root radius or disk outer radius
s = blade local coordinates in the span direction
t = time (s)
T_c, T_s = distributed moments on the beam about the chord and longitudinal directions (per unit length)
T = total kinetic energy of the blade
U = total potential energy of the blade
$S_j(s), U_j(s), V_j(s)$ = sinusoidal shape functions for blade deformation ($j=1,2,3,\ldots,N$)
$\boldsymbol{V}$ = velocity vector of any typical point on the airfoil or beam
$W_j(t), X_j(t), Y_j(t), Z_j(t)$ = time-dependent generalized coordinates for dynamic deflection of the blade ($j=1,2,3,\ldots,N$)

Greek Symbols

α = sweep or blade lean angle with respect to the radial direction
β = blade twist or stagger angle (rad), i.e., angle between the blade chord and the engine axis (axis of rotation) at the blade tip
β_r, β_R = twist angle of the blade cross section at radii r and R
$(\beta_R - \beta_r)$ = total twist in the blade over the span length L
β' = rate of pretwist of the blade in the span direction
$\delta(s-L)$ = dirac delta unit impulse function for values at $s=L$
γ = radial clearance at the blade tip with respect to the case inner radius
$\varepsilon_c, \varepsilon_n$ = contact load eccentricity in the local chord and normal direction
ϕ, ψ, η, ζ = blade deformation due to twist, cross-section rotation, lateral deflection, and longitudinal deflections, respectively.
θ = angle of the rigid body rotation of the shaft about the spin axis at time t from time 0 ($\theta = \Omega t$ for constant angular velocity Ω)
κ = shear coefficient in the Timoshenko beam formulation
Λ = Lagrangian parameter
μ = coefficient of friction between the blade tip and the outer case
ν = Poisson's ratio of the blade material $(d\theta/dt)$
ξ = nondimensional beam frequency parameter
$\mathfrak{R}$ = rotation vector for small rotations of the airfoil cross section
χ = critical damping parameter of the blade material (nondimensional)
ω_N = natural frequency (rad/s)
Ω = blade spin velocity (rad/s)$=(d\theta/dt)$
φ_j = $(2j-1)\pi/2L$

Appendix: Effect of Beam Bending in Two Principal Directions

In the current formulation, we have considered that $I_{yy} \gg I_{xx}$, which inherently assumes that the lateral deformation in the beam will be dominated by (y-y) direction displacements. If one wants to explicitly include the deformations in the other principal direction (x-x) as well, it would result into two additional degrees-of-freedom for η and ψ. For example, in addition to twist ϕ and extension ζ of the cross section, we will have to deal with (η_y, ψ_x) due to bending about the principal (x-x) direction and (η_x, ψ_y) due to bending about the principal (y-y) direction. The resulting equations for the Timoshenko beam formulation due to fully coupled two-directional bending alone are too cumbersome to derive here and are shown in the following for reference purposes as:

(a) Bending moment balance about local (x-x) axis

$$-EI_{xx}(\psi_{x,ss}) - \kappa AG(\eta_{y,s} - \psi_x) + \beta' EI_{yy}(\beta' \psi_x) + \beta' \kappa AG \eta_x - \beta' EJ_0 \psi_{y,s} - \rho I_{xx}\Omega^2 \psi_x + \rho I_{xx}(\psi_{x,tt}) = T_c(s,t) \tag{A1}$$

(b) Shear force balance in the local (y-y) direction

$$-\kappa AG(\eta_{y,ss} - \psi_{x,s}) + 2\kappa AG\beta' \eta_{x,s} + \beta'^2 \kappa AG \eta_y + \beta' \kappa AG \psi y - F_{cf}\cos\alpha\, \eta_{y,ss} + \rho A\, \eta_{y,tt} = Q_n(s,t) \tag{A2}$$

(c) Bending moment balance about local (y-y) axis

$$-EI_{yy}(\psi_{y,ss}) - \kappa AG(\eta_{x,s} + \psi_y) + \beta' EI_{xx}(\beta' \psi_y) + \beta' \kappa AG \eta_y + \beta' EJ_0 \psi_{x,s} - \rho I_{yy}\Omega^2 \psi_y + \rho I_{yy}(\psi_{y,tt}) = T_n(s,t) \tag{A3}$$

(d) Shear force balance in the local (x-x) direction

$$-\kappa AG(\eta_{x,ss} + \psi_{y,s}) - 2\kappa AG\beta' \eta_{y,s} + \beta'^2 \kappa AG \eta_x + \beta' \kappa AG \psi_x - F_{cf}\cos\alpha\, \eta_{x,ss} + \rho A\, \eta_{x,tt} = Q_c(s,t) \tag{A4}$$

The corresponding new geometric boundary conditions are as follows:

$$\psi_x(0,t) = 0, \quad \eta_y(0,t) = 0, \quad \psi_y(0,t) = 0, \quad \eta_x(0,t) = 0 \tag{A5}$$

The additional four natural boundary conditions at the free end of the cantilever beam for ($s=L$) are expressed in terms of the contact-impact force vector $\boldsymbol{F}$ and moment vector $\boldsymbol{M}$ components (see Eq. (3)) as

$$(\text{Bending moment}_{xx})_{\text{at } s=L} = M_C = EI\psi_{x,s}|_{s=L} = F_a[\varepsilon_n + \varepsilon_c \sin(\beta' L)] \tag{A6}$$

$$(\text{Shear force}_{yy})_{\text{at s=L}} = F_N = -\kappa AG[\eta_{y,s} - \psi_x]_{s=L} = -\mu F_a \cos\beta_R \tag{A7}$$

$$(\text{Bending moment}_{yy})_{\text{at s=L}} = M_N = EI\psi_{y,s}|_{s=L} = -F_a[\varepsilon_c + \varepsilon_n \cos(\beta' L)] \tag{A8}$$

$$(\text{Shear force}_{xx})_{\text{at s=L}} = F_C = -\kappa AG[\eta_{x,s} - \psi_y]_{s=L} = -\mu F_a \sin\beta_R \tag{A9}$$

It should be noted that additional coupling terms will appear for the partial differential equations governing the beam motion for extension ζ and twist ϕ. Using the above equations, the interested researchers can easily expand Eqs. (21a)–(21d) to take into account of coupled bending effect about (y-y) axis.

References

[1] Carnegie, W., and Dawson, B., 1971, "Vibration Characteristics of Pre-Twisted Blades of Asymmetrical Aerofoil Cross-Section," Aeronaut. Q., **1**(3), pp. 257–273.
[2] Fu, C. C., 1974, "Computer Analysis of a Rotating Axial Turbo Machine Blade in Coupled Bending-Bending-Torsion Vibrations," Int. J. Numer. Methods Eng., **8**, pp. 569–588.
[3] Feldt, W. T., and Hermann, G., 1974, "Bending-Torsional Flutter of a Cantilevered Wing Containing a Tip Mass and Subjected to a Transverse Follower Force," J. Franklin Inst., **297**, pp. 467–478.
[4] Pigot, R., and Abel, J. M., 1975, "Vibrations and Stability of Turbine Blades at Stall," ASME J. Eng. Power, **96**, pp. 201–208.
[5] Lin, C. Y., and Chen, L. W., 2003, "Dynamic Stability of Rotating Pre-Twisted Blades with a Constrained Damping Layer," Compos. Struct., **61**, pp. 235–245.
[6] Coller, B. D., and Chamara, P. A., 2004, "Structural Non-Linearities and the

Nature of the Classic Flutter Instability," J. Sound Vib., **277**(4–5), pp. 711–739.
[7] Chen, L. W., and Peng, W. K., 1995, "Dynamic Stability of Rotating Blades With Geometric Non-Linearity," J. Sound Vib., **187**(3), pp. 421–433.
[8] Hodges, D. H., 2001, "Lateral-Torsional Flutter of a Deep Cantilever Loaded by a Lateral Follower Force at the Tip," J. Sound Vib., **247**(1), pp. 175–183.
[9] Sinha, S. K., 2005, "Non-Linear Dynamic Response of a Rotating Radial Timoshenko Beam With Periodic Pulse Loading at the Free-End," Int. J. Non-Linear Mech., **40**(1), pp. 113–149.
[10] Chen, C. K., and Ho, S. H., 1999, "Transverse Vibration of a Rotating Twisted Timoshenko Beams Under Axial Loading Using Differential Transform," Int. J. Mech. Sci., **41**, pp. 1339–1356.
[11] Timoshenko, S., Young, D. H., and Weaver, W., 1974, *Vibration Problems in Engineering*, 4th ed., Wiley, New York, pp. 432–434.
[12] Leissa, A., and Jacob, K. I., 1986, "Three-Dimensional Vibrations of Twisted Cantilevered Parallelepipeds," ASME J. Appl. Mech., **53**, pp. 614–618.
[13] Rosen, A., 1991, "Structural and Dynamic Behavior of Pretwisted Rods and Beams," Appl. Mech. Rev., **44**, pp. 483–515.
[14] Lin, S. M., Wang, W. R., and Lee, S. Y., 2001, "The Dynamic Analysis of Nonuniformly Pre-Twisted Timoshenko Beam With Elastic Boundary Conditions," Int. J. Mech. Sci., **43**, pp. 2385–2405.
[15] Petrov, E., and Geradin, M., 1998, "Finite Element Theory for Curved and Twisted Beams Based on Exact Solutions for Three Dimensional Solids. Part 1: Beam Concept and Geometrically Exact Nonlinear Formulation, Part 2: Anisotropic and Advanced Beam Models," Comput. Methods Appl. Mech. Eng., **165**, pp. 43–127.
[16] Tang, S., and Yu, A., 2004, "Generalized Variational Principle on Nonlinear Theory of Naturally Curved and Twisted Beams," Appl. Math. Comput., **153**, pp. 275–288.
[17] Lee, S. Y., and Lin, S. M., 1994, "Bending Vibrations of a Rotating Nonuniform Timoshenko Beam With an Elastically Restrained Root," ASME J. Appl. Mech., **61**, pp. 949–955.
[18] Oguamanam, D. C. D., and Heppler, G. R., 1998, "Geometric Stiffening of Timoshenko Beams," ASME J. Appl. Mech., **65**, pp. 923–929.
[19] Lin, S. M., 1999, "Dynamic Analysis of Rotating Nonuniform Timoshenko Beams With an Elastically Restrained Root," ASME J. Appl. Mech., **66**, pp. 742–748.
[20] Yardimoglu, B., and Yildirim, T., 2004, "Finite Element Model for Vibration Analysis of Pre-Twisted Timoshenko Beam," J. Sound Vib., **273**, pp. 741–754.
[21] Banerjee, J. R., 2004, "Development of an Exact Dynamic Stiffness Matrix for Free Vibration Analysis of a Twisted Timoshenko Beam," J. Sound Vib., **270**, pp. 379–401.
[22] Hu, X. X., Sakiyama, T., Matsuda, H., and Morita, C., 2004, "Fundamental Vibration of Rotating Cantilever Blades With Pre-Twist," J. Sound Vib., **271**, pp. 47–66.
[23] Chen, W. R., and Keer, L. M., 1993, "Transverse Vibration of a Rotating Twisted Timoshenko Beam Under Axial Loading," ASME J. Vibr. Acoust., **115**, pp. 285–294.
[24] Lee, H. P., 1994, "Buckling and Dynamic Stability of Spinning Pre-Twisted Beams Under Compressive Axial Loads," Int. J. Mech. Sci., **36**(11), pp. 1011–1026.
[25] Liao, C. L., and Huang, B. W., 1995, "Parametric Instability of a Spinning Pretwisted Beam Under Periodic Axial Force," Int. J. Mech. Sci., **37**(4), pp. 423–439.
[26] Sakar, G., and Sabuncu, M., 2004, "Buckling and Dynamic Stability of a Rotating Pretwisted Asymmetric Cross-Section Blade Subjected to an Axial Periodic Load," Finite Elem. Anal. Design, **40**, pp. 1399–1415.
[27] Yang, S. M., and Tsao, S. M., 1997, "Dynamics of a Pretwisted Blade Under Nonconstant Rotating Speed," Comput. Struct., **62**(4), pp. 643–651.
[28] Temel, B., and Calim, F. F., 2003, "Forced Vibration of Cylindrical Helical Rods Subjected to Impulsive Loads," ASME J. Appl. Mech., **70**, pp. 281–291.
[29] Temel, B., 2004, "Transient Analysis of Viscoelastic Helical Rods Subject to Time-Dependent Loads," Int. J. Solids Struct., **41**, pp. 1605–1624.
[30] Turhan, O., and Bulut, G., 2005, "Dynamic Stability of Rotating Blades (Beams) Eccentrically Clamped to a Shaft With Fluctuating Speed," J. Sound Vib., **280**(3–5), pp. 945–964.
[31] Timoshenko, S. P., and Goodier, J. N., 1970, *Theory of Elasticity*, 3rd ed., McGraw-Hill, New York, pp. 307–309.
[32] Timoshenko, S. P., and Gere, J. M., 1961, *Theory of Elastic Stability*, 2nd ed., McGraw-Hill, New York, pp. 212–250.
[33] Harris, C. M., and Crede, C. E., 1976, *Shock and Vibration Handbook*, 2nd ed., McGraw-Hill, New York, pp. 7–14.
[34] Padova, S., Barton, J., Dunn, M. G., Manwaring, S., Young, G., Adams, M., Jr., and Adams, M., 2004, "Development of an Experimental Capability to Produce Controlled Blade Tip/Shroud Rubs at Engine Speed," *Proceedings of IGTI: ASME/IGTI Turbo Expo 2004*, Vienna, Austria, June 14–17, Paper No. GT2004-53322.

R. P. Shimpi[1]
Professor
e-mail: rpshimpi@aero.iitb.ac.in

H. G. Patel
Research Scholar

H. Arya
Assistant Professor

Aerospace Engineering Department,
Indian Institute of Technology Bombay,
Powai, Mumbai 400076, India

New First-Order Shear Deformation Plate Theories

First-order shear deformation theories, one proposed by Reissner and another one by Mindlin, are widely in use, even today, because of their simplicity. In this paper, two new displacement based first-order shear deformation theories involving only two unknown functions, as against three functions in case of Reissner's and Mindlin's theories, are introduced. For static problems, governing equations of one of the proposed theories are uncoupled. And for dynamic problems, governing equations of one of the theories are only inertially coupled, whereas those of the other theory are only elastically coupled. Both the theories are variationally consistent. The effectiveness of the theories is brought out through illustrative examples. One of the theories has striking similarity with classical plate theory. [DOI: 10.1115/1.2423036]

1 Introduction

It is now well-known that in plate analysis, shear deformation effects become important not only for thick plates but even for thin plates vibrating at higher modes. As classical plate theory (CPT) does not take into account shear effects, many theories got evolved to address the deficiency.

First-order shear deformation theory (FSDT) proposed by Reissner and another one proposed by Mindlin are considered to be pioneering theories which take into account shear effects.

Reissner [1,2], used stress based approach to develop his theory. Later, while at the same level of approximation, Mindlin [3], in 1951, employed displacement based approach. As per Mindlin's theory, transverse shear stresses turn out to be constant through the thickness of the plate, but this violates the shear stress free surface conditions. Mindlin's theory satisfies constitutive relations for transverse shear stresses and shear strains only in an approximate manner by way of using shear correction factor. A good discussion about theories of Reissner and Mindlin is available in a paper by Wang et al. [4].

After FSDTs of Reissner and Mindlin, a class of theories, generally known as "higher order theories," was developed by various researchers.

Some important higher order theories are available in the literature. These include theories by various researchers: Murty [5] with 5, 7, 9, … unknowns; Lo, Christensen and Wu [6] with eleven unknowns; Kant [7] with six unknowns; Bhimaraddi and Stevens [8] with five unknowns; Reddy [9] with five unknowns; Soldatos [10] with three unknowns; Reddy [11] with eight unknowns; and Hanna and Leissa [12] with four unknowns.

It is important to note that Srinivas et al. [13] have carried out a three-dimensional, linear, small deformation theory of elasticity solution for thick plates.

A critical review of plate theories is given by Vasil'ev [14]. Whereas Liew et al. [15] surveyed plate theories particularly applied to thick plate vibration problems. A recent review paper is by Ghugal and Shimpi [16].

Even though more accurate higher order theories are available, half a century old FSDTs of Reissner and Mindlin are still in vogue, (e.g, Refs. [17–22]) because of their inherent simplicity. It needs to be noted that finite elements based on FSDT are prone to shear locking phenomenon.

The present paper proposes two new displacement based first-order shear deformation plate theories involving only two unknown functions. FSDTs of Reissner and of Mindlin involve use of three unknown functions.

It is to be noted that recently, a refined plate theory (RPT) was proposed [23] for shear-deformable isotropic plates. The theory, using only two unknown functions, gives rise to two governing equations, which are uncoupled for static problems.

Taking a cue from the just mentioned Ref. [23], two new displacement based FSDTs are proposed here. These two new FSDTs have distinct advantages over Reissner's and Mindlin's theories. Illustrative examples are given so as to demonstrate the efficacies of the theories.

2 Plate Under Consideration

The following is assumed about the plate under consideration

- The plate is of uniform thickness "h."
- The plate is of orthotropic material.
- Midsurface of the undeformed plate is taken as a reference plane.

  The right handed Cartesian coordinate system o-x-y-z would be utilized throughout the paper.

  The origin "o" of the coordinate system can be selected at a convenient location in this plane. Also the x-axis and y-axis are parallel to principal material axis.

  The xy-plane is assumed to coincide with midsurface of the undeformed plate.
- The midsurface of the plate is an area $\mathcal{R}(x,y)$ enclosed by a boundary curve $\Gamma(x,y)$, as shown in Fig. 1.
- The plate is subjected to a lateral load of $q(x,y,t)$ acting along the z-direction.
- Local directions n, s, and z' at a typical point on an edge are as shown. Where "n" is normal and "s" is tangent to a boundary at the point. n_x and n_y are the direction cosines of direction n with respect to the x-axis and y-axis, respectively. Direction z' is parallel to the z-axis of coordinate system.

The plate is made of orthotropic material. The notation used here for describing the material properties is the same as the one used by Reddy [9] as well as by Jones [24].

The orthotropic plate has following material properties: E_1, E_2 are elastic moduli, G_{12}, G_{23}, G_{31} are shear moduli, and μ_{12}, μ_{21} are Poisson's ratios (Here, in the notation, subscripts 1, 2, 3 pertain to x,y,z directions of Cartesian coordinate system, respectively).

In the case of isotropic plate, the material properties reduce to $E_1=E_2=E$, $G_{12}=G_{23}=G_{31}=G$, $\mu_{12}=\mu_{21}=\mu$.

[1]Corresponding author.

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received August 19, 2005; final manuscript received May 31, 2006. Review conducted by Sanjay Govindjee.

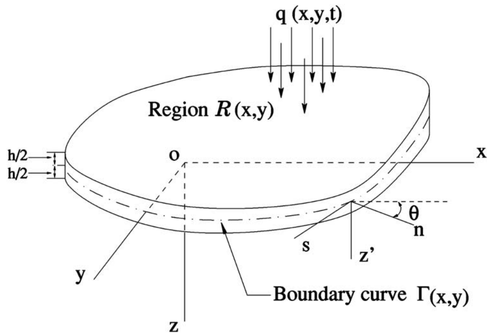


Fig. 1 Geometry of a plate

3 Assumptions of New First-Order Shear Deformation Theories

Assumptions of the proposed new FSDTs are of the same genre as those of Mindlin's theory and are as follows.

1. Displacements involved are small in comparison with the plate thickness and, therefore, strains involved are infinitesimal.

Comment on the assumption: As a result of this assumption, the direct strains ϵ_x, ϵ_y, ϵ_z, and shear strains γ_{xy}, γ_{yz}, γ_{zx} can be expressed in terms of displacements u along the x-direction, v along the y-direction, and w along the z-direction as follows:

$$\epsilon_x = \frac{\partial u}{\partial x}; \quad \epsilon_y = \frac{\partial v}{\partial y}; \quad \epsilon_z = \frac{\partial w}{\partial z}$$

$$\gamma_{xy} = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}; \quad \gamma_{yz} = \frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}; \quad \gamma_{zx} = \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \tag{1}$$

2. Transverse normal stress σ_z is negligible in comparison with inplane normal stresses σ_x and σ_y.

Comment on the assumption: As a result of this assumption, the direct stresses σ_x, σ_y and shear stresses τ_{xy}, τ_{yz}, τ_{zx} can be related to linear strains ϵ_x, ϵ_y and shear strains γ_{xy}, γ_{yz}, γ_{zx} by constitutive relations as follows:

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{Bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & 0 & 0 & 0 \\ Q_{12} & Q_{22} & 0 & 0 & 0 \\ 0 & 0 & Q_{66} & 0 & 0 \\ 0 & 0 & 0 & Q_{44} & 0 \\ 0 & 0 & 0 & 0 & Q_{55} \end{bmatrix} \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{Bmatrix} \tag{2}$$

where Q_{11}, Q_{12}, Q_{22}, Q_{44}, Q_{55}, Q_{66} are expressed in terms of properties of the orthotropic material as follows:

$$Q_{11} = \frac{E_1}{1-\mu_{12}\mu_{21}}; \quad Q_{12} = \frac{\mu_{12}E_2}{1-\mu_{12}\mu_{21}} = \frac{\mu_{21}E_1}{1-\mu_{12}\mu_{21}}$$

$$Q_{22} = \frac{E_2}{1-\mu_{12}\mu_{21}}$$

$$Q_{44} = G_{23}; \quad Q_{55} = G_{31}; \quad Q_{66} = G_{12} \tag{3}$$

3. A line, which is normal to the midsurface of plate before deformation, remains straight (i.e., may or may not be normal to the midsurface of the plate) after deformation.

Comment on the assumption: Transverse strains γ_{zx} and γ_{yz} are constant through the thickness, and, by implication, transverse stresses τ_{zx} and τ_{yz} will also be constant through the thickness.

This contradicts the well known fact that the transverse shear stresses vary more or less parabolically through the plate thickness; and therefore correction becomes necessary, one can write the constitutive relationship (2) in modified form as follows:

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{Bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & 0 & 0 & 0 \\ Q_{12} & Q_{22} & 0 & 0 & 0 \\ 0 & 0 & Q_{66} & 0 & 0 \\ 0 & 0 & 0 & kQ_{44} & 0 \\ 0 & 0 & 0 & 0 & kQ_{55} \end{bmatrix} \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{Bmatrix} \tag{4}$$

where k is a shear correction factor which is analogous to shear correction factor proposed by Mindlin [3].

As the constitutive relationship is modified as shown in Eq. (4), unlike FSDT proposed by Mindlin [3], moments M_x, M_y, M_{xy} and shear forces Q_x, Q_y can now be defined as follows:

$$\begin{Bmatrix} M_x \\ M_y \\ M_{xy} \\ Q_x \\ Q_y \end{Bmatrix} = \int_{z=-h/2}^{z=h/2} \begin{Bmatrix} \sigma_x z \\ \sigma_y z \\ \tau_{xy} z \\ \tau_{zx} \\ \tau_{yz} \end{Bmatrix} dz \tag{5}$$

Based on the assumptions, two new displacement based theories, namely "new first-order shear deformation plate theory—I" and "new first-order shear deformation plate theory—II" would now be presented.

4 New First-Order Shear Deformation Plate Theory—I

4.1 Expressions for Displacements, Moments, Shear Forces in the New First-Order Shear Deformation Plate Theory—I. Based on the assumptions made earlier in Sec. 3 and going by the previous experience (Ref. [23]), it is possible to write the displacements as follows:

$$u = -z\frac{\partial w_b}{\partial x} \tag{6}$$

$$v = -z\frac{\partial w_b}{\partial y} \tag{7}$$

$$w = w_b(x,y,t) + w_s(x,y,t) \tag{8}$$

Using expressions for displacements (6)–(8) in strain-displacement relations (1), the expressions for strains can be obtained as follows.:

$$\epsilon_x = -z\frac{\partial^2 w_b}{\partial x^2} \tag{9}$$

$$\epsilon_y = -z\frac{\partial^2 w_b}{\partial y^2} \tag{10}$$

$$\epsilon_z = 0 \tag{11}$$

$$\gamma_{xy} = -2z\frac{\partial^2 w_b}{\partial x \partial y} \tag{12}$$

$$\gamma_{yz} = \frac{\partial w_s}{\partial y} \tag{13}$$

$$\gamma_{zx} = \frac{\partial w_s}{\partial x} \tag{14}$$

Using strain expressions (9)–(14) in constitutive relations (4), stresses σ_x, σ_y, τ_{xy}, τ_{yz}, and τ_{zx} can be obtained. Using these stresses in definition (5) for moments and shear forces, expressions for moments M_x, M_y, M_{xy} and shear forces Q_x, Q_y can be obtained.

These expressions are

$$M_x = -\left[D_{11}\frac{\partial^2 w_b}{\partial x^2} + D_{12}\frac{\partial^2 w_b}{\partial y^2}\right] \tag{15}$$

$$M_y = -\left[D_{22}\frac{\partial^2 w_b}{\partial y^2} + D_{12}\frac{\partial^2 w_b}{\partial x^2}\right] \tag{16}$$

$$M_{xy} = -2\left[D_{66}\frac{\partial^2 w_b}{\partial x \partial y}\right] \tag{17}$$

$$Q_x = A_{55}\left[\frac{\partial w_s}{\partial x}\right] \tag{18}$$

$$Q_y = A_{44}\left[\frac{\partial w_s}{\partial y}\right] \tag{19}$$

where

$$D_{11} = \frac{Q_{11}h^3}{12}; \quad D_{22} = \frac{Q_{22}h^3}{12}; \quad D_{12} = \frac{Q_{12}h^3}{12}; \quad D_{66} = \frac{Q_{66}h^3}{12}$$

$$A_{44} = kQ_{44}h; \quad A_{55} = kQ_{55}h \tag{20}$$

It may be noted that expressions for moments M_x, M_y and M_{xy} contain only w_b as an unknown function. Also, The expressions for shear forces Q_x and Q_y contain only w_s as an unknown function.

4.2 Expressions for Kinetic and Total Potential Energies in Respect of the New First-Order Shear Deformation Plate Theory—I.

It should be noted that displacement w, given by Eq. (8), is not a function of z. As a result of this, normal strain ϵ_z comes out to be zero. Therefore, the expressions for kinetic energy T and total potential energy Π for the plate can be written as

$$T = \int\int_{\mathcal{R}}\int_{z=-h/2}^{z=h/2} \frac{1}{2}\rho\left[\left(\frac{\partial u}{\partial t}\right)^2 + \left(\frac{\partial v}{\partial t}\right)^2 + \left(\frac{\partial w}{\partial t}\right)^2\right]dzdxdy \tag{21}$$

$$\Pi = \int\int_{\mathcal{R}}\int_{z=-h/2}^{z=h/2} \frac{1}{2}[\sigma_x\epsilon_x + \sigma_y\epsilon_y + \tau_{xy}\gamma_{xy} + \tau_{yz}\gamma_{yz} + \tau_{zx}\gamma_{zx}]dzdxdy - \int\int_{\mathcal{R}} qwdxdy \tag{22}$$

Using expressions (6)–(14) and (4) in Eqs. (21) and (22), expressions for kinetic energy T and total potential energy Π can be written as

$$T = \frac{\rho h^3}{24}\int\int_{\mathcal{R}}\left\{\left[\frac{\partial}{\partial t}\left(\frac{\partial w_b}{\partial x}\right)\right]^2 + \left[\frac{\partial}{\partial t}\left(\frac{\partial w_b}{\partial y}\right)\right]^2\right\}dxdy + \frac{\rho h}{2}\int\int_{\mathcal{R}}\left\{\frac{\partial w_b}{\partial t} + \frac{\partial w_s}{\partial t}\right\}^2 dxdy \tag{23}$$

$$\Pi = \frac{1}{2}\int\int_{\mathcal{R}}\left[D_{11}\left(\frac{\partial^2 w_b}{\partial x^2}\right)^2 + D_{22}\left(\frac{\partial^2 w_b}{\partial y^2}\right)^2 + 2D_{12}\frac{\partial^2 w_b}{\partial x^2}\frac{\partial^2 w_b}{\partial y^2} + 4D_{66}\left(\frac{\partial^2 w_b}{\partial x\partial y}\right)^2 + A_{44}\left(\frac{\partial w_s}{\partial y}\right)^2 + A_{55}\left(\frac{\partial w_s}{\partial x}\right)^2\right]dxdy - \int\int_{\mathcal{R}} q[w_b + w_s]dxdy \tag{24}$$

4.3 Obtaining Governing Equations and Boundary Conditions of the New First-Order Shear Deformation Plate Theory—I Using Hamilton's Principle.

Governing differential equations and boundary conditions can be obtained using well known Hamilton's principle

$$\int_{t=t_1}^{t=t_2} \delta(T - \Pi)dt = 0 \tag{25}$$

where δ indicates a variation with respect to x and y only; t_1, t_2 are values of time variable t at the start and at the end of time interval (in the context of Hamilton's Principle), respectively.

Using expressions (23) and (24) in the preceding Eq. (25) and integrating the equation by parts, taking into account the independent variations of w_b and w_s, yields the governing differential equations and boundary conditions in respect of the new first-order shear deformation theory-I (NFSDT-I), which are as follows.

4.3.1 *Governing Equations of NFSDT-I.*

$$D_{11}\frac{\partial^4 w_b}{\partial x^4} + 2(D_{12} + 2D_{66})\frac{\partial^4 w_b}{\partial x^2\partial y^2} + D_{22}\frac{\partial^4 w_b}{\partial y^4} - \frac{\rho h^3}{12}\frac{\partial^2}{\partial t^2}(\nabla^2 w_b) + \rho h\left(\frac{\partial^2 w_b}{\partial t^2} + \frac{\partial^2 w_s}{\partial t^2}\right) = q \tag{26}$$

$$-\left(A_{55}\frac{\partial^2 w_s}{\partial x^2} + A_{44}\frac{\partial^2 w_s}{\partial y^2}\right) + \rho h\left(\frac{\partial^2 w_b}{\partial t^2} + \frac{\partial^2 w_s}{\partial t^2}\right) = q \tag{27}$$

Alternate form Governing equations can be written in alternate form as follows:

$$-\left(\frac{\partial^2 M_x}{\partial x^2} + 2\frac{\partial^2 M_{xy}}{\partial x\partial y} + \frac{\partial^2 M_y}{\partial y^2}\right) - \frac{\rho h^3}{12}\frac{\partial^2}{\partial t^2}(\nabla^2 w_b) + \rho h\left(\frac{\partial^2 w_b}{\partial t^2} + \frac{\partial^2 w_s}{\partial t^2}\right) = q \tag{28}$$

$$-\left(\frac{\partial Q_x}{\partial x} + \frac{\partial Q_y}{\partial y}\right) + \rho h\left(\frac{\partial^2 w_b}{\partial t^2} + \frac{\partial^2 w_s}{\partial t^2}\right) = q \tag{29}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{30}$$

4.3.2 *Boundary Conditions of NFSDT-I.*

1. The conditions involving w_b (i.e., bending component of lateral displacement)

$$V_n + \frac{\rho h^3}{12}\left[\frac{\partial^2}{\partial t^2}\left(\frac{\partial w_b}{\partial n}\right)\right] = 0 \text{ or } w_b \text{ is specified} \tag{31}$$

$$M_n = 0 \text{ or } \frac{\partial w_b}{\partial n} \text{ is specified} \tag{32}$$

where

$$M_n = \int_{-h/2}^{h/2}\sigma_n z dz = n_x^2 M_x + 2n_x n_y M_{xy} + n_y^2 M_y$$

$$M_{ns} = \int_{-h/2}^{h/2}\tau_{ns} z dz = n_x n_y (M_y - M_x) + (n_x^2 - n_y^2)M_{xy}$$

$$V_n = \frac{\partial M_n}{\partial n} + 2\frac{\partial M_{ns}}{\partial s}$$

$$\frac{\partial w_b}{\partial n} = \frac{\partial w_b}{\partial x}n_x + \frac{\partial w_b}{\partial y}n_y$$

2. The condition involving w_s (i.e., shear component of lateral displacement)

$$Q_n = 0 \text{ or } w_s \text{ is specified} \tag{33}$$

where

$$Q_n = \int_{-h/2}^{h/2} \tau_{z'n} dz = Q_x n_x + Q_y n_y$$

(It is to be noted that normal stress σ_n and shear stresses τ_{ns}, $\tau_{z'n}$ are with respect to n, s, z' coordinate system.)

If the integral is evaluated for a part of the curve $\Gamma(x,y)$ say between point 1 and point 2, one gets the term $[M_{ns}\delta w_b]_1^2$ (where 1 is a starting point and 2 is a end point of smooth curve), which is zero for a closed smooth curve, as the starting point 1 and end point 2 of the curve coincides.

However, in case of polygonal plates, the term indicates concentrated forces at corners. If a particular case of rectangular plate is considered, the magnitude of corner force will be "$2M_{ns}$."

This reasoning is exactly on the same lines as given in Ref. [25], in the context of classical plate theory.

For illustration, commonly occurring boundary conditions are stated below.

4.3.3 Commonly Occurring Boundary Conditions using NFSDT-I.

Simply-supported plate

$$w_b = 0$$

$$M_n = 0$$

$$w_s = 0$$

Clamped plate

$$w_b = 0$$

$$\frac{\partial w_b}{\partial n} = 0$$

$$w_s = 0$$

Free plate

$$V_n + \frac{\rho h^3}{12}\left[\frac{\partial^2}{\partial t^2}\left(\frac{\partial w_b}{\partial n}\right)\right] = 0$$

$$M_n = 0$$

$$Q_n = 0$$

Comments on governing equations and boundary conditions of the new first-order shear deformation plate theory—I are given later.

5 New First-Order Shear Deformation Plate Theory—II

5.1 Expressions for Displacements, Moments, Shear Forces in the New First-Order Shear Deformation Plate Theory—II. Based on the assumptions made earlier in the Sec. 3, it is possible to write the displacements, which are different from those given by expressions (6)–(8), as follows:

$$u = -z\frac{\partial \Phi}{\partial x} \tag{34}$$

$$v = -z\frac{\partial \Phi}{\partial y} \tag{35}$$

$$w = w(x,y,t) \tag{36}$$

where Φ is function of coordinates x,y and time t.

Using expressions for displacements (34)–(36) in strain-displacement relations (1), the expressions for strains can be obtained, further using constitutive relations (4), stresses σ_x, σ_y, τ_{xy}, τ_{yz}, and τ_{zx} can be obtained.

Using these stresses in definition (5) for moments and shear forces, expressions for moments M_x, M_y, M_{xy} and shear forces Q_x, Q_y can be obtained. These expressions are

$$M_x = -\left[D_{11}\frac{\partial^2\Phi}{\partial x^2} + D_{12}\frac{\partial^2\Phi}{\partial y^2}\right] \tag{37}$$

$$M_y = -\left[D_{22}\frac{\partial^2\Phi}{\partial y^2} + D_{12}\frac{\partial^2\Phi}{\partial x^2}\right] \tag{38}$$

$$M_{xy} = -2\left[D_{66}\frac{\partial^2\Phi}{\partial x\partial y}\right] \tag{39}$$

$$Q_x = A_{55}\left[\frac{\partial w}{\partial x} - \frac{\partial \Phi}{\partial x}\right] \tag{40}$$

$$Q_y = A_{44}\left[\frac{\partial w}{\partial y} - \frac{\partial \Phi}{\partial y}\right] \tag{41}$$

It may be noted that expressions for moments M_x, M_y, and M_{xy} contain only Φ as an unknown function. Terms D_{11}, D_{22}, D_{66}, A_{44}, and A_{55} are given in Eq. (20).

5.2 Expressions for Kinetic and Total Potential Energies in Respect of the New First-Order Shear Deformation Plate Theory—II. Using expressions (34)–(36), (1), and (4) in Eqs. (21) and (22), expressions for kinetic energy and strain energy can be written as

$$T = \frac{\rho h^3}{24}\int\int_{\mathcal{R}}\left\{\left[\frac{\partial}{\partial t}\left(\frac{\partial\Phi}{\partial x}\right)\right]^2 + \left[\frac{\partial}{\partial t}\left(\frac{\partial\Phi}{\partial y}\right)\right]^2\right\}dxdy + \frac{\rho h}{2}\int\int_{\mathcal{R}}\left\{\frac{\partial w}{\partial t}\right\}^2 dxdy \tag{42}$$

$$\Pi = \frac{1}{2}\int\int_{\mathcal{R}}\left[D_{11}\left(\frac{\partial^2\Phi}{\partial x^2}\right)^2 + D_{22}\left(\frac{\partial^2\Phi}{\partial y^2}\right)^2 + 2D_{12}\frac{\partial^2\Phi}{\partial x^2}\frac{\partial^2\Phi}{\partial y^2} + 4D_{66}\left(\frac{\partial^2\Phi}{\partial x\partial y}\right)^2\right]dxdy + \frac{1}{2}\int\int_{\mathcal{R}}\left[A_{44}\left(\frac{\partial w}{\partial y} - \frac{\partial\Phi}{\partial y}\right)^2 + A_{55}\left(\frac{\partial w}{\partial x} - \frac{\partial\Phi}{\partial x}\right)^2\right]dxdy - \int\int_{\mathcal{R}} qw dxdy \tag{43}$$

5.3 Obtaining Governing Equations and Boundary Conditions of the New First-Order Shear Deformation Plate Theory—II Using Hamilton's Principle. Applying the Hamilton's principle as stated in Eq. (25), and using expressions for kinetic energy T and total potential energy Π (i.e., expressions (42) and (43)), the governing differential equations and boundary conditions in respect of the new first-order shear deformation plate theory—II (NFSDT-II) can be obtained for the plate, which are as follows.

5.3.1 Governing Equations of NFSDT-II.

$$D_{11}\frac{\partial^4\Phi}{\partial x^4} + 2(D_{12}+2D_{66})\frac{\partial^4\Phi}{\partial x^2\partial y^2} + D_{22}\frac{\partial^4\Phi}{\partial y^4} - \left(A_{55}\frac{\partial^2\Phi}{\partial x^2} + A_{44}\frac{\partial^2\Phi}{\partial y^2}\right) - \frac{\rho h^3}{12}\frac{\partial^2}{\partial t^2}(\nabla^2\Phi) + \left(A_{55}\frac{\partial^2 w}{\partial x^2} + A_{44}\frac{\partial^2 w}{\partial y^2}\right) = 0 \tag{44}$$

$$-\left(A_{55}\frac{\partial^2 w}{\partial x^2}+A_{44}\frac{\partial^2 w}{\partial y^2}\right)+\rho h\left(\frac{\partial^2 w}{\partial t^2}\right)+\left(A_{55}\frac{\partial^2 \Phi}{\partial x^2}+A_{44}\frac{\partial^2 \Phi}{\partial y^2}\right)=q \tag{45}$$

Alternate form Governing equations can be written in alternate form as follows

$$-\left(\frac{\partial^2 M_x}{\partial x^2}+2\frac{\partial^2 M_{xy}}{\partial x\partial y}+\frac{\partial^2 M_y}{\partial y^2}\right)+\left(\frac{\partial Q_x}{\partial x}+\frac{\partial Q_y}{\partial y}\right)-\frac{\rho h^3}{12}\frac{\partial^2}{\partial t^2}(\nabla^2\Phi)=0 \tag{46}$$

$$-\left(\frac{\partial Q_x}{\partial x}+\frac{\partial Q_y}{\partial y}\right)+\rho h\left(\frac{\partial^2 w}{\partial t^2}\right)=q \tag{47}$$

5.3.2 Boundary Conditions of NFSDT-II.

$$V_n+Q_n+\frac{\rho h^3}{12}\left[\frac{\partial^2}{\partial t^2}\left(\frac{\partial \Phi}{\partial n}\right)\right]=0 \text{ or } \Phi \text{ is specified} \tag{48}$$

$$M_n=0 \text{ or } \frac{\partial \Phi}{\partial n} \text{ is specified} \tag{49}$$

$$Q_n=0 \text{ or } w \text{ is specified} \tag{50}$$

where

$$M_n=\int_{-h/2}^{h/2}\sigma_n z dz=n_x^2 M_x+2n_x n_y M_{xy}+n_y^2 M_y$$

$$M_{ns}=\int_{-h/2}^{h/2}\tau_{ns} z dz=n_x n_y(M_y-M_x)+(n_x^2-n_y^2)M_{xy}$$

$$V_n=\frac{\partial M_n}{\partial n}+2\frac{\partial M_{ns}}{\partial s}$$

$$\frac{\partial \Phi}{\partial n}=\frac{\partial \Phi}{\partial x}n_x+\frac{\partial \Phi}{\partial y}n_y$$

$$Q_n=\int_{-h/2}^{h/2}\tau_{z'n} dz=Q_x n_x+Q_y n_y$$

(It is to be noted that normal stress σ_n and shear stresses τ_{ns}, $\tau_{z'n}$ are with respect to n, s, z' co-ordinate system.) While deriving the boundary conditions, logic similar to that adopted for NFSDT-I was used in case of NFSDT-II also. Hence, for a rectangular plate there would be corner reactions of magnitude "$2M_{ns}$". For illustration, commonly occurring boundary conditions are stated below.

5.3.3 Commonly Occurring Boundary Conditions using NFSDT-II.

Simply-supported plate

$$\Phi=0$$

$$M_n=0$$

$$w=0$$

Clampled plate

$$\Phi=0$$

$$\frac{\partial \Phi}{\partial n}=0$$

$$w=0$$

Free plate

$$V_n+Q_n+\frac{\rho h^3}{12}\left[\frac{\partial^2}{\partial t^2}\left(\frac{\partial \Phi}{\partial n}\right)\right]=0$$

$$M_n=0$$

$$Q_n=0$$

Comments on governing equations and boundary conditions of NFSDT-I and NFSDT-II are given below.

6 Comments on NFSDT-I and NFSDT-II

1. With respect to governing equations, the following is to be noted:
   - (a) In each theory, one governing equation is a fourth order one and the second is a second order one.
   - (b) The governing equations of both theories, involve only two unknown functions (w_b, w_s in the case of NFSDT-I and w, Φ in the case of NFSDT-II). Even first-order shear deformation theories of Reissner [2] and Mindlin [3], which are considered to be simple ones, involve three unknown functions.
   - (c) (i) In the case of a static problem, the governing equations are uncoupled in NFSDT-I.
     (ii) In the case of a dynamic problem,
       - A. governing equations of NFSDT-I, are only inertially coupled, and there is no elastic coupling at all, whereas
       - B. governing equations of NFSDT-II, are only elastically coupled, and there is no inertial coupling at all.

2. (a) With respect to boundary conditions for NFSDT-I, the following is to be noted:

   There are three conditions along the boundary $\Gamma(x,y)$. Out of these, two conditions (Eqs. (31) and (32)) are stated in terms of w_b and its derivatives only.
   The remaining one condition (Eq. (33)) is stated in terms of w_s and its derivative only.

   (b) With respect to boundary conditions for NFSDT-II, the following is to be noted:

   There are three conditions along the boundary $\Gamma(x,y)$ (i.e., Eqs. (48)–(50)), which are stated in terms of Φ and w and its derivative only.

3. Some entities of NFSDT-I and NFSDT-II have strong similarity with those of CPT:
   - (a) The following entities of NFSDT-I and NFSDT-II are identical, save for the appearance of w_b or Φ instead of w, to the corresponding entities of the CPT:
     - (i) Expressions for inplane strains ϵ_x, ϵ_y, γ_{xy}, stresses σ_x, σ_y, τ_{xy} and expressions for moments M_x, M_y, M_{xy}
     - (ii) Conditions on boundary curve $\Gamma(x,y)$ involving w_b for NFSDT-I (i.e., conditions (31) and (32)).
     - (iii) Conditions on boundary curve $\Gamma(x,y)$ involving Φ for NFSDT-II (i.e., conditions (49)).

     (The bending component w_b of lateral displacement figures in the just mentioned equations of NFSDT-I, and similarly Φ figures in the just mentioned equations of NFSDT-II, whereas lateral displacement w figures in the corresponding equations of the CPT.)
   - (b) The governing Eq. (26), of NFSDT-I is very similar to the governing equation of CPT. (If in Eq. (26) the

term $\partial^2 w_s / \partial t^2$ is ignored, and if w_b is replaced by w, then the resulting equation is identical to the governing equation of CPT.)

4. CPT comes out as a special case of NFSDT-I whereas NFSDT-I comes out as a special case of RPT (see Ref. [23]) as the governing equations and boundary conditions of NFSDT-I are identical to those of RPT-Variant I of Ref. [23].

5. (a) It should be noted that, the common feature with refined plate theory [23] and present NFSDT-I and NFSDT-II theories is that, they have same number of variables. However, the theories presented are quiet different right from the assumption stage. RPT is higher order plate theory, gives cubic shear stress distribution, whereas NFSDT-I and NFSDT-II are first-order shear deformation theories.
   (b) The assumptions of proposed displacement theories are on the same lines as those of FSDT of Mindlin, so it will also require shear correction factor which is taken as $5/6$. There is good discussion on shear correction factor in Mindlin's paper [3]. It is to be noted that the same number appears in Reissner [1] as well as RPT-Variant I (See Ref. [23]) theories.

6. Finite element based on FSDT of Mindlin is prone to shear locking for thin plates. Literature shows, even today, quiet a lot efforts put in by researchers (e.g, Ref. [26–30]) on shear locking free finite elements.

   Whereas in the case of present theories, inspection of the governing equations (Eqs. (26) and (27) for NFSDT-I and Eqs. (44) and (45) for NFSDT-II) reveals that, finite elements based on these theories would require C^0 continuity for shear component w_s of displacement, whereas C^1 continuity would require for bending component w_b of displacement, therefore, shear locking will not occur.

   Simultaneous use of C^0 and C^1 continuity requirements was also utilized by Reddy [31]. Use of only two variables, in proposed theories, also especially only inertially decoupled governing equations in the case of NFSDT-I, will make these finite elements really interesting to study.

7. If NFSDT-I is applied to laminated plate, it will give rise to elastically decoupled equations, which is not possible by any other known theory.

Now, well known examples, available in the literature [9,13,22,23,32] would be taken up to demonstrate the effectiveness of presented theories.

7 Illustrative Examples for Static Analysis of a Simply Supported Rectangular Plate Carrying Lateral Load

Consider a rectangular plate (of length a, width b, and thickness h) of orthotropic material. The plate occupies (in o-x-y-z right-handed Cartesian coordinate system) a region

$$0 \leqslant x \leqslant a; \quad 0 \leqslant y \leqslant b; \quad -h/2 \leqslant z \leqslant h/2 \tag{51}$$

The plate has simply supported boundary conditions at all four edges $x=0$, $x=a$, $y=0$ and $y=b$. The plate is loaded on surface $z=-h/2$, by a lateral load $q(x,y)$ acting in the z-direction.

Two specific cases of loading $q(x,y)$ would be considered. In one case, the loading would be a uniformly distributed load; and, in the second case, the loading would be a sinusoidal loading.

7.1 Application of NFSDT-I (Static Analysis)

7.1.1 Governing Equations for Illustrative Examples (Static Analysis). The governing equations for static analysis of plate can be obtained from Eqs. (26) and (27) by setting the kinetic energy terms to zero, as follows:

$$D_{11}\frac{\partial^4 w_b}{\partial x^4} + 2(D_{12}+2D_{66})\frac{\partial^4 w_b}{\partial x^2 \partial y^2} + D_{22}\frac{\partial^4 w_b}{\partial y^4} = q(x,y) \tag{52}$$

$$-\left[A_{55}\frac{\partial^2 w_s}{\partial x^2} + A_{44}\frac{\partial^2 w_s}{\partial y^2}\right] = q(x,y) \tag{53}$$

7.1.2 Boundary Conditions for Illustrative Examples (Static Analysis). The boundary conditions of the plate are given as follows:

1. At corners $(x=0, y=0)$, $(x=0, y=b)$, $(x=a, y=0)$, and $(x=a, y=b)$ the following condition holds:

$$w_b = 0 \tag{54}$$

2. On edges $x=0$ and $x=a$, the following conditions hold:

$$w_b = 0 \tag{55}$$

$$-\left[D_{11}\frac{\partial^2 w_b}{\partial x^2} + D_{12}\frac{\partial^2 w_b}{\partial y^2}\right] = 0 \tag{56}$$

$$w_s = 0 \tag{57}$$

3. On edges $y=0$ and $y=b$, the following conditions hold:

$$w_b = 0 \tag{58}$$

$$-\left[D_{22}\frac{\partial^2 w_b}{\partial y^2} + D_{12}\frac{\partial^2 w_b}{\partial x^2}\right] = 0 \tag{59}$$

$$w_s = 0 \tag{60}$$

7.1.3 Solution of Illustrative Examples (Static Analysis). Lateral load applied on the plate, in general, can be represented using a double Fourier series as follows:

$$q(x,y) = \sum_{m=1,2,3,\ldots}^{\infty}\sum_{n=1,2,3,\ldots}^{\infty} q_{mn}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{61}$$

where q_{mn} are Fourier constants, which depend on loading.

The following displacement functions w_b and w_s satisfy the boundary conditions (54)–(60)

$$w_b = \sum_{m=1,2,3,\ldots}^{\infty}\sum_{n=1,2,3,\ldots}^{\infty} W_{b_{mn}}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{62}$$

$$w_s = \sum_{m=1,2,3,\ldots}^{\infty}\sum_{n=1,2,3,\ldots}^{\infty} W_{s_{mn}}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{63}$$

where $W_{b_{mn}}$, $W_{s_{mn}}$ are constant coefficients associated with w_b and w_s, respectively.

Using expressions (62), (63), and (61) in the governing Eqs. (52) and (53), one obtains two completely uncoupled equations which can be written in matrix form, as follows:

$$\begin{bmatrix} K_{11_{mn}} & 0 \\ 0 & K_{22_{mn}} \end{bmatrix}\begin{Bmatrix} W_{b_{mn}} \\ W_{s_{mn}} \end{Bmatrix} = \begin{Bmatrix} q_{mn} \\ q_{mn} \end{Bmatrix} \tag{64}$$

where

$$K_{11_{mn}} = D_{11}\left(\frac{m\pi}{a}\right)^4 + 2(D_{12}+2D_{66})\left(\frac{m\pi}{a}\right)^2\left(\frac{n\pi}{b}\right)^2 + D_{22}\left(\frac{n\pi}{b}\right)^4$$

Table 1 Comparison of various nondimensional parameters of simply supported isotropic square plate ($b/a=1.0, h/a=0.1$) under sinusoidal transverse load: $\hat{w}=wE/hq_0$, $\bar{\sigma}_x=\sigma_x/q_0$, $\bar{\tau}_{zx}=\tau_{zx}/q_0$, ($E_2=E_1=E, G_{12}=G_{13}=G_{23}=G=[E/2(1+\mu)], \mu_{12}=\mu_{21}=\mu=0.3$)

Parameter	EXACT[a]	RPT[a]	CPT[a]	Present[b]
Nondimensional displacement $\hat{w}$ at $x=a/2$, $y=b/2$	294.2375	296.0568	280.2613	296.0674
Nondimensional stress $\bar{\sigma}_x$ at $x=a/2$, $y=b/2$, $z=h/2$	20.04426	19.94322	19.75763	19.75763
Nondimensional stress $\bar{\tau}_{zx}$ at $x=0$, $y=b/2$, $z=0$	Not quoted	2.385722	2.387324	2.387324

[a]Taken from Ref. [23].
[b]Results using NFSDT-I and NFSDT-II.

$$K_{22_{mn}} = A_{55}\left(\frac{m\pi}{a}\right)^2 + A_{44}\left(\frac{n\pi}{b}\right)^2$$

It can be seen from Eq. (64) that $W_{b_{mn}}=q_{mn}/K_{11_{mn}}$ and $W_{s_{mn}}=q_{mn}/K_{22_{mn}}$. Using this information in Eqs. (62) and (63) one gets displacements as

$$w_b = \sum_{m=1,2,3,\ldots}^{\infty} \sum_{n=1,2,3,\ldots}^{\infty} \frac{q_{mn}}{K_{11_{mn}}} \sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \quad (65)$$

$$w_s = \sum_{m=1,2,3,\ldots}^{\infty} \sum_{n=1,2,3,\ldots}^{\infty} \frac{q_{mn}}{K_{22_{mn}}} \sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \quad (66)$$

Here it is observed from the Eq. (64) that bending and shearing components are uncoupled.

- **Sinusoidal Lateral Load**
  For a sinusoidal load $q_0 \sin(m\pi x/a)\sin(n\pi y/b)$, the Fourier constants q_{mn} are given by

$$q_{mn} = q_0 \quad \text{for } m=1 \quad \text{and } n=1$$

$$=0 \quad \text{for } m>1 \quad \text{and } n>1 \quad (67)$$

  Results for plate with sinusoidal loading are tabulated in Table 1.
- **Uniformly Distributed Lateral Load**

For a uniformly distributed load q_0, the Fourier constants q_{mn} are given by

$$q_{mn} = \frac{16q_0}{\pi^2 mn} \quad \text{for } m=1,3,5,\ldots, \quad \text{and } n=1,3,5,\ldots$$

$$=0 \quad \text{for } m=2,4,6,\ldots, \quad \text{and } n=2,4,6,\ldots \quad (68)$$

Results for plate with uniformly distributed load are tabulated in Tables 2 and 3.

Results in terms of nondimensionalized deflections $\hat{w}$, $\tilde{w}$, $\bar{w}$ and stresses $\bar{\sigma}_x$, $\bar{\sigma}_y$, $\bar{\tau}_{zx}$ are tabulated in Tables 1–3.

Definitions used for nondimensionalized deflections and stresses are as follows:

$\hat{w}$	=	$wE/(hq_0)$	for isotropic plate with sinusoidal loading;
$\tilde{w}$	=	$wG/(hq_0)$	for isotropic plate with uniformly distributed loading;
$\bar{w}$	=	$wQ_{11}/(hq_0)$	for orthotropic plate with uniformly distributed loading;
$\bar{\sigma}_x$	=	σ_x/q_0	for isotropic and orthotropic plate;
$\bar{\sigma}_y$	=	σ_y/q_0	for isotropic and orthotropic plate;
$\bar{\tau}_{zx}$	=	τ_{zx}/q_0	for orthotropic plate with uniformly distributed loading.

Now, NFSDT-II will be applied to the just discussed static problem.

7.2 Application of NFSDT-II (Static Analysis)

7.2.1 Governing Equations for Illustrative Examples (Static Analysis). The governing equations for static analysis of plate can be obtained from Eqs. (44) and (45) by setting the kinetic energy terms to zero, and these equations are as follows:

$$D_{11}\frac{\partial^4\Phi}{\partial x^4} + 2(D_{12}+2D_{66})\frac{\partial^4\Phi}{\partial x^2\partial y^2} + D_{22}\frac{\partial^4\Phi}{\partial y^4} - \left(A_{55}\frac{\partial^2\Phi}{\partial x^2} + A_{44}\frac{\partial^2\Phi}{\partial y^2}\right) + \left(A_{55}\frac{\partial^2 w}{\partial x^2} + A_{44}\frac{\partial^2 w}{\partial y^2}\right) = 0 \quad (69)$$

$$-\left[A_{55}\frac{\partial^2 w}{\partial x^2} + A_{44}\frac{\partial^2 w}{\partial y^2}\right] + \left[A_{55}\frac{\partial^2\Phi}{\partial x^2} + A_{44}\frac{\partial^2\Phi}{\partial y^2}\right] = q(x,y) \quad (70)$$

Table 2 Comparison of various nondimensional parameters of simply supported isotropic square plate ($b/a=1$) under uniformly distributed transverse load: $\tilde{w}=wG/hq_0$, $\bar{\sigma}_x=\sigma_x/q_0$, $\bar{\sigma}_y=\sigma_y/q_0$, $\bar{\tau}_{zx}=\tau_{zx}/q_0$, ($E_2=E_1=E, G_{12}=G_{13}=G_{23}=G=[E/2(1+\mu)], \mu_{12}=\mu_{21}=\mu=0.3$)

h/a	Results by various theories			
	Nondimensional displacement $\tilde{w}$ at $x=a/2$, $y=b/2$			
	EXACT[a]	Reissner[a]	CPT[a]	Present[b]
0.05	2761.3	2760.00	2729.9	2765.27
0.1	178.45	178.13	170.62	179.46
0.14	48.40	48.247	44.414	48.92
	Nondimensional stress $\bar{\sigma}_x$ at $x=a/2$, $y=b/2$, $z=h/2$			
	EXACT[a]	Reissner[a]	CPT[a]	Present[b]
0.05	115.21	115.02	114.93	114.93
0.1	29.012	28.821	28.733	28.733
0.14	14.939	14.749	14.660	14.660
	Nondimensional stress $\bar{\sigma}_y$ at $x=a/2$, $y=b/2$, $z=h/2$			
	EXACT[a]	Reissner[a]	CPT[a]	Present[b]
0.05	115.21	115.02	114.93	114.93
0.1	29.012	28.821	28.733	28.733
0.14	14.939	14.749	14.660	14.660
	Nondimensional stress $\bar{\tau}_{zx}$ at $x=0$, $y=b/2$, $z=0$			
	EXACT[a]	Reissner[a]	CPT[a]	Present[b]
0.05	9.8339	9.8656	9.8656	9.8656
0.1	4.8816	5.0232	5.0232	5.0232
0.14	3.4345	3.6021	3.6021	3.6021

[a]Taken from Ref. [33].
[b]Results using NFSDT-I and NFSDT-II.

Table 3 Comparison of various nondimensional parameters of simply supported orthotropic square plate ($b/a=1$) under uniformly distributed transverse load: $\bar{w}=wQ_{11}/hq_0$, $\bar{\sigma}_x=\sigma_x/q_0$, $\bar{\sigma}_y=\sigma_y/q_0$, $\bar{\tau}_{zx}=\tau_{zx}/q_0$, ($E_2/E_1=0.52500, G_{12}/E_1=0.29281, G_{13}/E_1=0.17809, G_{23}/E_1=0.29713, \mu_{12}=0.44046, \mu_{21}=0.23124$)

h/a	Results by various theories			
	Nondimensional displacement $\bar{w}$ at $x=a/2$, $y=b/2$			
	EXACT[a]	Reissner[a]	CPT[a]	Present[b]
0.05	10443	10442	10246	10413.4
0.1	688.57	688.37	640.39	681.75
0.14	191.07	191.02	166.7	187.77
	Nondimensional stress $\bar{\sigma}_x$ at $x=a/2$, $y=b/2$, $z=h/2$			
	EXACT[a]	Reissner[a]	CPT[a]	Present[b]
0.05	144.31	143.87	144.39	144.39
0.1	36.021	35.578	36.098	36.098
0.14	18.346	17.906	18.417	18.417
	Nondimensional stress $\bar{\sigma}_y$ at $x=a/2$, $y=b/2$, $z=h/2$			
	EXACT[a]	Reissner[a]	CPT[a]	Present[b]
0.05	87.08	86.921	86.487	86.487
0.1	22.21	22.048	21.622	21.622
0.14	11.615	11.453	11.031	11.031
	Nondimensional stress $\bar{\tau}_{zx}$ at $x=0$, $y=b/2$, $z=0$			
	EXACT[a]	Reissner[a]	CPT[a]	Present[b]
0.05	10.873	10.864	10.972	10.972
0.1	5.3411	5.4267	5.5642	5.5642
0.14	3.7313	3.8741	3.986	3.986

[a]Taken from Ref. [32].
[b]Results using NFSDT-I and NFSDT-II.

7.2.2 Boundary Conditions for Illustrative Examples (Static Analysis). The boundary conditions of the plate are given as follows:

1. At corners $(x=0, y=0)$, $(x=0, y=b)$, $(x=a, y=0)$, and $(x=a, y=b)$ the following condition holds:

$$\Phi = 0 \tag{71}$$

2. On edges $x=0$ and $x=a$, the following conditions hold:

$$\Phi = 0 \tag{72}$$

$$-\left[D_{11}\frac{\partial^2\Phi}{\partial x^2} + D_{12}\frac{\partial^2\Phi}{\partial y^2}\right] = 0 \tag{73}$$

$$w = 0 \tag{74}$$

3. On edges $y=0$ and $y=b$, the following conditions hold:

$$\Phi = 0 \tag{75}$$

$$-\left[D_{22}\frac{\partial^2\Phi}{\partial y^2} + D_{12}\frac{\partial^2\Phi}{\partial x^2}\right] = 0 \tag{76}$$

$$w = 0 \tag{77}$$

7.2.3 Solution of the Illustrative Examples (Static Analysis). Lateral load applied on the plate, in general, can be represented by Eq. (61).

The following displacement functions Φ and w satisfy the boundary conditions (71)–(77)

$$\Phi = \sum_{m=1,2,\ldots}^{\infty}\sum_{n=1,2,\ldots}^{\infty} \phi_{mn}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{78}$$

$$w = \sum_{m=1,2,\ldots}^{\infty}\sum_{m=1,2,\ldots}^{\infty} W_{mn}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{79}$$

where ϕ_{mn}, W_{mn} are constant coefficients associated with Φ and w, respectively.

Using expressions (78), (79), and (61) in the governing Eqs. (69) and (70), one obtains two equations which can be written in matrix form, as follows:

$$\begin{bmatrix} k_{11_{mn}} & k_{12_{mn}} \\ k_{12_{mn}} & k_{22_{mn}} \end{bmatrix}\begin{Bmatrix} \phi_{mn} \\ W_{mn} \end{Bmatrix} = \begin{Bmatrix} 0 \\ q_{mn} \end{Bmatrix} \tag{80}$$

where

$$k_{11_{mn}} = D_{11}\left(\frac{m\pi}{a}\right)^4 + 2(D_{12}+2D_{66})\left(\frac{m\pi}{a}\right)^2\left(\frac{n\pi}{b}\right)^2 + D_{22}\left(\frac{n\pi}{b}\right)^4 + A_{55}\left(\frac{m\pi}{a}\right)^2 + A_{44}\left(\frac{n\pi}{b}\right)^2$$

$$k_{12_{mn}} = -\left[A_{55}\left(\frac{m\pi}{a}\right)^2 + A_{44}\left(\frac{n\pi}{b}\right)^2\right]$$

$$k_{22_{mn}} = A_{55}\left(\frac{m\pi}{a}\right)^2 + A_{44}\left(\frac{n\pi}{b}\right)^2$$

It can be seen from Eq. (80) that $\phi_{mn}=-q_{mn}k_{12}/(k_{11}k_{22}-k_{12}^2)$ and $W_{mn}=q_{mn}k_{11}/(k_{11}k_{22}-k_{12}^2)$. Using this information in Eqs. (78) and (79) one gets displacements as:

$$\Phi = \sum_{m=1,2,3,\ldots}^{\infty}\sum_{n=1,2,3,\ldots}^{\infty} \frac{-q_{mn}k_{12}}{(k_{11}k_{22}-k_{12}^2)}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{81}$$

$$w = \sum_{m=1,2,3,\ldots}^{\infty}\sum_{n=1,2,3,\ldots}^{\infty} \frac{q_{mn}k_{11}}{(k_{11}k_{22}-k_{12}^2)}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \tag{82}$$

Here it is observed from the Eq. (80) that Φ and w components are coupled unlike the case of NFSDT-I solution.

- **Sinusoidal Lateral Load**
  For a sinusoidal load $q_0\sin(m\pi x/a)\sin(n\pi y/b)$, the Fourier constants q_{mn} are given by Eq. (67).
  Results for plate with sinusoidal loading are tabulated in Table 1.
- **Uniformly Distributed Lateral Load**

For a uniformly distributed load q_0, the Fourier constants q_{mn} are given by Eq. (68).

Results for plate with uniformly distributed load are tabulated in Tables 2 and 3.

Results in terms of nondimensionalized deflections $\hat{w}$, $\tilde{w}$, $\bar{w}$ and stresses $\bar{\sigma}_x$, $\bar{\sigma}_y$, $\bar{\tau}_{zx}$ are tabulated in Tables 1–3.

Definitions used for nondimensionalized deflections and stresses are same as defined previously in Sec. 7.1.3.

8 Illustrative Examples for Free Vibration Analysis of a Simply Supported Rectangular Plate

An illustrative example would be taken up to demonstrate the effectiveness of present theories for vibration analysis.

Consider a plate (of length a, width b, and thickness h) of orthotropic material. The plate occupies (in o-x-y-z right-handed Cartesian coordinate system) a region defined by Eq. (51). The plate has simply supported boundary conditions at all four edges $x=0$, $x=a$, $y=0$, and $y=b$. The plate is in a state of free vibrations.

8.1 Application of NSFDT-I (Free Vibration Analysis)

8.1.1 Governing Equations for Illustrative Examples (Free Vibration Analysis). The governing equations for free vibration of plate can be obtained from Eqs. (26) and (27) by setting the external load (here, lateral load q) to zero

$$D_{11}\frac{\partial^4 w_b}{\partial x^4}+2(D_{12}+2D_{66})\frac{\partial^4 w_b}{\partial x^2\partial y^2}+D_{22}\frac{\partial^4 w_b}{\partial y^4}-\frac{\rho h^3}{12}\frac{\partial^2}{\partial t^2}(\nabla^2 w_b)+\rho h\left(\frac{\partial^2 w_b}{\partial t^2}+\frac{\partial^2 w_s}{\partial t^2}\right)=0 \quad (83)$$

$$-\left[A_{55}\frac{\partial^2 w_s}{\partial x^2}+A_{44}\frac{\partial^2 w_s}{\partial y^2}\right]+\rho h\left(\frac{\partial^2 w_b}{\partial t^2}+\frac{\partial^2 w_s}{\partial t^2}\right)=0 \quad (84)$$

8.1.2 Boundary Conditions for Illustrative Examples (Free Vibration Analysis). The boundary conditions are same as those given by conditions (54)–(60).

8.1.3 Solution of the Illustrative Examples (Free Vibration Analysis). The following displacement functions Φ and w satisfy the boundary conditions (54)–(60).

$$w_b=\sum_{m=1,2\ldots}^{\infty}\sum_{n=1,2\ldots}^{\infty}W_{b_{mn}}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)\sin(\omega_{mn}t) \quad (85)$$

$$w_s=\sum_{m=1,2\ldots}^{\infty}\sum_{n=1,2\ldots}^{\infty}W_{s_{mn}}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)\sin(\omega_{mn}t) \quad (86)$$

where $W_{b_{mn}}$, $W_{s_{mn}}$ are constants and ω_{mn} is the circular frequency of vibration associated with mth mode in the x-direction and nth mode in the y-direction.

Using expressions (85) and (86) in the governing Eqs. (83) and (84), one obtains two equations which are written in matrix form as follows:

$$\begin{bmatrix} K_{11_{mn}} & 0 \\ 0 & K_{22_{mn}}\end{bmatrix}\begin{Bmatrix} W_{b_{mn}} \\ W_{s_{mn}}\end{Bmatrix}-\omega_{mn}^2\begin{bmatrix} M_{11_{mn}} & M_{12_{mn}} \\ M_{12_{mn}} & M_{22_{mn}}\end{bmatrix}\begin{Bmatrix} W_{b_{mn}} \\ W_{s_{mn}}\end{Bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \quad (87)$$

where

$$M_{11_{mn}}=\left\{\frac{\rho h^3}{12}\left[\left(\frac{m\pi}{a}\right)^2+\left(\frac{n\pi}{b}\right)^2\right]+\rho h\right\}$$

$$M_{12_{mn}}=[\rho h]$$

$$M_{22_{mn}}=[\rho h]$$

Equation (87) is of a standard eigenvalue form, and solving it one gets free vibration frequencies.

Here, it is observed from the Eq. (87) that bending and shearing components are elastically uncoupled, but are inertially coupled.

Results are tabulated in Tables 4 and 5 and discussed later.

Now, NSFDT-II will be applied to the just discussed vibration problem.

8.2 Application of NSFDT-II (Free Vibration Analysis)

8.2.1 Governing Equations for Illustrative Examples (Free Vibration Analysis). The governing equations for free vibration of plate can be obtained from Eqs. (44) and (45) by setting the external load (here, lateral load q_0) to zero

$$D_{11}\frac{\partial^4\Phi}{\partial x^4}+2(D_{12}+2D_{66})\frac{\partial^4\Phi}{\partial x^2\partial y^2}+D_{22}\frac{\partial^4\Phi}{\partial y^4}-\left(A_{55}\frac{\partial^2\Phi}{\partial x^2}+A_{44}\frac{\partial^2\Phi}{\partial y^2}\right)-\frac{\rho h^3}{12}\frac{\partial^2}{\partial t^2}(\nabla^2\Phi)+\left(A_{55}\frac{\partial^2 w}{\partial x^2}+A_{44}\frac{\partial^2 w}{\partial y^2}\right)=0 \quad (88)$$

Table 4 Comparison of nondimensional natural frequencies $\hat{\omega}_{mn}$ of simply supported isotropic square plate: $\hat{\omega}_{mn}=\omega_{mn}h\sqrt{\rho/G}$; $h/a=0.1$, $b/a=1.0$; ($E_2=E_1=E$, $G_{12}=G_{13}=G_{23}=G=[E/2(1+\mu)]$, $\mu_{12}=\mu_{21}=\mu=0.3$)

		Nondimensional natural frequency $\hat{\omega}_{mn}$ by various theories			
m	n	EXACT[a]	Mindlin[a]	CPT[a]	Present[b]
1	1	0.0932	0.0930	0.0955	0.0930
1	2	0.2226	0.2219	0.2360	0.2219
2	2	0.3421	0.3406	0.3732	0.3406
1	3	0.4171	0.4149	0.4629	0.4149
2	3	0.5239	0.5206	0.5951	0.5206
1	4	—	0.6520	0.7668	0.6520
3	3	0.6889	0.6834	0.8090	0.6834
2	4	0.7511	0.7446	0.8926	0.7447
3	4	—	0.8896	1.0965	0.8897
1	5	0.9268	0.9174	1.1365	0.9174

[a]Taken from Ref. [13].
[b]Results using NFSDT-I and NFSDT-II.

$$-\left[A_{55}\frac{\partial^2 w}{\partial x^2}+A_{44}\frac{\partial^2 w}{\partial y^2}\right]+\rho h\left(\frac{\partial^2 w}{\partial t^2}\right)+\left[A_{55}\frac{\partial^2\Phi}{\partial x^2}+A_{44}\frac{\partial^2\Phi}{\partial y^2}\right]=0 \quad (89)$$

8.2.2 Boundary Conditions for Illustrative Examples (Free Vibration Analysis). The boundary conditions are same as those given by conditions (71)–(77).

8.2.3 Solution of the Illustrative Examples (Free Vibration Analysis). The following functions Φ and w satisfy the boundary conditions (71)–(77)

$$\Phi=\sum_{m=1,2}^{\infty}\sum_{m=1,2}^{\infty}\phi_{mn}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)\sin(\omega_{mn}t) \quad (90)$$

$$w=\sum_{m=1,2}^{\infty}\sum_{m=1,2}^{\infty}W_{mn}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)\sin(\omega_{mn}t) \quad (91)$$

where ϕ_{mn}, W_{mn} are constants and ω_{mn} is the circular frequency of vibration associated with mth mode in the x-direction and nth

Table 5 Comparison of nondimensional natural frequencies $\bar{\omega}_{mn}$ of simply supported orthotropic rectangular plate: $\bar{\omega}_{mn}=\omega_{mn}h\sqrt{\rho/Q_{11}}$; $h/a=0.1$, $b/a=1.0$; ($E_2/E_1=0.52500$, $G_{12}/E_1=0.29281$, $G_{13}/E_1=0.17809$, $G_{23}/E_1=0.29713$, $\mu_{12}=0.44046$, $\mu_{21}=0.23124$)

		Nondimensional natural frequency $\bar{\omega}_{mn}$ by various theories			
m	n	EXACT[a]	Mindlin[a]	CPT[a]	Present[b]
1	1	0.0474	0.0474	0.0497	0.0477
1	2	0.1033	0.1032	0.1120	0.1040
2	1	0.1188	0.1187	0.1354	0.1197
2	2	0.1694	0.1692	0.1987	0.1721
1	3	0.1888	0.1884	0.2154	0.1898
3	1	0.2180	0.2178	0.2779	0.2195
2	3	0.2475	0.2469	0.3029	0.2519
3	2	0.2624	0.2619	0.3418	0.2672
1	4	0.2969	0.2959	0.3599	0.2978
4	1	0.3319	0.3311	0.4773	0.3332

[a]Taken from Ref. [32].
[b]Results using NFSDT-I and NFSDT-II.

mode in the y-direction.

Using expressions (90) and (91) in the governing Eqs. (88) and (89) one obtains two equations which are written in matrix form as follows:

$$\begin{bmatrix} k_{11_{mn}} & k_{12_{mn}} \\ k_{12_{mn}} & k_{22_{mn}} \end{bmatrix} \begin{Bmatrix} \phi_{mn} \\ W_{mn} \end{Bmatrix} - \omega_{mn}^2 \begin{bmatrix} m_{11_{mn}} & 0 \\ 0 & m_{22_{mn}} \end{bmatrix} \begin{Bmatrix} \phi_{mn} \\ W_{mn} \end{Bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{92}$$

where

$$m_{11_{mn}} = \frac{\rho h^3}{12}\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]$$

$$m_{22_{mn}} = [\rho h]$$

Equation (92) is of a standard eigenvalue form, and solving it one gets free vibration frequencies.

Here it is observed from Eq. (92) that bending and shearing components are inertially uncoupled, but are elastically coupled.

Results in terms of nondimensionalized free vibration frequencies $\hat{\omega}_{mn}$, $\bar{\omega}_{mn}$ are tabulated in Tables 4 and 5.

The definitions used for nondimensionalized frequencies are

$\hat{\omega}_{mn} = \omega_{mn} h \sqrt{\rho / G}$ for isotropic plate and

$\bar{\omega}_{mn} = \omega_{mn} h \sqrt{\rho / Q_{11}}$ for orthotropic plate

9 Discussion on Results

The present theories have been applied to various problems to bring out their effectiveness. It is to be noted that square plate is an important case of rectangular plate analysis so square plate is considered for illustrative problems. The new theories are applied to solve the following problems.

1. Static analysis of simply-supported square plate
   - (a) Isotropic plate under sinusoidal load (results: Table 1)
   - (b) Isotropic plate under uniformly distributed load (results: Table 2)
   - (c) Orthotropic plate under uniformly distributed load (results: Table 3)
2. Free vibration analysis of simply-supported square plate
   - (a) Isotropic plate (Results: Table 4)
   - (b) Orthotropic plate (Results: Table 5)

It is to be noted that both the theories NFSDT-I and NFSDT-II give exactly the same results, therefore results given by present theories are presented under single column. Shear correction factor used for the analysis is 5/6.

9.1 On Static Analysis Results

9.1.1 On Stress Results. Tables 1–3 give the static analysis results for illustrative static analysis problems. Stresses σ_x, σ_y, and τ_{zx} (τ_{zx} is calculated using equations of equilibrium) obtained by present theories are same as those corresponding stresses obtained using CPT, and which are in excellent agreement with exact theory results.

9.1.2 On Displacement Results. Table 1 gives results for problem of simply supported isotropic square plate ($b/a=1, h/a=0.1$) under sinusoidal loading. Displacement results are in very good agreement with exact theory results (having maximum error of 0.62% only), whereas CPT has maximum error of −4.82% with respect to exact theory results. RPT gives almost same results as those given by present theories.

Table 2 gives results for problem of simply supported isotropic square plate under uniformly distributed loading. Errors (with respect to exact theory) in displacement results increase with increase in plate thickness for all the theories (i.e., CPT, Reissner's theory, and present theories) as expected. Results of displacements for Reissner's theory and present theories are nearly same and are in good agreement with results by exact theory.

Table 3 gives results for problem of simply supported orthotropic square plate under uniformly distributed loading. CPT results are not satisfactory as expected. Results for present theories are in good agreement with exact theory results. Here, results by Reissner's theory are marginally better than the results from present theories.

9.2 On Free Vibration Analysis Results

9.2.1 On Isotropic Plate Results. Results are tabulated in Table 4. As mode number increases the error with respect to exact theory increases for all three plate theories (i.e., CPT, FSDT, and present theories) as expected. Frequency values for FSDT and present theories are more or less same, which are in good agreement with exact theory results. Whereas, results using CPT are not satisfactory.

9.2.2 On Orthotropic Plate Results. Results are tabulated in Table 5. In case of orthotropic plate, the results are having similar trend as observed for isotropic plate. But, for orthotropic plate the results using FSDT are on lower side of exact theory results, whereas results using present formulation are on higher side of exact theory results.

10 Concluding Remarks

In this paper, two new displacement based, first-order shear deformation theories have been introduced. The effectiveness of the theories is brought out by applying them for static as well as dynamic analysis. The results obtained using both the theories are found to be in excellent agreement with the exact theory [33].

The following points need to be noted in respect of present NFSDT-I and NFSDT-II theories.

1. Use of the present theories results in two governing differential equations, wherein one is a fourth order differential equation, and another one is a second order differential equation.
   - (a) In the case of NFSDT-I, the equations are only inertially coupled and there is no elastic coupling at all.
   - (b) In the case of NFSDT-II, the equations are only elastically coupled and there is no inertial coupling at all.

   Yet, the results obtained by both theories are same, as the degrees of freedom allowed in displacement fields are same.
2. The number of unknown functions involved in NFSDT-I and NFSDT-II is only two. Even in the Mindlin's theory (a first-order shear deformation theory) three unknown functions are involved.
3. Both the displacement based theories are variationally consistent.
4. The theories (especially NFSDT-I) have strong similarity with the classical plate theory in many aspects (in respect of a governing equation, boundary conditions, moment expressions). Also, CPT comes out as a special case of NFSDT-I.
5. The finite elements based on NFSDT-I and NFSDT-II would be interesting to study.
6. The unique feature of NFSDT-I is that, if applied to laminated plate, it will give rise to elastically uncoupled equations, which is not possible by any other theory.

   But, to get proper shear contribution, in case of laminated plate analysis research needs to be carried out on, shear correction factor on the lines of Refs. [34–36].
7. It is to be noted that, CPT predicts stresses and moments accurately, where as free vibration frequency and displacement results using CPT are not satisfactory. Present theories

though give same expressions for stresses and moments as that of CPT, results for frequency and displacement have substantial improvement over CPT.

In conclusion, it can be said that simple new first-order shear deformation theories NFSDT-I and NFSDT-II can be effectively used for plate analysis.

References

[1] Reissner, E., 1944, "On the Theory of Bending of Elastic Plates," J. Math. Phys., **23**, pp. 184–191.

[2] Reissner, E., 1945, "The Effect of Transverse Shear Deformation on the Bending of Elastic Plates," Trans. ASME, **67**, pp. A-69–A-77.

[3] Mindlin, R. D., 1951, "Influence of Rotatory Inertia and Shear on Flexural Motions of Isotropic, Elastic Plates," Trans. ASME, **73**, pp. 31–38.

[4] Wang, C. M., Lim, G. T., Reddy, J. N., and Lee, K. H., 2001, "Relationships Between Bending Solutions of Reissner and Mindlin Plate Theories," Eng. Struct., **23**(7), pp. 838–849.

[5] Krishna Murty, A. V., 1977, "Higher Order Theory for Vibrations of Thick Plates," AIAA J., **15**(12), pp. 1823–1824.

[6] Lo, K. H., Christensen, R. M., and Wu, E. M., 1977, "A High-Order Theory of Plate Deformation Part 1: Homogeneous Plates," ASME J. Appl. Mech., **44**(4), pp. 663–668.

[7] Kant, T., 1982, "Numerical Analysis of Thick Plates," Comput. Methods Appl. Mech. Eng., **31**(1), pp. 1–18.

[8] Bhimaraddi, A., and Stevens, L. K., 1984, "A Higher Order Theory for Free Vibration of Orthotropic, Homogeneous, and Laminated Rectangular Plates," ASME J. Appl. Mech., **51**(1), pp. 195–198.

[9] Reddy, J. N., 1984, "A Refined Nonlinear Theory of Plates With Transverse Shear Deformation," Int. J. Solids Struct., **20**(9/10), pp. 881–896.

[10] Soldatos, K. P., 1988, "On Certain Refined Theories for Plate Bending," ASME J. Appl. Mech., **55**(4), pp. 994–995.

[11] Reddy, J. N., 1990, "A General Non-Linear Third-Order Theory of Plates With Moderate Thickness," Int. J. Non-Linear Mech., **25**(6), pp. 677–686.

[12] Hanna, N. F., and Leissa, A. W., 1994, "A Higher Order Shear Deformation Theory for the Vibration of Thick Plates," J. Sound Vib., **170**(4), pp. 545–555.

[13] Srinivas, S., Joga Rao, C. V., and Rao, A. K., 1970, "An Exact Analysis for Vibration of Simply-Supported Homogeneous and Laminated Thick Rectangular Plates," J. Sound Vib., **12**(2), pp. 187–199.

[14] Vasil'ev, V. V., 1992, "The Theory of Thin Plates," Mech. Solids, **27**(3), pp. 22–42.

[15] Liew, K. M., Xiang, Y., and Kitipornchai, S., 1995, "Research on Thick Plate Vibration: A Literature Survey," J. Sound Vib., **180**(1), pp. 163–176.

[16] Ghugal, Y. M., and Shimpi, R. P., 2002, "A Review of Refined Shear Deformation Theories of Isotropic and Anisotropic Laminated Plates," J. Reinf. Plast. Compos., **21**(9), pp. 775–813.

[17] Wang, J., Liew, K. M., Tan, M. J., and Rajendran, S., 2002, "Analysis of Rectangular Laminated Composite Plates via FSDT Meshless Method," Int. J. Mech. Sci., **44**(7), pp. 1275–1293.

[18] Xiang, Y., and Reddy, J. N., 2003, "Natural Vibration of Rectangular Plates With an Internal Line Hinge Using the First Order Shear Deformation Plate Theory," J. Sound Vib., **263**(2), pp. 285–297.

[19] Liew, K. M., Huang, Y. Q., and Reddy, J. N., 2003, "Vibration Analysis of Symmetrically Laminated Plates Based on FSDT Using the Moving Least Squares Differential Quadrature Method," Comput. Methods Appl. Mech. Eng., **192**(19), pp. 2203–2222.

[20] Liew, K. M., He, X. Q., Tan, M. J., and Lim, H. K., 2004, "Dynamic Analysis of Laminated Composite Plates With Piezoelectric Sensor/Actuator Patches Using the FSDT Mesh-Free Method," Int. J. Mech. Sci., **46**(3), pp. 411–431.

[21] Peng, L. X., Kitipornchai, S., and Liew, K. M., 2005, "Analysis of Rectangular Stiffened Plates Under Uniform Lateral Load Based on FSDT and Element-Free Galerkin Method," Int. J. Mech. Sci., **47**(2), pp. 251–276.

[22] Liew, K. M., Wang, C. M., Xiang, Y., and Kitipornchai, S., 1998, *Vibration of Mindlin Plates: Programming the P-Version Ritz Method*, 1st ed., Elsevier Science Oxford.

[23] Shimpi, R. P., 2002, "Refined Plate Theory and Its Variants," AIAA J., **40**(1), pp. 137–146.

[24] Jones, R. M., 1999, *Mechanics of Composite Materials*, 2nd ed., Taylor and Francis, Philadelphia, pp. 63–67.

[25] Shames, I. H., and Dym, C. L., 1975, *Energy and Finite Element Methods in Structural Mechanics*, Hemisphere Publishing, New York, pp. 257–274.

[26] Kabir, H. R. H., 1995, "A Shear-Locking Free Robust Isoparametric Three-Node Triangular Finite Element for Moderately-Thick and Thin Arbitrarily Laminated Plates," Comput. Struct., **57**(4), pp. 589–597.

[27] Reddy, J. N., 1997, "On Locking-Free Shear Deformable Beam Finite Elements," Comput. Methods Appl. Mech. Eng., **149**(1–4), pp. 113–132.

[28] Ainsworth, M., and Pinchedez, K., 2002, "The *hp*-MITC Finite Element Method for the Reissner-Mindlin Plate Problem," J. Comput. Appl. Math., **148**(2), pp. 429–462.

[29] Xiao, J. R., and McCarthy, M. A., 2003, "Meshless Analysis of Timoshenko Beams Based on a Locking-Free Formulation and Variational Approaches," Comput. Methods Appl. Mech. Eng., **192**(39–40), pp. 4403–4424.

[30] Brezzi, F., and Marini, L. D., 2003, "A Nonconforming Element for the Reissner-Mindlin Plate," Comput. Struct., **81**(8–11), pp. 515–522.

[31] Averill, R. C., and Reddy, J. N., 1992, "An Assessment of Four-Noded Plate Finite Elements Based On a Generalized Third-Order Theory," Int. J. Numer. Methods Eng., **33**(8), pp. 1553–1572.

[32] Srinivas, S., and Rao, A. K., 1970, "Bending, Vibration and Buckling of Simply-Supported Thick Orthotropic Rectangular Plates and Laminates," Int. J. Solids Struct., **6**(11), pp. 1463–1481.

[33] Srinivas, S., 1970, "*Three Dimensional Analysis of Some Plates and Laminates and a Study of Thickness Effects*," Ph.D. thesis, Dept. of Aeronautical Engineering, Indian Institute of Science, Bangalore, India.

[34] Whitney, J. M., 1973, "Shear Correction Factors for Orthotropic Laminates Under Static Load," ASME J. Appl. Mech., **40**(1), pp. 302–304.

[35] Vlachoutsis, S., 1992, "Shear Correction Factors for Plates and Shells," Int. J. Numer. Methods Eng., **33**(7), pp. 1537–1552.

[36] Birman, V., and Bert, C. W., 2002, "On the Choice of Shear Correction Factor in Sandwich Structures," J. Sandwich Struct. Mat., **4**(1), pp. 83–95.

An Efficient Approach to Estimate Critical Value of Friction Coefficient in Brake Squeal Analysis

Jinchun Huang
e-mail: huangjinchun@tsinghua.org.cn

Charles M. Krousgrill
e-mail: krousgri@ecn.purdue.edu

Anil K. Bajaj
e-mail: bajaj@ecn.purdue.edu

School of Mechanical Engineering,
Purdue University,
585 Purdue Mall,
West Lafayette, IN 47907-2088

Automotive brake squeal generated during brake applications has become a major concern in automotive industry. Warranty costs for brake noise related complaints have been greatly increasing in recent years. Brake noise and vibration control are also important for the improvement of vehicle quietness and passenger comfort. In this work, the mode coupling instability mechanism is discussed and a method to estimate the critical value of friction coefficient identifying the onset of brake squeal is presented. This is achieved through a sequence of steps. In the first step, a modal expansion method is developed to calculate eigenvalue and eigenvector sensitivities. Different types of mode couplings and their relationships with possible onset of squeal are discussed. Then, a reduced-order characteristic equation method based on the elastically coupled system eigenvalues and their derivatives is presented to estimate the critical value of friction coefficient. The significance of this method is that the critical value of friction coefficient can be predicted accurately without the need for a full complex eigenvalue analysis, making it possible to determine the sensitivity of system stability with respect to design parameters directly. [DOI: 10.1115/1.2423037]

Keywords: brake squeal, mode coupling instability, eigenvalue veering, mode merging, reduced-order characteristic equation

1 Introduction

For the last several decades, research on brake vibration and noise has been conducted using theoretical, experimental, and numerical approaches. In theoretical investigations, the complicated brake system has to be considerably simplified. Though these simplified models can provide some good insight into understanding the mechanism of brake squeal, they often do not provide accurate simulation for practical braking systems applications. Experimental approaches have also been widely used to measure the brake frequencies and mode shapes for the system in squeal, to investigate the effects of different parameters and operating conditions, and to verify possible solutions that can eliminate or significantly delay the onset of squeal. Accelerometer and double-pulsed laser holographic interferometry are two effective tools for determining the vibration mode shapes and forced response. However, the experiments are mostly expensive and time consuming. Furthermore, the remedies found from experimental study on one specific brake system may not be applicable to another type of brake systems. The numerical methods, on the other hand, are able to include more detailed brake component models along with the presence of friction. The frequencies and mode shapes of the components are first obtained through a finite element analysis [1,2] or assumed modes method [3,4], and the component models thus constructed are then coupled together into one brake system model through the normal contact and friction tractions at the braking surface. The resulting linear model equations are solved by a complex eigenvalue analysis and the influences of design parameters on the onset of brake squeal are investigated. Hamabe et al. [5] and Nack [6] directly performed a complex eigenvalue analysis of a finite element model of a brake assembly including the friction force, without reducing the total number of degrees of freedom. An excellent review on the status of brake squeal research is available in the work of Kinkaid et al. [7].

Substantial research on brake squeal has been conducted through numerical approaches. However, there are only a few works that investigate the brake squeal problem with sensitivity analysis though this powerful tool is widely used in many fields. The obstacle to conducting a comprehensive sensitivity analysis partially lies in the unknown explicit relationships between physical design variables and the level of system instability. Generally, the existence of complex eigenvalues with positive real parts indicates the presence of instability and the magnitude of the real part is used to represent the level of system instability (or squeal propensity). Since the eigenvalues are determined by a complex eigenvalue analysis and the eigenvalues with positive real parts are complex, it is very difficult to establish quantitative relationships between design parameters and the real parts of complex eigenvalues. By evaluating the stability from the mode coupling standpoint, Chung et al. [8] presented a new analysis method for brake squeal and defined the convergence rate as the stability metric.

It is a well known fact that a high friction coefficient increases the likelihood of squeal occurrence. It is found that the critical value of friction coefficient (μ_{cr}) can be used as a measure of the degree of system instability. The smaller the critical value of μ, the more is the system prone to instability and likely to exhibit unstable response. In addition, the fact that μ_{cr} is always real can be used to advantage to simplify the sensitivity analysis.

In this work, a reduced-order characteristic equation method is presented to estimate the critical value of friction coefficient. First a modal expansion method is developed to calculate the partial derivatives of eigenvalues with respect to two parameters. The couplings between the system modes as a function of the system parameters are discussed on the basis of these partial derivatives. Modes showing "curve crossing" and "veering away" in eigen-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received March 30, 2005; final manuscript received June 13, 2006. Review conducted by Oliver M. O'Reilly.

value loci [9] are shown to be unlikely to merge in a modal coupling and thus be immune from squeal, while modes showing "veering towards" are candidates for producing squeal. A reduced-order characteristic equation method is then developed to accurately predict the onset of squeal based on the elastically coupled structure's frequencies and their derivatives. Here, an "elastically coupled system" refers to the case when the friction coefficient is assumed zero, thus signifying that the brake components are not undergoing any relative motion. Once the changes in frequencies and their derivatives due to variation of system parameters are obtained at $\mu=0$, μ_{cr} can be estimated without constructing new system models. This makes it possible to save great efforts in the study of sensitivity of system stability boundary with respect to design parameters. The examples of using this method to estimate the frequency loci and stability boundaries are given for a drum brake system. We should note that the reduced-order characteristic equation can be viewed as a way to develop an asymptotic approximation to the eigenvalue loci for eigenvalues that are likely to coalesce. We should also note that the finite element packages ABAQUS and NASTRAN have become quite capable of performing a complex eigenvalue analysis. The value of the present work lies in performing an analysis as a function of the model system parameters, thereby providing insights into the dependence of propensity to squeal on system parameters. As an example, the approach can clearly identify the modes that can merge to lead to coupled-mode instability and how the design parameters influence this modal coupling. It also provides sensitivity of eigenvalues and modes to design parameters and thus an ability to perform design optimization.

2 Derivatives of the Eigensolutions

The nominal eigenvalue problem corresponding to a coupled drum brake system for a set of parameters was derived in Ref. [10] and is shown below

$$Lu_n^0=\lambda_n^0 u_n^0, \quad n=1,2,\ldots,N \tag{1}$$

Here u_n^0 and λ_n^0 are the eigensolutions, and

$$L=K+k_0A+\mu_0 k_0 B \tag{2}$$

In this equation, matrix K is an $N\times N$ stiffness matrix for the uncoupled brake components (e.g., the drum and the two shoes), matrix A is the stiffness contribution due to the lining and matrix B results from the friction coupling between the drum and the shoes. The parameters k_0 and μ_0 denote the lining stiffness and friction coefficient of the nominal system, respectively. Here the lining stiffness distribution at the drum-shoe interface can be nonuniform, and the factor k_0 is scaled to the order of 1 [10]. It should be noted that K and A are symmetric whereas B is asymmetric due to the nonconservative nature of the friction coupling forces. Furthermore, the system with matrix B neglected, or μ_0 set to zero, corresponds to the elastically coupled system case when the brake is not rotating and the components are only coupled by normal stiffness of the lining. Such eigenvalue problems that are linearly dependent on parameters also arise in other engineering applications such as flow-induced vibrations in structures conveying fluids [11] and in aeroelastic instabilities of structures [12].

The adjoint eigenvalue problem associated with the nominal system in Eq. (1) is

$$L^*\nu_n^0=\lambda_n^0\nu_n^0, \quad n=1,2,\ldots,N \tag{3}$$

where $L^*=K+k_0A+\mu_0k_0B^T$. It is assumed that the eigenvectors and their adjoints are normalized to satisfy the biorthogonality condition

$$\nu_m^{0T}u_n^0=\delta_{mn}=\begin{cases}1, & m=n\\ 0, & m\neq n\end{cases} \tag{4}$$

The sensitivity analysis requires derivatives of the eigenvalues and eigenvectors with respect to the system parameters of interest. Rudisill and Chu [13] presented two numerical methods for an eigenvalue sensitivity analysis: an iterative method and an algebraic method. Perkins and Mote [9] used a modal expansion method to calculate the derivatives with respect to one parameter. Though the algebraic method is efficient, the modal expansion method provides some physical interpretations and may be used to investigate the coupling of the modes. Here the modal expansion approach is extended to calculate the derivatives of the eigensolutions with respect to two parameters. Note that the same approach can be used for systems with more than two parameters.

The elastically coupled system will be used in this work as the nominal system. The eigensolutions of this self-adjoint system (since $B=0$) can be computed reliably and verified by experimental approaches. Hence, the unperturbed system is chosen to be the system without friction, the derivatives of the eigenvalues at $\mu_0=0$ will be calculated, and the sensitivity analysis will be conducted based on the information for an elastically coupled model.

Let $k=k_0+\varepsilon_k$ and $\mu=\mu_0+\varepsilon_\mu=\varepsilon_\mu$, where ε_k and ε_μ denote small perturbations in lining stiffness and the friction coefficient, respectively, so that the system is still stable. The change in the mass matrix is assumed to be negligible. The perturbed eigenvalue problem then becomes

$$(L^0+l)u_n=\lambda_n u_n \tag{5}$$

where $L^0=K+k_0A$, and the perturbation matrix l is

$$l=\varepsilon_kA+\varepsilon_\mu k_0B+\varepsilon_k\varepsilon_\mu B \tag{6}$$

For a precritical value of μ, let the perturbed eigenvectors be expressed as a linear combination of unperturbed eigenvectors

$$u_n=\sum_{j=1}^{N}c_{nj}u_j^0 \tag{7}$$

Substituting Eq. (7) into Eq. (5), premultiplying by v_m^T, and using Eq. (4) give

$$c_{nm}(\lambda_n-\lambda_m^0)=\sum_{j=1}^{N}c_{nj}(\varepsilon_kE_{kmj}+\varepsilon_\mu E_{\mu mj}+\varepsilon_k\varepsilon_\mu E_{\mu kmj}) \tag{8}$$

where

$$\begin{aligned}E_{kmj}&=\nu_m^{0T}Au_j^0\\ E_{\mu mj}&=\nu_m^{0T}(k_0B)u_j^0\\ E_{\mu kmj}&=\nu_m^{0T}Bu_j^0\end{aligned} \tag{9}$$

From above, it is clear that $E_{kmj}=E_{kjm}$ due to the symmetry of matrix A.

Consider the following power series expansions for c_{nm} and λ_n with respect to the parameters k and μ

$$\begin{aligned}c_{nm}=c_{nm}^0&+(\varepsilon_\mu c_{nm,\mu}+\varepsilon_kc_{nm,k})\\ &+(\varepsilon_\mu^2c_{nm,\mu^2}+\varepsilon_\mu\varepsilon_kc_{nm,\mu k}+\varepsilon_k^2c_{nm,k^2})+\cdots\end{aligned} \tag{10}$$

$$\lambda_n=\lambda_n^0+(\varepsilon_\mu\lambda_{n,\mu}+\varepsilon_k\lambda_{n,k})+(\varepsilon_\mu^2\lambda_{n,\mu^2}+\varepsilon_\mu\varepsilon_k\lambda_{n,\mu k}+\varepsilon_k^2\lambda_{n,k^2})+\cdots \tag{11}$$

where for $s\geqslant 0, t\geqslant 0$

$$\lambda_{n,\mu^sk^t}=\frac{1}{s!t!}\frac{\partial^{(s+t)}\lambda_n}{\partial\mu^s\partial k^t}$$

Substitution of Eqs. (10) and (11) into Eq. (8) yields

$$\begin{aligned}&[c_{nm}^0+(\varepsilon_\mu c_{nm,\mu}+\varepsilon_kc_{nm,k})\\ &\quad+(\varepsilon_\mu^2c_{nm,\mu^2}+\varepsilon_\mu\varepsilon_kc_{nm,\mu k}+\varepsilon_k^2c_{nm,k^2})+\cdots]\\ &\quad\times[(\lambda_n^0-\lambda_m^0)+(\varepsilon_\mu\lambda_{n,\mu}+\varepsilon_k\lambda_{n,k})\\ &\quad+(\varepsilon_\mu^2\lambda_{n,\mu^2}+\varepsilon_\mu\varepsilon_k\lambda_{n,\mu k}+\varepsilon_k^2\lambda_{n,k^2})+\cdots]\end{aligned}$$

$$= \sum_{j=1}^{N} [c_{nj}^0 + (\varepsilon_\mu c_{nj,\mu} + \varepsilon_k c_{nj,k}) + (\varepsilon_\mu^2 c_{nj,\mu^2} + \varepsilon_\mu \varepsilon_k c_{nj,\mu k} + \varepsilon_k^2 c_{nj,k^2}) + \cdots] \times [\varepsilon_k E_{kmj} + \varepsilon_\mu E_{\mu mj} + \varepsilon_k \varepsilon_\mu E_{\mu kmj}] \quad (12)$$

Equating the terms of the same orders in Eq. (12) gives

$$c_{nm}^0 = c_{nm,\mu^0 k^0} = \delta_{nm} \quad (13)$$

and for $s>0, t\geqslant 0$ or $t>0, s\geqslant 0$

$$c_{nn,\mu^s k^t} = 0 \quad (14)$$

$$c_{nm,\mu^s k^t} = \frac{1}{\lambda_n^0 - \lambda_m^0} \left[\sum_{j=1}^{N} (c_{nj,\mu^{s-1}k^t} E_{\mu mj} + c_{nj,\mu^s k^{t-1}} E_{kmj} + c_{nj,\mu^{s-1}k^{t-1}} E_{\mu kmj}) - \sum_{i,j}{}^{*} c_{nm,\mu^i k^j} \lambda_{n,\mu^{s-i}k^{t-j}} \right], \quad n \neq m \quad (15)$$

$$\lambda_{n,\mu^s k^t} = \sum_{j=1}^{N} (c_{nj,\mu^{s-1}k^t} E_{\mu nj} + c_{nj,\mu^s k^{t-1}} E_{knj} + c_{nj,\mu^{s-1}k^{t-1}} E_{\mu knj}) \quad (16)$$

where $\Sigma_{i,j}^{*}$ in Eq. (15) implies summation with the indices over the range: $0\leqslant i\leqslant s, 0\leqslant j\leqslant t$, and $0<i+j<s+t$. It is also implied that $c_{nm,\mu^i k^j}=0, \lambda_{n,\mu^i k^j}=0$, if $i<0$ or $j<0$ on the right hand side of Eqs. (15) and (16).

It is assumed that there are no repeated eigenvalues for the unperturbed system, i.e., $\lambda_n^0 \neq \lambda_m^0$, which is typically true for coupled brake systems at $\mu=0$. If the eigenvalues are repeated, other methods such as revised series expansions [14] and the extension of Nelson's method [15] may be used to calculate the derivatives of eigensolutions. Detailed references can be found in Refs. [16,17].

With a way to compute eigenvalues and eigenvectors as a function of system parameters, we now turn to a discussion of coupling of modes that is the essential mechanism for coupled-mode instability in elastic systems with follower and gyroscopic forces [16,17].

3 Mode Coupling in the Discretized System

In the presence of both lining stiffness and friction coupling (i.e., k_0 and μ_0 are greater than zero in Eq. (1)), the eigenvalues λ_n are no longer guaranteed to be real since matrix B is not symmetrical. The existence of complex roots with positive real parts indicates the presence of a "mode-merging" instability, which can lead to brake squeal vibrations. A typical situation in eigenloci merging is illustrated in Fig. 1 [10]. As the friction coefficient is increased, the two undamped frequencies come close to coalesce and the corresponding mode shapes of the brake system become identical at $\mu=\mu_{cr}$. The value of friction coefficient μ that demarcates the possible stable and unstable oscillations will be referred to as a critical value of friction coefficient μ_{cr}.

Mode merging is typically observed to occur between two neighboring or close modes. Their frequencies usually have small separation at $\mu=0$. However, not all pairs of modes with small separations come to merge even for a large value of μ. The series expansions in Eqs. (7), (10), and (11) for the perturbed eigensolutions can help explain these phenomena.

3.1 Mode Coupling as a Function of Lining Stiffness. For a pair of close eigenvalues λ_r^0 and λ_s^0 ($\lambda_r^0<\lambda_s^0$) of the unperturbed system, let us consider the coupling between this pair of modes and neglect the coupling between these two modes and the rest of the structural modes. Then, the summations can be dropped in the

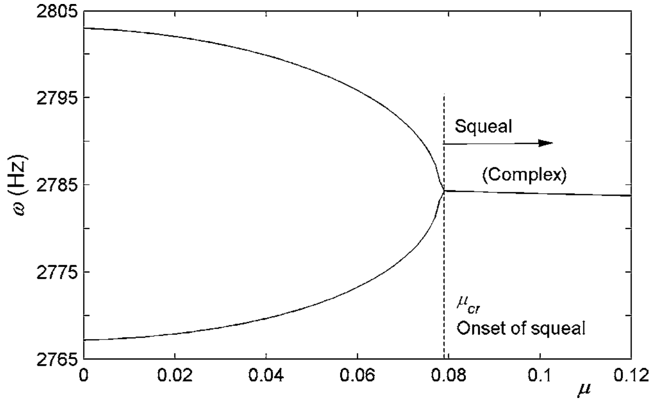


Fig. 1 A typical occurrence of mode merging for a pair of modes that leads to "coupled-mode" instability. The frequency loci for the two frequencies as a function of the friction coefficient μ merge at $\mu=\mu_{cr}$ and beyond this point, the two frequencies are a complexconjugate pair. Here, only the real part of the two frequencies is shown.

expressions for the second derivatives of eigenvalues. With this simplification, the power series expansions for perturbed eigensolutions with respect to the lining stiffness k for a precritical value of μ can be shown to become

$$\lambda_r = \lambda_r^0 + \varepsilon_k E_{krr} + \varepsilon_k^2 \frac{E_{krs}E_{ksr}}{\lambda_r^0 - \lambda_s^0} + \text{h.o.t.} \quad (17)$$

$$\lambda_s = \lambda_s^0 + \varepsilon_k E_{kss} + \varepsilon_k^2 \frac{E_{krs}E_{ksr}}{\lambda_s^0 - \lambda_r^0} + \text{h.o.t.} \quad (18)$$

$$u_r = u_r^0 + \varepsilon_k \frac{E_{ksr}}{\lambda_r^0 - \lambda_s^0} u_s^0 + \text{h.o.t.} \quad (19)$$

$$u_s = u_s^0 + \varepsilon_k \frac{E_{krs}}{\lambda_s^0 - \lambda_r^0} u_r^0 + \text{h.o.t.} \quad (20)$$

From Eqs. (19) and (20) it can be seen that E_{ksr} provides a measure of coupling between the rth and the sth perturbed modes for the symmetric system. If $E_{ksr}=E_{krs}=0$, the rth mode and the sth mode are uncoupled at the lowest order in ε_k as the lining stiffness k is varied. From Eqs. (17) and (18), if the two eigenfrequency loci or curves approach and cross as k changes, they simply intersect in a locally linear fashion, which is illustrated in Fig. 2(a). In Fig. 2(b), the inner products of the perturbed and unperturbed eigenvectors are plotted, which clearly shows that the two modes are totally independent of each other. This case is referred to as curve crossing. Since there is no coupling between the modes, they are not expected to merge or lead to instability when friction, and thus an asymmetric component in the stiffness matrix, is included in the system model.

On the other hand, if $E_{ksr}=E_{krs}\neq 0$, the rth mode and the sth mode are coupled as lining stiffness k is varied. From Eqs. (17) and (18), if the two frequency loci approach each other, the lower frequency locus $\lambda_r(\varepsilon_k)$ becomes increasingly concave downward and the higher frequency locus $\lambda_s(\varepsilon_k)$ becomes increasingly concave upward, as shown in Fig. 3(a). The smaller the initial separation between the frequency loci, the stronger the repulsion effect is. The two loci approach and then diverge without intersection. This locus divergence is referred to as curve veering away. Figure 3(b) shows that the two eigenvectors are almost interchanged during veering in a rapid and continuous way, which is consistent with Eqs. (19) and (20). Since there exists a strong coupling between the modes, they are likely candidates for merging when friction exists in the system (or is included in the model).

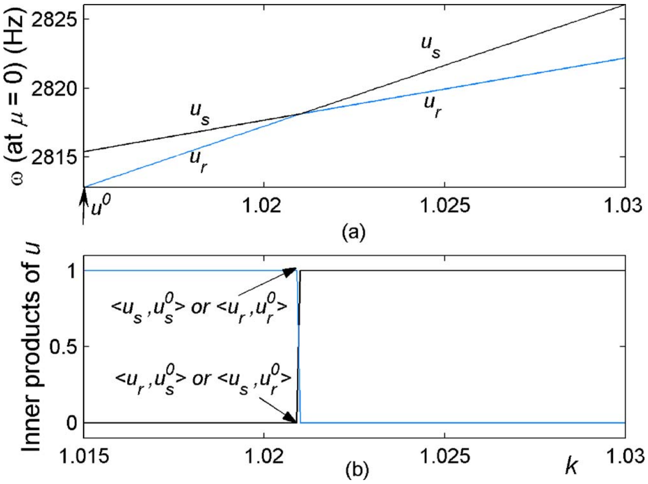


Fig. 2 Frequency loci and inner products of eigenvectors as a function of the lining stiffness *k*, showing curve crossing for modes *r* and *s*. This corresponds to the case where $E_{ksr}=E_{krs}=0$ in eigenvalue expressions in Eqs. (17) and (18).

3.2 Mode Coupling as a Function of Friction Coefficient. Now for a pair of close eigenvalues λ_r^0 and λ_s^0 ($\lambda_r^0<\lambda_s^0$), consider the expansions of the perturbed eigensolutions with respect to the friction coefficient μ for a precritical case

$$\lambda_r=\lambda_r^0+\varepsilon_\mu E_{\mu rr}+\varepsilon_\mu^2\frac{E_{\mu rs}E_{\mu sr}}{\lambda_r^0-\lambda_s^0}+\text{h.o.t.} \tag{21}$$

$$\lambda_s=\lambda_s^0+\varepsilon_\mu E_{\mu ss}+\varepsilon_\mu^2\frac{E_{\mu rs}E_{\mu sr}}{\lambda_s^0-\lambda_r^0}+\text{h.o.t.} \tag{22}$$

$$u_r=u_r^0+\varepsilon_\mu\frac{E_{\mu sr}}{\lambda_r^0-\lambda_s^0}u_s^0+\text{h.o.t.} \tag{23}$$

$$u_s=u_s^0+\varepsilon_\mu\frac{E_{\mu rs}}{\lambda_s^0-\lambda_r^0}u_r^0+\text{h.o.t.} \tag{24}$$

From Eqs. (9), $E_{\mu sr}$ and $E_{\mu rs}$ are not guaranteed to be equal in either magnitude or sign since matrix B is not symmetric. In the absence of friction coupling between the pair of modes, we have $E_{\mu sr}=E_{\mu rs}=0$. In this case, if two eigenfrequency loci or curves approach each other as μ is varied, they just intersect without merging, as shown in Fig. 4(*a*).

In the presence of friction coupling, two different cases must be treated. When $E_{\mu sr}$ and $E_{\mu rs}$ are of same sign, $E_{\mu rs}E_{\mu sr}/(\lambda_r^0-\lambda_s^0)<0$ and $E_{\mu rs}E_{\mu sr}/(\lambda_s^0-\lambda_r^0)>0$. The two frequency loci veer away from each other as in the self-adjoint case. It is a repulsive coupling and the modes do not lead to instability, as shown in Fig. 4(*b*). When $E_{\mu sr}$ and $E_{\mu rs}$ are of opposite sign, the two frequency loci veer towards each other when the friction is increased. In this case the two modes can merge and lead to squeal, as shown in Fig. 4(*c*). The rate of merging is influenced by two factors: the sepa-

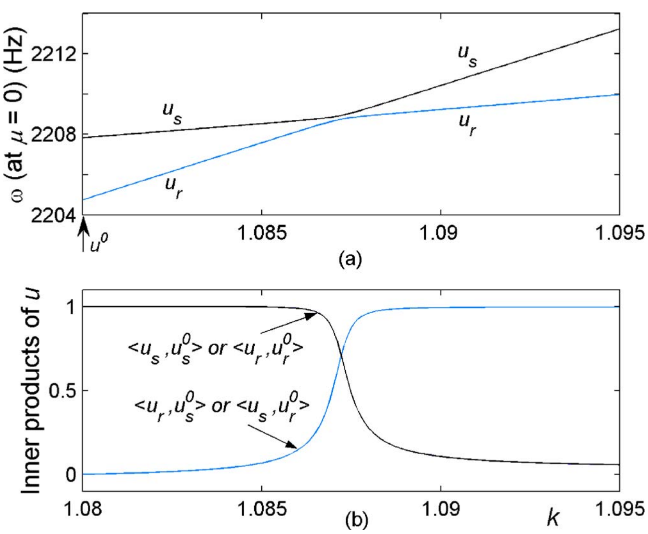


Fig. 3 Frequency loci and inner products of eigenvectors as a function of the lining stiffness *k*, showing "curve veering away" for modes *r* and *s* when $E_{ksr}=E_{krs}\neq 0$ in Eqs. (17) and (18)

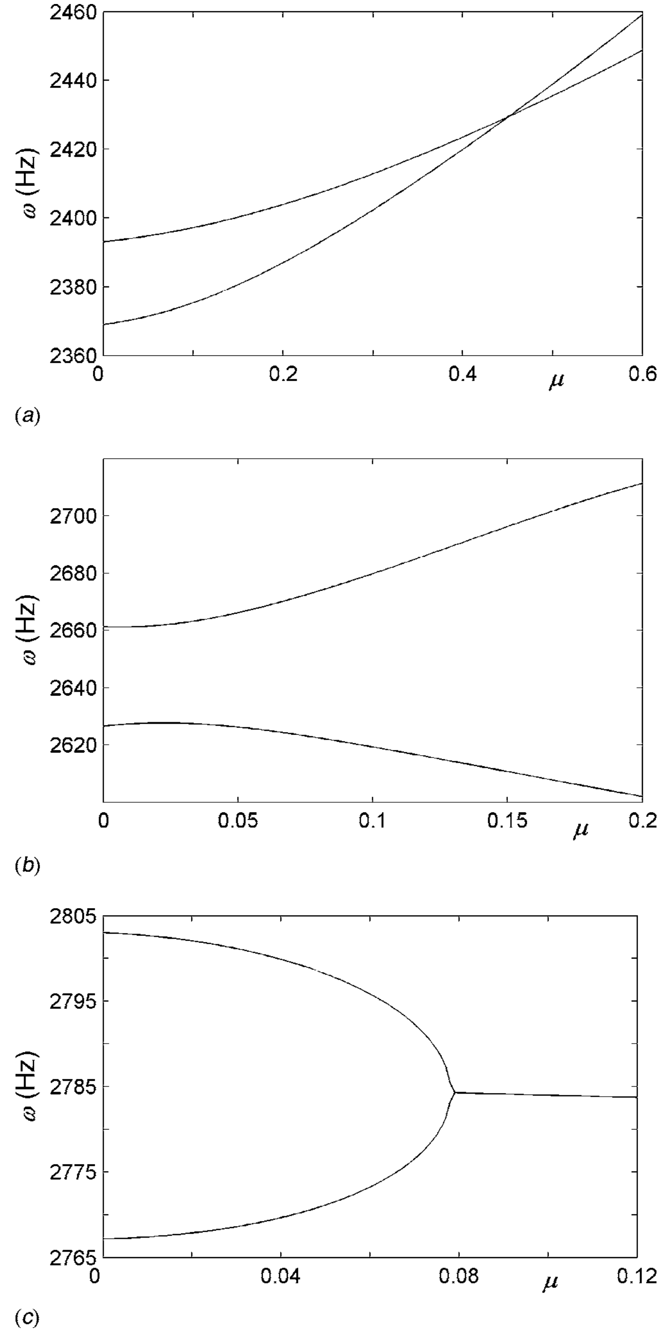


Fig. 4 Examples of eigenfrequency loci interactions for two modes as the friction coefficient μ is varied: (*a*) curve crossing, (*b*) veering away, and (*c*) veering towards

ration of the unperturbed frequencies and the magnitudes of $E_{\mu sr}$ and $E_{\mu rs}$. The latter quantities represent strength of friction coupling between the two interacting modes. With small separation and large magnitudes of $E_{\mu sr}$ and $E_{\mu rs}$, the pair of modes is prone to merging at a small value of $\mu_{\rm cr}$. It can be seen from Eqs. (23) and (24) that a pair of eigenvectors u_r and u_s can become more similar as μ is increased only when $E_{\mu sr}/(\lambda_r^0-\lambda_s^0)$ and $E_{\mu rs}/(\lambda_s^0-\lambda_r^0)$ are of same sign.

Hence, when μ is varied, the two modes showing curve crossing and veering away do not merge, while the modes showing curve veering towards are very likely to merge to result in squeal if their separation is small and/or if sufficient friction coupling exists.

It should be noted that Eqs. (17)–(24) are simplified versions of the original expansions where the summations in terms such as $\lambda_{r,\mu^2}=\Sigma_{j=1,\neq r}^{N}[E_{\mu jr}E_{\mu rj}/(\lambda_r^0-\lambda_j^0)]$ were dropped when the coupling with other modes is neglected. In some cases this simplification is not valid if other modes have a strong influence on this pair of modes. This shows that the mode couplings among the modes can become very complicated. In some cases, the higher order perturbations also play an important role in mode merging.

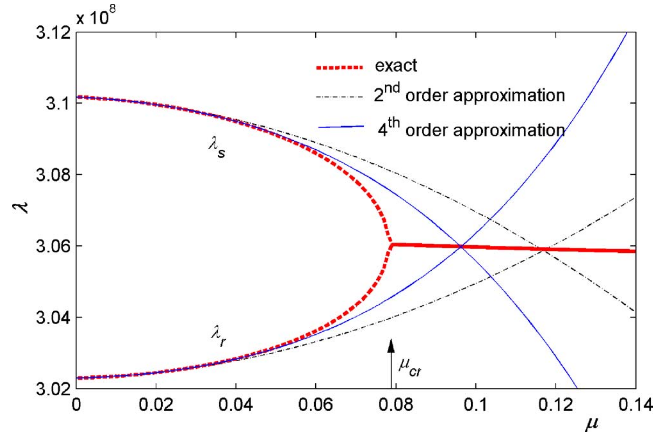


Fig. 5 Estimates of $\mu_{\rm cr}$ by the power series expansion approach. Here $\mu_{\rm cr}$ is defined by the intersection of approximations to the two eigenfrequency loci.

4 The Reduced-Order Characteristic Equation Method

Typically the order of the modal-based brake system model is quite large ($N>100$). Note that this number is different from the degrees of freedom for any finite element model that may be created for a brake system to investigate the stability of the system and its propensity to squeal. This creates a formidable task of studying coupling among a large number of modes for the prediction of onset of squeal. The relationships between several system parameters and brake squeal are also complicated. However, in most cases the magnitudes and signs of $E_{\mu sr}$ and $E_{\mu rs}$ can be used to predict whether the rth and the sth modes would merge (veer towards) or not (cross or veer away). The interactions between a single pair of modes seem to have strong influence on mode merging and the two modes that are likely candidates for merging can be identified by an analysis explained in Sec. 3. This inspires us to now develop a method for squeal prediction by using the frequencies and their derivatives at $\mu=0$, focusing on the eigenmodes that exhibit the property of veering towards for the two modes under consideration. In order to get a good prediction, the derivatives from the full model will be used, as shown in Eq. (16).

The natural thought of using frequency and derivative information to predict $\mu_{\rm cr}$ is to use the power series expansions of the eigenvalues

$$\begin{aligned}\lambda_r &= \lambda_r^0+\mu\lambda_{r,\mu}+\mu^2\lambda_{r,\mu^2}+\mu^3\lambda_{r,\mu^3}+\cdots\\ \lambda_s &= \lambda_s^0+\mu\lambda_{s,\mu}+\mu^2\lambda_{s,\mu^2}+\mu^3\lambda_{s,\mu^3}+\cdots\end{aligned}\qquad(25)$$

through the eigenvalue derivatives found in the last section.

The intersection of the two curves produces an approximation of the critical value of friction coefficient $\mu_{\rm cr}$. An example is shown in Fig. 5, in which the frequency loci are estimated by series expansions up to the second order and fourth order in μ, respectively. As the order of the eigenfrequency loci expansion is increased, a better approximation is seen to be achieved.

The drawback of this approach is obvious: it does not reflect the mode coupling between the modes, nor does it capture $d\lambda_r/d\mu=\infty$ at $\mu=\mu_{\rm cr}$. Consequently, the two frequency loci are free to cross instead of merging. In addition, a critical value of μ may be predicted even for the cases in which the two modes may not actually be coupled sufficiently to lead to mode merging. This is because the eigenvalue expansions do not explicitly account for the eigenvalue veering towards properties for the two frequency loci. As a result, the power series approach does not appear to be very useful for the prediction of mode merging and the estimate of $\mu_{\rm cr}$ is poor at best and erroneous in the worst case. A better method to estimate $\mu_{\rm cr}$ is needed.

The proposed approach in the following is based on the observation that the merging of two eigenfrequency loci is a situation in which a double root arises as a function of the parameter μ. Beyond the critical value of the parameter at which the frequencies merge, the eigenfrequencies become complex. Thus, the eigenloci pair that merges as a function of μ must in some way form roots of a quadratic equation that vary as a function of the other system parameters. It is this quadratic equation that we need to extract from the system characteristic equation and then approximate it as a function of system parameters.

The characteristic equation for the full N-mode model is an Nth order polynomial in λ that can be factored into the following form:

$$f(\lambda)g(\lambda)=[\lambda^2+b(\mu)\lambda+c(\mu)]\left(\sum_{j=2}^{N}d_j\lambda^{N-j}\right)=0\qquad(26)$$

where the coefficients $b(\mu)$ and $c(\mu)$ are analytical functions of the friction coefficient μ as well as the other model parameters (the dependence on these parameters is suppressed here for notational convenience).

The coefficients $b(\mu)$ and $c(\mu)$ can be expressed in terms of their truncated power series about $\mu=0$ as

$$b(\mu)\approx\sum_{i=1}^{M}b_i\mu^{i-1}\qquad(27)$$

$$c(\mu)\approx\sum_{i=1}^{M}c_i\mu^{i-1}\qquad(28)$$

where b_i and c_i are sets of M yet-to-be-determined coefficients, with (hopefully) $M\ll N$. With this expansion, the factored polynomial $f(\lambda)$ with roots λ_r and λ_s is approximated by

$$f(\lambda)=\lambda^2+\sum_{i=1}^{M}(b_i\mu^{i-1})\lambda+\sum_{i=1}^{M}(c_i\mu^{i-1})=0\qquad(29)$$

Since $f(\lambda)=0$ at every value of μ, its total derivatives with respect to μ also vanish. Note that these expressions also involve the derivatives of the eigenvalues as well as the eigenvalues and the coefficients b_i and c_i.

The coefficients b_i and c_i can be determined by matching known values of λ_r, λ_s and their derivatives with respect to μ at $\mu=0$

$$f(\lambda_r)=\lambda_r^2+b_1\lambda_r+c_1=0$$

$$\frac{df(\lambda_r)}{d\mu}=2\lambda_r\lambda_{r,\mu}+b_1\lambda_{r,\mu}+b_2\lambda_r+c_2=0$$

$$\frac{d^2f(\lambda_r)}{d\mu^2}=2\lambda_{r,\mu}^2+2\lambda_r\lambda_{r,\mu^2}+b_1\lambda_{r,\mu^2}+2b_2\lambda_{r,\mu}+2b_3\lambda_r+2c_3=0$$

Combining these equations for the pair of roots λ_r and λ_s gives the following:

$$\begin{bmatrix} 1 & \lambda_r & 0 & 0 & 0 & 0 \\ 1 & \lambda_s & 0 & 0 & 0 & 0 \\ 0 & \lambda_{r,\mu} & 1 & \lambda_r & 0 & 0 \\ 0 & \lambda_{s,\mu} & 1 & \lambda_s & 0 & 0 \\ 0 & \lambda_{r,\mu^2} & 0 & 2\lambda_{r,\mu} & 2 & 2\lambda_r \\ 0 & \lambda_{s,\mu^2} & 0 & 2\lambda_{s,\mu} & 2 & 2\lambda_s \\ \vdots & & & & & \end{bmatrix}_{2M\times 2M} \begin{Bmatrix} c_1 \\ b_1 \\ c_2 \\ b_2 \\ c_3 \\ b_3 \\ \vdots \end{Bmatrix}_{2M\times 1} = \begin{Bmatrix} -\lambda_r^2 \\ -\lambda_s^2 \\ -2\lambda_r\lambda_{r,\mu} \\ -2\lambda_s\lambda_{s,\mu} \\ -2(\lambda_{r,\mu})^2-2\lambda_r\lambda_{r,\mu^2} \\ -2(\lambda_{s,\mu})^2-2\lambda_s\lambda_{s,\mu^2} \\ \vdots \end{Bmatrix}_{2M\times 1} \tag{30}$$

or

$$[D]\begin{Bmatrix} c_i \\ b_i \end{Bmatrix}=\{H\} \tag{31}$$

The unknown coefficients b_i and c_i can be solved explicitly from Eq. (30) to produce for $i=1,2,\ldots,M$

$$b_i=-(\lambda_{r,\mu^{i-1}}+\lambda_{s,\mu^{i-1}})$$

$$c_i=\sum_{j=0}^{i-1}\lambda_{r,\mu^j}\lambda_{s,\mu^{(i-1-j)}} \tag{32}$$

Using the above, the roots of Eq. (29) define the two eigenvalue loci. The coupling between modes r and s is clearly demonstrated by this reduced-order characteristic equation. This coupling is clearly seen to depend on the information about the roots and their derivatives at $\mu=0$. With this approach, it is now very easy to see the influences on mode merging of the separations between the two frequencies and the derivatives at $\mu=0$.

Since $\partial\lambda/\partial\mu=\infty$ at $\mu=\mu_{\mathrm{cr}}$ [4], the critical value of μ is one for which

$$\left.\frac{\partial\mu}{\partial\lambda}\right|_{\mu_{\mathrm{cr}}}=0 \tag{33}$$

Taking the derivatives of the reduced-order characteristic Eq. (29) with respect to λ and using Eq. (33) yield

$$\lambda_{\mathrm{cr}}=-\frac{1}{2}\sum_{i=1}^{M}b_i\mu_{\mathrm{cr}}^{i-1} \tag{34}$$

Substituting Eq. (34) into Eq. (29) yields an equation for μ_{cr}

$$\left[\sum_{i=1}^{M}(b_i\mu_{\mathrm{cr}}^{i-1})\right]^2=4\sum_{i=1}^{M}(c_i\mu_{\mathrm{cr}}^{i-1}) \tag{35}$$

Thus, μ_{cr} can be solved for numerically from the roots of polynomial expression in Eq. (35), giving an explicit approach to determine the critical value of μ. Note that the value of μ_{cr} could be large and is not related to the smallness of the perturbation ε_μ used in Sec. 3. Furthermore, Eq. (35) can also be directly derived from Eq. (28) as a condition of double roots without the intermediate steps in Eqs. (33) and (34).

Setting $M=3$ in Eq. (35) and neglecting the terms which are higher than second order in μ_{cr}, a simple expression for estimate of μ_{cr} is obtained from the reduced-order characteristic equation

$$\mu_{\mathrm{cr}}=\frac{1}{4}\frac{\lambda_s^0-\lambda_r^0}{\sqrt{-E_{\mu rs}E_{\mu sr}}}\left[2+\frac{E_{\mu ss}-E_{\mu rr}}{\sqrt{-E_{\mu rs}E_{\mu sr}}}\right] \tag{36}$$

This equation clearly shows that the squeal propensity of interacting modes is strongly related to the mode shape coefficients with respect to the friction coupling matrix (i.e., $E_{\mu sr}$ as defined in Eq. (9)) and the initial eigenvalues separation at $\mu=0$. As discussed earlier, $E_{\mu rs}E_{\mu sr}<0$ is a necessary condition for merging of a pair of modes. Here $E_{\mu ss}-E_{\mu rr}$ is the difference in the slopes of λ_s and λ_r at $\mu=0$, and typically the term $(E_{\mu ss}-E_{\mu rr})/\sqrt{-E_{\mu rs}E_{\mu sr}}$ is negligible compared to 2. Hence, Eq. (36) can be further simplified to:

$$\mu_{\mathrm{cr}}\approx\frac{1}{2}\frac{\lambda_s^0-\lambda_r^0}{\sqrt{-E_{\mu rs}E_{\mu sr}}} \tag{37}$$

This relation shows that qualitatively, μ_{cr} is proportional to the initial eigenvalue separation and is inversely related to the strength of friction coupling.

5 Examples of Application of Reduced-Order Characteristic Equation for Brake Squeal Prediction

The technique developed above for estimating the critical value of friction coefficient μ_{cr} will now be applied to the mathematical model of a light-duty truck drum brake system described in Ref. [10]. This brake system model is based on the modal information extracted from finite element models for the individual drum and shoe components. The component models of the drum and brake shoes are coupled through the shoe lining material, as detailed in Ref. [10]. For this system, a total of 150 component modes were incorporated into the coupled model. These include the first 50 modes each for the drum and two shoes. For this system, the stiffness matrix K in Eq. (1) is composed of the individual component stiffness or frequencies. The matrix A is the stiffness matrix representing coupling due to the shoe lining that couples modes of the drum to those of the two shoes. The matrix B results from the tangential forces that are produced by friction at the drum-lining interfaces.

The complex eigenvalue analysis of Eq. (1) was conducted using lining stiffness $k=1.0$. Here k is a normalized ratio of lining

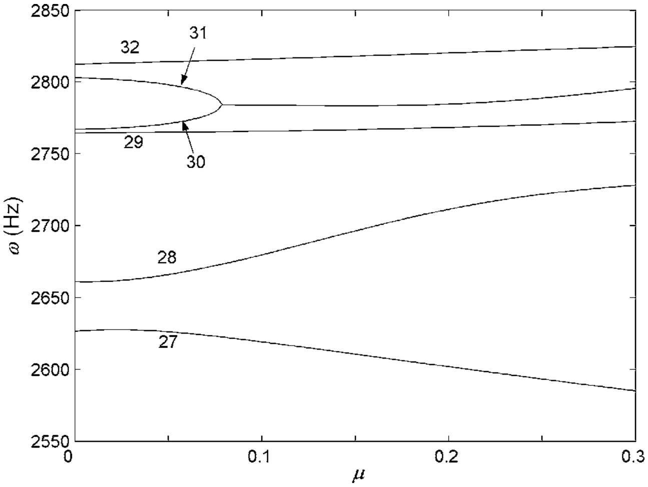


Fig. 6 Variation of the frequencies of modes 27 to 32 with μ for a drum brake model (See Ref. [10])

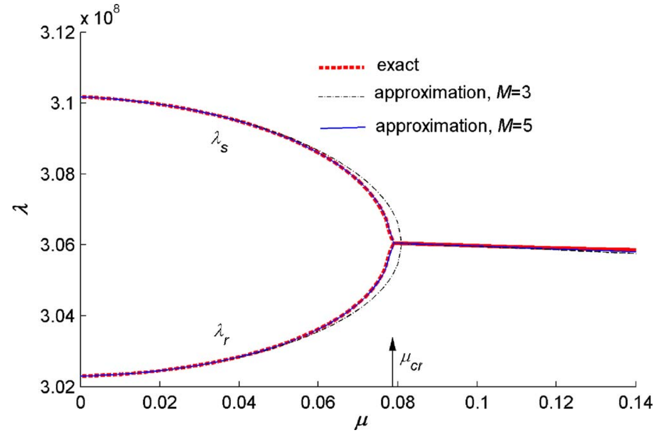


Fig. 7 Estimate of μ_{cr} by the reduced-order characteristic equation for modes 30 and 31 in Fig. 6 for the drum brake system (see Ref. [10]). The exact results are from the complex eigenvalue analysis, whereas the approximate eigenfrequency loci are based on approximation in Eq. (29).

stiffness to nominal lining stiffness per unit area. The variation of the frequencies (or the imaginary parts of eigenvalues in case of complex frequencies) for modes 27–32 with the friction coefficient μ is shown in Fig. 6.

Friction coupling coefficients $E_{\mu sr}$ and $E_{\mu rs}$ are also calculated for several pairs of modes at $k=1.0$, and the results are listed in Table 1. The predictions of the perturbation analysis above regarding the merging or nonmerging of any pairs of modes were verified by the complex eigenvalue analysis. The coupling between the pairs of modes under consideration is determined by the magnitudes and signs of $E_{\mu sr}$ and $E_{\mu rs}$. Note that, even though the frequency separation between the modes 29 and 30 at $\mu=0$ is very small, the mode 30 merges with mode 31 rather than with mode 29, as shown in Fig. 6. This is because the eigenloci for modes 30 and 31 veer towards and there is no coupling between the modes 29 and 30, as clearly shown in Table 1. Compared to modes 30 and 31, the frequency separation of modes 29 and 32 is larger and their values of $E_{\mu sr}$ and $E_{\mu rs}$ are smaller. As a result, modes 29 and 32 merge at a critical value of μ which is much larger than that for modes 30 and 31. For modes 27 and 28, $E_{\mu sr}$ and $E_{\mu rs}$ are of same sign and they veer away without merging.

Figure 7 shows the frequency loci for modes 30 and 31 at $k=1$, as estimated by the reduced-order characteristic equation, with the polynomials up to the second order ($M=3$) and fourth order ($M=5$) in μ. Compared to Fig. 5, the results of the same order approximations are improved significantly. In this example, the fourth order characteristic equation predicts almost exact frequency loci, offering a very good estimation to μ_{cr}.

A more complicated example is shown in Fig. 8 for $k=1.3$. Here, two pairs of modes merge at relatively large values of μ. The frequency loci for each pair of modes is estimated by the reduced-order characteristic equation with $M=5$ in Eq. (29).

The technique developed above can now be used to discuss the system stability and its dependence on system parameters. First note that at a given set of system parameters, the critical values of μ for all pairs of modes which tend to merge in the frequency range of interest can be estimated, and the minimum μ_{cr} defines the stability boundary for the drum brake model.

The lining stiffness is an important parameter which is known to vary with changes in applied pressure and wear of the shoe lining. Since lining stiffness affects the squeal propensity, it is helpful to plot the stability boundaries over a range of lining stiffness. The critical values of μ that are solved from a full complex eigenvalue analysis of the system equations are called "exact" solutions here. These are shown as dashed curves in Fig. 9. The regions above the curves are unstable since the corresponding friction coefficients are greater than μ_{cr}. The critical values of μ estimated by using the reduced-order characteristic equation with $M=5$ are shown as solid curves in Fig 9. The overall stability

Table 1 Mode coupling characteristics of different pairs of modes for the drum brake model in Ref. [10]

Mode pair	$E_{\mu sr}$	$E_{\mu rs}$	Coupling	Merging
27, 28	4.2e7	9.6e7	Veering away	No
29, 30	0.16	1.13	No coupling	No
29, 32	2.0e7	−7.1e6	Veering towards	Yes
30, 31	−3.3e7	4.6e7	Veering towards	Yes
31, 32	0.79	−0.11	No coupling	No

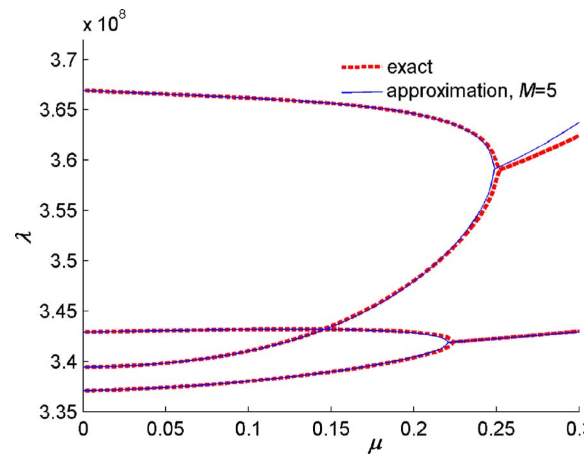


Fig. 8 Estimate of μ_{cr} by the reduced-order characteristic equation for the drum brake model in Ref. [10] with increased lining stiffness. $k=1.3$.

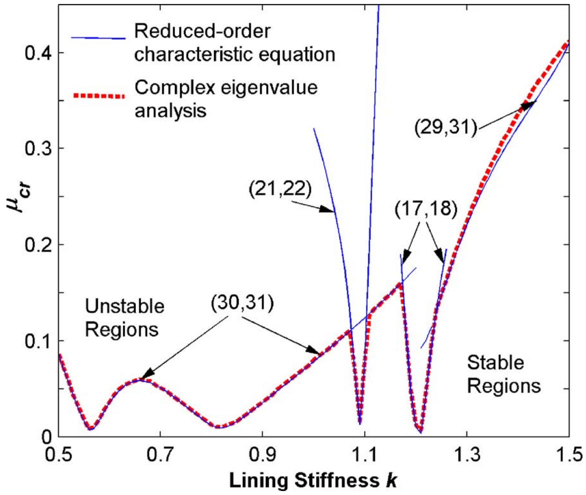


Fig. 9 Exact and estimated stability boundaries as a function of the lining stiffness for the drum brake system in Ref. [10]. The stability boundaries are for different combinations of modes that merge.

boundary is made up of several segments corresponding to different pairs of modes merging to cause squeal. The estimated curves of μ_{cr} over the lining stiffness range closely match the exact curves. Generally the results may be improved with a larger M. However, this reduced-order characteristic equation method gives satisfactory estimates to the frequency loci and μ_{cr} with a relatively small M.

To further illustrate the value of this approach for estimating the stability boundaries, the estimated μ_{cr} using Eq. (36) and exact μ_{cr} from the complex eigenvalue analysis for a few sample values of lining stiffness are listed in Table 2. Since Eq. (36) neglects the influences of higher order derivatives, this simple expression works well if the influence of higher order derivatives on the merging is small.

6 Conclusions

Brake squeal is often analytically studied by either a complex eigenvalue analysis of linearized models of the brake assembly or by a transient analysis of the nonlinear system. In this work, a new method called the reduced-order characteristic equation method is presented based on the eigenvalues and their derivatives at $\mu=0$. The method is then used to predict the onset of squeal. To evaluate the partial derivatives of eigenvalues and eigenvectors, a modal expansion method is utilized. The following conclusions can be drawn from the study.

(1) The eigenvalue derivatives can be used to study mode couplings between the various modes of the linear brake model. Generally, for a pair of close eigenvalues, two modes showing curve crossing have no coupling, and two modes showing curve veering away have coupling but it does not lead to coupled-mode instability. The modes showing curve veering towards are very likely to merge to cause squeal if there is sufficient friction coupling. Eigenvector sensitivity clearly shows the difference between eigenvalue loci crossing and loci veering.

(2) The reduced-order characteristic equation gives very good estimates to μ_{cr} without requiring complete solution of the full complex eigenvalue problem. This method can accurately predict the stability boundaries. It can be used to study the influences of frequency separation and eigenvalue derivatives at $\mu=0$ on the occurrence of brake squeal. When the changes in frequencies and

Table 2 A comparison of estimated and exact μ_{cr} for the drum brake model in Ref. [10]. The estimates are based on the reduced-order characteristic equation method where as the exact value is based on the complex eigenvalue analysis of the full drum brake model.

Lining stiffness	Estimated μ_{cr}	Exact μ_{cr}
0.7	0.051	0.051
0.8	0.011	0.012
0.9	0.035	0.036
1.0	0.080	0.079
1.1	0.133	0.124
1.2	0.202	0.176

their derivatives due to variation of system parameters are determined at $\mu=0$, μ_{cr} can be estimated without constructing new system models. This greatly facilitates the sensitivity analysis of system stability with respect to physical design parameters.

(3) A simple explicit expression of μ_{cr} involving eigenvalues and derivatives at $\mu=0$ up to the second order clearly shows how eigenvalue separation and friction coupling influence the magnitudes of μ_{cr}.

Acknowledgment

The first stage of this work was supported by Isuzu Motor Company, Ltd. The authors are also grateful for the support provided by Purdue Research Foundation and to the reviewers for a careful reading of the manuscript. The suggestions in reviews very much helped clarify the discussion in the paper.

References

[1] Liles, G. D., 1989, "Analysis of Disc Brake Squeal Using Finite Element Methods," SAE Paper No. 891150.

[2] Guan, D., and Jiang, D., 1998, "A Study on Disc Brake Squeal Using Finite Element Methods," SAE Paper No. 980597.

[3] Chowdhary, H. V., Bajaj, A. K., and Krousgrill, C. M., 2001, "An Analytical Approach to Model Disc Brake System for Squeal Prediction," *Proceedings of the ASME Design Engineering Technical Conferences*, DETC 2001/VIB-21560.

[4] Servis, B. K., 2000, "The Onset of Squeal Vibrations in Drum Brake Systems Resulting From a Coupled Mode Instability," Ph.D. dissertation, Purdue University, West Lafayette, IN.

[5] Hamabe, T., Yamazaki, I., Yamada, K., Matsui, H., Nakagawa, S., and Kawamura, M., 1999, "Study of a Method for Reducing Drum Brake Squeal," SAE Paper No. 1999–01–0144.

[6] Nack, W. V., 2000, "Brake Squeal Analysis by Finite Element," Int. J. Veh. Des., **23**(3/4), pp. 263–275.

[7] Kinkaid, N. M., O'Reilly, O. M., and Papadopoulos, P., 2003, "Automotive Disc Brake Squeal: A Review," J. Sound Vib., **267**, pp. 105–166.

[8] Chung, C., Steed, W., Kobayashi, K., and Nakata, H., 2001, "A New Analysis Method for Brake Squeal Part I: Theory for Modal Domain Formulation and Stability Analysis," SAE Paper No. 2001–01–1600.

[9] Perkins, N. C., and Mote, C. D., Jr., 1986, "Comments on Curve Veering in Eigenvalue Problems," J. Sound Vib., **106**, pp. 451–463.

[10] Huang, J., Krousgrill, C. M., and Bajaj, A. K., 2006, "Modeling of Automotive Drum Brakes for Squeal and Parameter Sensitivity Analysis," J. Sound Vib., **289**(1–2), pp. 245–263.

[11] Blevins, R. D., 1990, *Flow-Induced Vibration*, Van Nostrand Reinhold, New York.

[12] Dowell, E. H., Crawley, E. F., Curtiss, H. C., Peters, D. A., Scanlan, R. H., and Sisto, F., 1995, *A Modern Course in Aeroelasticity*, Kluwer Academic, Dordrecht.

[13] Rudisill, C. S., and Chu, Y. Y., 1975, "Numerical Methods for Evaluating the Derivatives of Eigenvalues and Eigenvectors," AIAA J., **13**, pp. 834–837.

[14] Seyranian, A. P., 1993, "Sensitivity Analysis of Multiple Eigenvalues," Mech. Struct. Mach. **21**(2), pp. 261–284.

[15] Friswell, M. I., 1996, "The Derivatives of Repeated Eigenvalues and Their Associated Eigenvectors," ASME J. Vibr. Acoust., **118**, pp. 390–397.

[16] Huseyin, K., 1978, *Vibrations and Stability of Multiple Parameter Systems*, Noordhoff Publishing, Groningen.

[17] Seyranian, A. P., and Mailybaev, A. A., 2003, *Multiparameter Stability Theory With Mechanical Applications*, World Scientific, Singapore.

Friction Induced Vibrations in Moving Continua and Their Application to Brake Squeal

Daniel Hochlenert[1]
e-mail: hochlenert@dyn.tu-darmstadt.de

Gottfried Spelsberg-Korspeter
e-mail: speko@dyn.tu-darmstadt.de

Peter Hagedorn
e-mail: peter.hagedorn@dyn.tu-darmstadt.de

Department of Mechanical Engineering,
Dynamics and Vibrations Group,
Technische Universität Darmstadt,
Hochschulstrasse 1,
64289 Darmstadt, Germany

Considerable effort is spent in the design and testing of disk brake systems installed in modern passenger cars. This effort can be reduced if appropriate mathematical–mechanical models are used for studying the dynamics of these brakes. In this context, the mechanism generating brake squeal in particular deserves closer attention. The present paper is devoted to the modeling of self-excited vibrations of moving continua generated by frictional forces. Special regard is given to an accurate formulation of the kinematics of the frictional contact in two and three dimensions. On the basis of a travelling Euler–Bernoulli beam and a rotating annular Kirchhoff plate with frictional point contact the essential properties of the contact kinematics leading to self-excited vibrations are worked out. A Ritz discretization is applied and the obtained approximate solution is compared to the exact one of the traveling beam. A minimal disk brake model consisting of the discretized rotating Kirchhoff plate and idealized brake pads is analyzed with respect to its stability behavior resulting in traceable design proposals for a disk brake. [DOI: 10.1115/1.2424239]

Keywords: friction induced vibrations, moving media, brake squeal, minimal model

1 Introduction

Most of the commonly used brake systems in cars, trains, airplanes, and several industrial machines generate the brake force by friction. Noises possibly generated by the brake are usually unwanted; they may occur due to a transfer of the kinetic energy of the part being decelerated to vibrations of the brake system. It has become generally accepted by engineers and researchers working in the field of brake noise that brake squeal is generated by friction induced self-excited vibrations of the brake system. The noise-free configuration of the self-excited brake system loses its stability, the system then starts oscillating with audible frequencies and reaches a limit cycle.

The cause of the energy transfer and therefore the mechanism generating brake squeal has been put forward on different grounds. A broad overview of brake squeal and possible excitation mechanisms is given in Ref. [1]. A more general review of friction induced vibrations is the comprehensive review paper by Ibrahim [2]. The most popular explanations for brake squeal are the excitation due to nonconstant friction coefficients, especially a friction characteristic depending on the relative velocity between the contact points and flutter type instabilities due to nonconservative forces in the contact area which may even occur with constant parameters. The authors believe that the latter one is a more realistic cause of brake squeal even though it is known from experiments that the parameters of the system may vary with time, temperature, and wear. Given that an instability can be modeled with constant parameters, it is assumable that the instability may even occur in a real brake system with constant friction coefficient once the other parameters reach the unstable region and vary only slowly compared to the oscillation period of the system.

Commonly used disk brake models, i.e., multi-body systems and finite element models (FEM), have high numbers of degrees of freedom. However one of the problems in modeling brake squeal is the basic formulation of the contact mechanism, in particular also of its kinematics, and its integration into the models. There are several papers describing instability mechanisms due to frictional contact involving vibrations of continuous systems. Ouyang and Mottershead [3] studied the dynamic instability of an elastic disk under the action of a rotating friction couple and proposed the model to demonstrate a mechanism for unstable vibrations of a car disk brake. The main restrictions of this model are the simplified kinematics at the contact point and the exclusion of friction forces acting in the radial direction of the disk. Ono et al. [4] analyzed a head–disk interface in a flexible disk drive and considered friction at the contact points. Since this work is focused on disk storage systems, the considered parameters and some assumptions are far from the corresponding ones of a disk brake. Furthermore, the kinematics of the disk and the evaluation of the friction force are simplified as well. The transient dynamics of a multi-degree of freedom friction oscillator and the application to disk brake squeal was studied by Kinkaid et al. [5]. They modeled the brake disk as a rigid body sliding under a brake pad and approximated the kinematics by neglecting any tilting motion of the disk.

In Ref. [6] the authors gave a review of minimal models for the explanation of disk brake squeal and introduced a new two degree of freedom disk brake model which can easily be associated to a real brake, to work out the details of an excitation mechanism. This model represents the brake rotor as a rigid wobbling disk. The intention of the present paper is to replace the wobbling rigid disk by a rotating annular Kirchhoff plate without making any simplifying assumptions concerning the kinematics of the frictional contact, except for the assumption of point contact (this model will be generalized in a later publication to the case of distributed contact).

An axially moving beam under frictional point contact is first considered as a preliminary two-dimensional problem, to work out the essential properties of the contact kinematics leading to self-excited vibrations. Furthermore, the model will be used to introduce a discretization scheme which will be applied to the rotating Kirchhoff plate later on. Additionally, by analyzing the equations of motion of the beam and the plate it becomes clear that a two-dimensional consideration of the kinematics of a disk

[1]Corresponding author.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received September 14, 2005; final manuscript received June 22, 2006. Review conducted by Oliver M. O'Reilly.

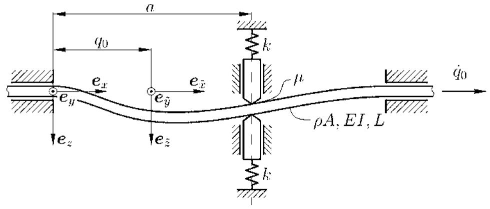


Fig. 1 Axially moving beam

brake is not adequate, i.e., the three-dimensional contact kinematics yield dissipative terms that cannot be observed in the existing two-dimensional models.

2 Axially Moving Beam

In this section we describe the modeling of an axially moving beam as depicted in Fig. 1. The beam is moving with a given velocity $\dot{q}_0(t)$ through two idealized massless "brake pads" supported by two prestressed linear springs (spring constant k, prestress force N_0). The brake pads are constrained to move in the vertical direction only. Coulomb friction occurs between the beam and the brake pads. The traveling speed of the beam is assumed to be large enough to avoid stick–slip phenomena. We discuss a Ritz discretization method which will later be used in analogous form for the rotating Kirchhoff plate. As the convergence of the Ritz method is not obvious for systems with gyroscopic terms, such as the axially moving beam or the rotating plate, a comparison of the approximate solution with an exact solution seems useful. The cartesian reference frame $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ (coordinates x,y,z) is inertially fixed and the frame $\boldsymbol{e}_{\tilde{x}}$, $\boldsymbol{e}_{\tilde{y}}$, $\boldsymbol{e}_{\tilde{z}}$ (coordinates $\tilde{x},\tilde{y},\tilde{z}$) moves axially with velocity $\dot{q}_0$. The transverse displacement of points on the neutral fiber of the beam is denoted by $w(x,t)$ in the inertial frame and by $\tilde{w}(\tilde{x},t)$ in the moving frame. The relation $x=\tilde{x}+q_0$ holds and the following identities can be established easily

$$w(x,t) = \tilde{w}(\tilde{x},t) \tag{1}$$

$$\frac{\partial}{\partial \tilde{x}}\tilde{w}(\tilde{x},t) = \frac{\partial}{\partial x}w(x,t) \tag{2}$$

$$\frac{d}{dt}\tilde{w}(\tilde{x},t) = \frac{\partial}{\partial t}w(x,t) + \dot{q}_0\frac{\partial}{\partial x}w(x,t) = \dot{w}(x,t) + \dot{q}_0 w'(x,t) \tag{3}$$

where the partial derivatives are abbreviated by $\partial/\partial x = ()'$ and $\partial/\partial t = (\dot{})$. The neutral fiber of the beam is parameterized by

$$f(x,z,t) = z - w(x,t) = 0 \tag{4}$$

of which we frequently use the gradient

$$\nabla f(x,z,t) = -\frac{\partial}{\partial x}w(x,t)\boldsymbol{e}_x + \boldsymbol{e}_z \tag{5}$$

and the unit vector in its direction

$$\boldsymbol{e}_\nabla(x,t) = \frac{\nabla f(x,z,t)}{|\nabla f(x,z,t)|} \tag{6}$$

The equations of motion can be derived from Hamilton's principle

$$\delta\int_{t_1}^{t_2}(T-U)dt = -\int_{t_1}^{t_2}\delta W\, dt \tag{7}$$

where

$$\delta W = \int_0^L \boldsymbol{F}\cdot\delta\boldsymbol{p}\, dx \tag{8}$$

is the virtual work of all forces not considered in U. Note that in order to obtain the correct linear equations of motion, the position vectors $\boldsymbol{p}$ to the points of actions of the forces $\boldsymbol{F}$ have to be exact up to second order in w and its derivatives. The kinetic and the potential energy of the beam are given by

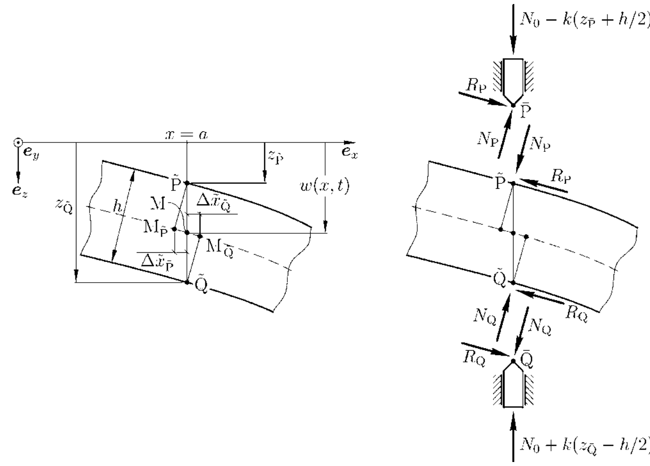


Fig. 2 Kinematics and forces acting on the beam and on the pads

$$T = \frac{1}{2}\rho A\int_0^L \boldsymbol{v}_M(x,t)^2\, dx \tag{9}$$

and

$$U = \frac{1}{2}EI\int_0^L w''(x,t)^2\, dx \tag{10}$$

where $\boldsymbol{v}_M$ is the velocity of a material point on the beam's neutral fiber.

2.1 Kinematics. Figure 2 shows the kinematics of the axially moving beam in the region of the brake pads with the forces acting on the beam and on the pads. In the following we will derive the kinematics of the beam in a rather formal way. Even though some geometric relations are intuitively determinable and evident from the figures, the formal procedure is valuable since it will be reused for the more complex kinematics of the rotating plate later on.

The position vector from the origin O of the inertial system to an arbitrary material point M on the neutral fiber of the beam is given by

$$\boldsymbol{p}_M = \overline{OM} = x\boldsymbol{e}_x + \tilde{w}(\tilde{x},t)\boldsymbol{e}_z \tag{11}$$

Therefore the velocity of M is the derivative of $\boldsymbol{p}_M$ with respect to time in the inertial coordinate system

$$\boldsymbol{v}_M = \frac{{}^N d}{dt}\boldsymbol{p}_M = \dot{q}_0\boldsymbol{e}_x + \frac{d}{dt}\tilde{w}(\tilde{x},t)\boldsymbol{e}_z \tag{12}$$

where the superscript N denotes differentiation in the Newtonian frame.

To calculate the virtual work Eq. (8) we need the position vectors of the contact points and the contact forces. We concentrate on the upper contact point P between beam and brake pad. The point $\tilde{P}$ is the material point of the beam in actual contact with material point $\bar{P}$ of the brake pad. The notation and the calculations for the lower contact point Q are similar. The position vector from the origin O to $\bar{P}$ on the upper brake pad at $x=a$ is given by

$$\boldsymbol{p}_{\bar{P}} = \overline{O\bar{P}} = a\boldsymbol{e}_x + z_{\bar{P}}\boldsymbol{e}_z \tag{13}$$

and the position vector to the corresponding point $\tilde{P}$ of the beam being in actual contact with the pad is

$$\boldsymbol{p}_{\bar{P}} = \overline{O\tilde{P}} = (a + \Delta\tilde{x}_{\tilde{P}})\boldsymbol{e}_x + \tilde{w}(a + \Delta\tilde{x}_{\tilde{P}},t)\boldsymbol{e}_z - \frac{h}{2}\boldsymbol{e}_\nabla(a + \Delta\tilde{x}_{\tilde{P}},t) \tag{14}$$

Setting $\boldsymbol{p}_{\bar{P}}=\boldsymbol{p}_{\tilde{P}}$ and solving for the two unknowns $\Delta\tilde{x}_{\tilde{P}}$ and $z_{\bar{P}}$, yields

$$\Delta\tilde{x}_{\tilde{P}} = -\frac{h}{2}w'(a,t) + \mathcal{O}(w^2) \tag{15}$$

$$z_{\bar{P}} = -\frac{h}{2} + w(a,t) - \frac{h}{4}w'^2(a,t) + \mathcal{O}(w^3) \tag{16}$$

where $\mathcal{O}(w^n)$ contains all terms which are of order n or greater in w and its derivatives. Using the fact that in Euler–Bernoulli theory planar cross sections of the beam remain planar and orthogonal to the neutral axis the velocity of $\tilde{P}$ is consequently given by

$$\boldsymbol{v}_{\tilde{P}} = \boldsymbol{v}_M + \Delta\tilde{x}_{\tilde{P}}\frac{\partial}{\partial x}\boldsymbol{v}_M - \frac{h}{2}\left[\frac{{}^N d}{dt}\boldsymbol{e}_\nabla(a,t) + \Delta\tilde{x}_{\tilde{P}}\frac{\partial}{\partial x}\frac{{}^N d}{dt}\boldsymbol{e}_\nabla(a,t)\right] + \mathcal{O}(w^3) \tag{17}$$

The velocity of the material point of the pad $\boldsymbol{v}_{\bar{P}}$ simply is

$$\boldsymbol{v}_{\bar{P}} = \frac{{}^N d}{dt}\boldsymbol{p}_{\bar{P}} = \left[\dot{w}(a,t) - \frac{h}{2}\dot{w}'(a,t)\dot{w}(a,t)\right]\boldsymbol{e}_z \tag{18}$$

Analogously for Q we determine $z_{\bar{Q}}$, $\Delta x_{\tilde{Q}}$, $\boldsymbol{v}_{\tilde{Q}}$, and $\boldsymbol{v}_{\bar{Q}}$.

2.2 Contact Forces. The forces at the contact point P can be determined from a force balance at the brake pad and Coulomb's law of friction

$$\boldsymbol{R}_{\bar{P}} = -\boldsymbol{R}_{\tilde{P}} = \mu|\boldsymbol{N}_{\bar{P}}|\boldsymbol{e}_{R_{\bar{P}}} \tag{19}$$

where the direction of the friction force is given by the unit vector

$$\boldsymbol{e}_{R_{\bar{P}}} = \frac{\boldsymbol{v}_{\tilde{P}} - \boldsymbol{v}_{\bar{P}}}{|\boldsymbol{v}_{\tilde{P}} - \boldsymbol{v}_{\bar{P}}|} \tag{20}$$

The normal forces are perpendicular to the surface of the beam so that

$$\boldsymbol{e}_{N_{\bar{P}}} = \boldsymbol{e}_\nabla(a + \Delta\tilde{x}_{\tilde{P}},t) \tag{21}$$

is the unit vector in direction of the normal force. Note that the forces N_P and R_P are defined as the normal and tangential component of the contact force in the deformed configuration and consequently read

$$N_P = |\boldsymbol{N}_{\bar{P}}| = N_0 - kw(a,t) + \mu N_0 w'(a,t) + \mathcal{O}(w^2) \tag{22}$$

$$R_P = |\boldsymbol{R}_{\bar{P}}| = \mu N_P \tag{23}$$

The corresponding forces acting on $\tilde{Q}$ can be determined analogously. The expressions containing $\boldsymbol{e}_{R_{\bar{P}}}$ can only be linearized if $\dot{q}_0$ is sufficiently large, i.e., $h/2|\dot{w}'(a,t)| \ll \dot{q}_0$ is assumed in the following. This can be seen from the linearized relative velocity between the pad and the traveling beam at the contact point

$$\boldsymbol{v}_{\bar{P}} - \boldsymbol{v}_{\tilde{P}} = -\left[\dot{q}_0 + \frac{h}{2}\dot{w}'(a,t)\right]\boldsymbol{e}_x - \dot{q}_0 w'(a,t)\boldsymbol{e}_z \tag{24}$$

This assumption also implies that no stick–slip occurs since

$$|\dot{w}'(a)|\frac{h}{2} < \dot{q}_0 \tag{25}$$

ensures that the relative velocity between the pad and the beam does not vanish.

2.3 Solution. In this section we give a solution for the problem using a Ritz discretization approach. To get an idea of the speed of convergence, we first solve the problem for the traveling beam omitting the brake pads. In this case for $\dot{q}_0$=const the problem can be solved in closed form (see for example Ref. [7] for a spectral or Ref. [8] for a wave propagation approach). This exact solution will be used to check the accuracy of the approximate solution given below, to gain confidence in the convergence of the method for gyroscopic systems.

We now introduce a method to determine the eigenvalues and eigenfunctions approximately. We substitute the expression

$$w(x,t) = \sum_{i=1}^{I} W_i(x)q_i(t) = \sum_{i=1}^{I} W_i q_i \tag{26}$$

with suitably assumed shape functions $W_i(x)$ into Eq. (7), obtaining discretized expressions for the kinetic and potential energy on the left hand side of Eq. (7). The discretized equations of motion are obtained from Lagrange's equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = f_j = \sum_k \boldsymbol{F}_k \cdot \frac{\partial \boldsymbol{v}_k}{\partial \dot{q}_j} \quad j = 1, \ldots, I \tag{27}$$

where $L=T-U$ is the Lagrangian. The complete linearized equations of motion read

$$\begin{aligned}\sum_{i=1}^{I}\Bigg\{&\left(\rho A\int_0^L W_j W_i\,dx\right)\ddot{q}_i + \left[\dot{q}_0\rho A\int_0^L (W_j W_i' - W_j' W_i)dx\right]\dot{q}_i \\ &+ \left[\int_0^L (EIW_j'' W_i'' - \dot{q}_0^2\rho A W_j' W_i')dx\right]q_i \\ &= (\boldsymbol{N}_{\bar{P}} + \boldsymbol{R}_{\bar{P}})\cdot\frac{\partial \boldsymbol{v}_{\tilde{P}}}{\partial \dot{q}_j} + (\boldsymbol{N}_{\bar{Q}} + \boldsymbol{R}_{\bar{Q}})\cdot\frac{\partial \boldsymbol{v}_{\tilde{Q}}}{\partial \dot{q}_j}\end{aligned} \tag{28}$$

which is a system of equations of the form

$$\boldsymbol{M}\ddot{\boldsymbol{q}} + \boldsymbol{G}\dot{\boldsymbol{q}} + \boldsymbol{K}\boldsymbol{q} = \boldsymbol{0} \tag{29}$$

where the elements of the matrices are

$$m_{ji} = \rho A\int_0^L W_j W_i\,dx \tag{30}$$

$$g_{ji} = \rho A\dot{q}_0\int_0^L (W_j W_i' - W_j' W_i)dx \tag{31}$$

$$\begin{aligned}k_{ji} = &-\rho A\dot{q}_0^2\int_0^L W_j' W_i'\,dx + EI\int_0^L W_j'' W_i''\,dx + 2kW_j(a)W_i(a) \\ &- hk\mu W_j'(a)W_i(a) + hN_0(1+\mu^2)W_j'(a)W_i'(a) \\ &- \frac{h^2 N_0\mu}{2}W_j''(a)W_i'(a)\end{aligned} \tag{32}$$

As will be seen later on, the terms in the matrices of the beam's and plate's equation of motion are quite similar. In order to compare the results with the rotating Kirchhoff plate later on we carry out the calculations using the first two eigenfunctions of the corresponding non moving clamped beam as shape functions. By choosing $a=L/2$ we get

$$\boldsymbol{M} = \rho A\begin{bmatrix}\int_0^L W_1^2\,dx & 0 \\ 0 & \int_0^L W_2^2\,dx\end{bmatrix} \tag{33}$$

$$
\boldsymbol{G} = \rho A \dot{q}_0 \begin{bmatrix} 0 & -\int_0^L (W_2 W_1' - W_2' W_1) dx \\ \int_0^L (W_2 W_1' - W_2' W_1) dx & 0 \end{bmatrix} \tag{34}
$$

$$
\boldsymbol{K} = \begin{bmatrix} -\rho A \dot{q}_0^2 \int_0^L W_1'^2\, dx + EI \int_0^L W_1''^2\, dx + 2k W_1^2(a) & \cdots \\ -hk\mu W_2'(a) W_1(a) - \dfrac{h^2 N_0 \mu}{2} W_1''(a) W_2'(a) & \\ \cdots & -\rho A \dot{q}_0^2 \int_0^L W_2'^2\, dx + EI \int_0^L W_2''^2\, dx + hN_0(1+\mu^2) W_2'^2(a) \end{bmatrix} \tag{35}
$$

since $W_1'(L/2)=0$, $W_2(L/2)=0$, and $W_2''(L/2)=0$. In Eq. (29) the matrix $\boldsymbol{M}$ is symmetric, $\boldsymbol{G}$ is skew symmetric, and $\boldsymbol{K}$ contains asymmetric coupling terms, making the system susceptible to self-excited vibrations.

2.4 Stability Analysis. In order to study the stability of the trivial solution of Eq. (29) we substitute $\boldsymbol{q}(t)=\hat{\boldsymbol{q}}e^{\lambda t}$ and obtain the characteristic equation

$$
\det(\boldsymbol{M}\lambda^2 + \boldsymbol{G}\lambda + \boldsymbol{K}) = 0 \tag{36}
$$

From the expansion theorem of the determinant it follows that the coefficient a_{2I-1} of the characteristic polynomial

$$
a_{2I}\lambda^{2I} + a_{2I-1}\lambda^{2I-1} + \cdots + a_2\lambda^2 + a_1\lambda + a_0 = 0 \tag{37}
$$

vanishes. This can be seen by considering that a_{2I} originates from the product of the terms on the main diagonal of $(\boldsymbol{M}\lambda^2+\boldsymbol{G}\lambda+\boldsymbol{K})$. Since the main diagonal of $\boldsymbol{G}$ consists of zeros the second nonzero coefficient of Eq. (37) is a_{2I-2}. Thus the Routh–Hurwitz criterion states that the trivial solution of the system cannot be asymptotically stable and that not all of the eigenvalues of Eq. (29) can have a negative real part. Provided that not all eigenvalues are purely imaginary, we now show that the trivial solution is unstable. This conclusion follows from a theorem proved by Karapetjan [9] and Lakhadanov [10] independently. The ideas of the proof relevant for the present problem will be stated in the following.

In the case where Eq. (37) has no purely imaginary roots, there has to be at least one root with a positive real part. If there is a purely imaginary root, Eq. (37) can be written as

$$
(\lambda^2 + \alpha)(b_{2I-2}\lambda^{2I-2} + b_{2I-3}\lambda^{2I-3} + \ldots + b_0) = 0 \tag{38}
$$

where $\alpha>0$ and $b_{2I-3}=0$. Now the same investigation can be performed for

$$
b_{2I-2}\lambda^{2I-2} + b_{2I-3}\lambda^{2I-3} + \ldots + b_0 = 0 \tag{39}
$$

Following this argument in an inductive manner and factoring out all purely imaginary roots, we see from the Routh–Hurwitz criterion that the remaining polynomial has to have a root with positive real part, i.e., the trivial solution is unstable.

A sufficient condition that the characteristic equation cannot have only purely imaginary roots is a nonvanishing coefficient corresponding to an odd power of λ. For the case $I=2$ the coefficients of the characteristic polynomial are

$$
a_0 = k_{11}k_{22} - k_{12}k_{21} \tag{40}
$$

$$
a_1 = (k_{21} - k_{12})g_{21} \tag{41}
$$

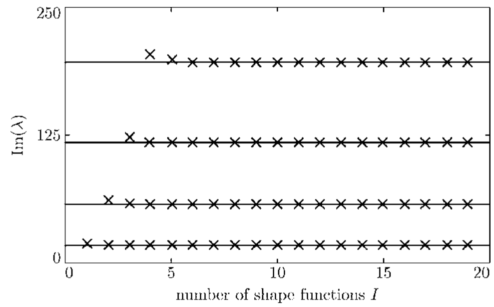


Fig. 3 Comparison of the exact (lines) and the approximate solution (×)

$$
a_2 = m_{11}k_{22} + m_{22}k_{11} \tag{42}
$$

$$
a_3 = 0 \tag{43}
$$

$$
a_4 = m_{11}m_{22} \tag{44}
$$

where a_1 is nonzero when $\boldsymbol{K}$ is asymmetric. We now readily conclude for $I=2$ that the trivial solution of Eq. (29) is unstable for any given parameter combination μ, h, N_0, $k>0$, requiring only that $\boldsymbol{K}$ is asymmetric. Using more shape functions $(I>2)$ the trivial solution is unstable except for particular parameter combinations restricted by the condition $a_i=0$ for all odd i. This holds in general for circulatory gyroscopic systems without damping. Note that the absence of damping is an essential point in the argumentation. This is the reason for the essential differences in the stability behavior in comparison with the rotating Kirchhoff plate investigated in Sec. 3.

2.5 Comparison of the Exact and the Discretized Solution for the Traveling Beam Without Brake Pads. For both approaches we introduce the dimensionless time $\tau=\omega t$ and length $\bar{x}=x/L$ where $\omega^2=EI/\rho AL^4$ and choose the dimensionless parameter $\bar{v}=\dot{q}_0/L\omega=4$. The discretized equations of motion for the traveling beam are obtained from Eq. (28) setting the right-hand side equal to zero. They are solved using $\boldsymbol{q}(\tau)=\boldsymbol{q}e^{\lambda\tau}$. The exact solution is calculated as given in Ref. [7], solving the boundary value problem

$$
\ddot{w} + 2\bar{v}\dot{w}' + \bar{v}^2 w'' + w^{\mathrm{IV}} = 0 \tag{45}
$$

with boundary conditions

$$
w(0,\tau) = 0, \quad w'(0,\tau) = 0
$$

$$
w(1,\tau) = 0, \quad w'(1,\tau) = 0 \tag{46}
$$

and using

$$
w(\bar{x},\tau) = e^{\beta\bar{x}}e^{\lambda\tau} \tag{47}
$$

Figure 3 shows a plot of the eigenvalues obtained from the discretized method over the number of shape functions used. Obviously the eigenvalues are purely imaginary hence the real part is not shown. The first frequencies already match closely if only a few shape functions are used. Also when varying the dimensionless speed of the beam from $\bar{v}=0$ to $\bar{v}=2\pi$ (the critical speed, Ref. [7]) the eigenvalues of the exact and the discretized approach still coincide, as can be seen from Fig. 4, where only $I=6$ shape functions are used for the discretized solution. Summarizing, the method converges rapidly for the moving beam and due to the similar structure, the same can be expected for the rotating plate.

3 Rotating Kirchhoff Plate

As mentioned above, the intention is to work out a possible excitation mechanism of brake squeal with a simple model, which still has an obvious relation to a technical disk brake system. In Ref. [6] the authors presented a minimal model for the explana-

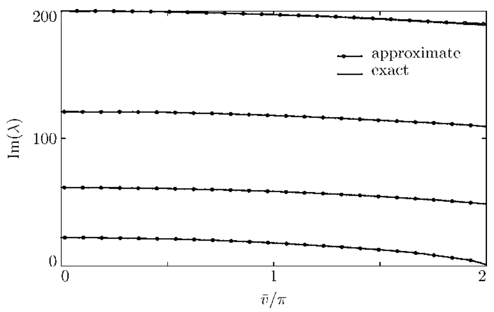


Fig. 4 Comparison of the exact and the approximate solution for varying $\bar{v}$

tion of brake squeal with the brake disk modeled as a wobbling rigid body and showed that the self-excitation mechanism needs at least two degrees of freedom. As mentioned in Ref. [6], a possible and obvious extension of the model is to replace the wobbling disk by a rotating Kirchhoff plate discretized in an appropriate way.

Figure 5 shows a rotating annular Kirchhoff plate in contact with two idealized massless brake pads. The plate is driven by a torque $M_A \boldsymbol{e}_z$, such that frame $\boldsymbol{e}_{\tilde{x}}$, $\boldsymbol{e}_{\tilde{y}}$, $\boldsymbol{e}_{\tilde{z}}$ (rigidly attached to the nondeformed plate) rotates with the angular velocity $\dot{q}_0 \boldsymbol{e}_z$ relatively to the inertial frame $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$. Again, the brake pads are supported by two prestressed linear springs (prestress N_0, spring constant k) and Coulomb friction occurs between the disk and the brake pads. The displacement of points on the neutral plane can be described by $\tilde{w}(r,\alpha,t)$ with independent variables r, α, t and by $w(r,\varphi,t)$ with independent variables r, φ, t. The relation $\varphi=\alpha+q_0$ connects both representations. The identities

$$w(r,\varphi,t) = \tilde{w}(r,\alpha,t) \tag{48}$$

and

$$\frac{\partial}{\partial r}\tilde{w}(r,\alpha,t) = \frac{\partial}{\partial r}w(r,\varphi,t)$$

$$\frac{\partial}{\partial \alpha}\tilde{w}(r,\alpha,t) = \frac{\partial}{\partial \varphi}w(r,\varphi,t)$$

$$\frac{d}{dt}\tilde{w}(r,\alpha,t) = \frac{\partial}{\partial t}w(r,\varphi,t) + \dot{q}_0\frac{\partial}{\partial \varphi}w(r,\varphi,t)$$

will be used to transform appearing expressions from material to spatial coordinates. The neutral plane of the plate is parameterized by the function $f(r,\alpha,z,t)=z-\tilde{w}(r,\alpha,t)=0$, of which we will use the gradient

$$\nabla f(r,\alpha,z,t) = -\frac{\partial}{\partial r}\tilde{w}(r,\alpha,t)\boldsymbol{e}_r - \frac{1}{r}\frac{\partial}{\partial \alpha}\tilde{w}(r,\alpha,t)\boldsymbol{e}_\varphi + \boldsymbol{e}_z \tag{49}$$

and the unit vector in its direction

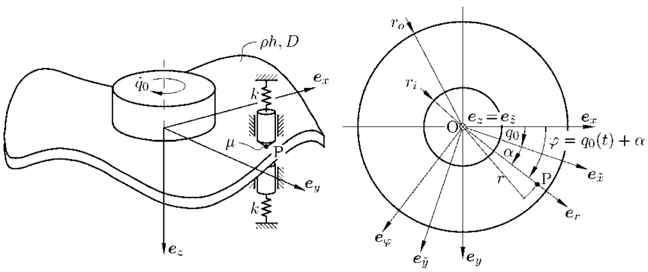


Fig. 5 Rotating Kirchhoff plate: model (left) and top view (right)

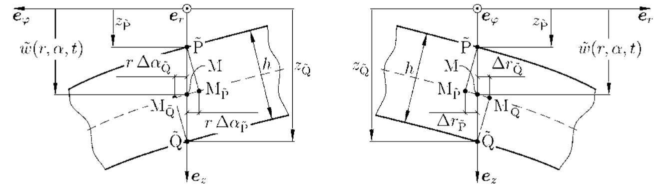


Fig. 6 Kinematics of the plate (section in $\boldsymbol{e}_\varphi$-$\boldsymbol{e}_z$- and $\boldsymbol{e}_r$-$\boldsymbol{e}_z$-plane)

$$\boldsymbol{e}_\nabla(r,\alpha,t) = \frac{\nabla f(r,\alpha,z,t)}{|\nabla f(r,\alpha,z,t)|} \tag{50}$$

We now systematically derive the equations of motion using Eq. (7). Therefore we need the potential and kinetic energy of the rotating plate

$$U = \frac{1}{2}D\int_A (\nabla^2\tilde{w})^2 - 2(1-\nu)\left[\frac{\partial^2\tilde{w}}{\partial r^2}\left(\frac{1}{r}\frac{\partial\tilde{w}}{\partial r} + \frac{1}{r^2}\frac{\partial^2\tilde{w}}{\partial\alpha^2}\right) - \left(\frac{1}{r}\frac{\partial^2\tilde{w}}{\partial r\,\partial\alpha} - \frac{1}{r^2}\frac{\partial\tilde{w}}{\partial\alpha}\right)^2\right]dA \tag{51}$$

and

$$T = \frac{1}{2}\rho h\int_A \boldsymbol{v}_M^2\, dA \tag{52}$$

where $\boldsymbol{v}_M$ is the velocity of a material point on the plate's neutral plane [11]. In order to calculate the virtual work Eq. (8), which now reads

$$\delta W = \int_A \boldsymbol{F}\cdot\delta\boldsymbol{p}\, dA \tag{53}$$

the virtual displacements $\delta\boldsymbol{p}$ of the upper and lower contact point and the forces acting at these points will be determined in the following subsections.

3.1 Kinematics. The kinematics of the rotating Kirchhoff plate shown in Fig. 6 is basically a three-dimensional version of the beam's kinematics. The vector from the origin O to a point M on the neutral plane is given by

$$\boldsymbol{p}_M = \overline{OM} = r\boldsymbol{e}_r + \tilde{w}(r,\alpha,t)\boldsymbol{e}_z \tag{54}$$

and the velocity of M consequently is

$$\boldsymbol{v}_M = \frac{{}^N d}{dt}\boldsymbol{p}_M \tag{55}$$

$$= \dot{q}_0 r\boldsymbol{e}_\varphi + \frac{d}{dt}\tilde{w}(r,\alpha,t)\boldsymbol{e}_z \tag{56}$$

Similarly as for the axially moving beam, the calculation of the virtual work Eq. (53) requires the determination of the vectors $\boldsymbol{p}$ up to second order. The following calculations are exemplarily shown for the contact point P between the disk and the brake pad. In our notation $\tilde{P}$ is the material point of the disk in actual contact with the point $\bar{P}$ of the brake pad. The vector pointing from the origin O to $\bar{P}$ at $r=r_0$ can be expressed as

$$\boldsymbol{p}_{\bar{P}} = \overline{O\bar{P}} = r_0\boldsymbol{e}_r + z_{\bar{P}}\boldsymbol{e}_z \tag{57}$$

The position vector from the origin to the material point $\tilde{P}$ coinciding with $\bar{P}$ at a certain instant

$$\boldsymbol{p}_{\tilde{P}} = \overline{O\tilde{P}} = (r+\Delta r_{\tilde{P}})\boldsymbol{e}_r + r\Delta\alpha_{\tilde{P}}\boldsymbol{e}_\varphi + \tilde{w}(r+\Delta r_{\tilde{P}}, \alpha+\Delta\alpha_{\tilde{P}}, t)\boldsymbol{e}_z - \frac{h}{2}\boldsymbol{e}_{\nabla_{\tilde{P}}} \tag{58}$$

where

$$\boldsymbol{e}_{\nabla_{\tilde{P}}} = \boldsymbol{e}_\nabla(r+\Delta r_{\tilde{P}}, \alpha+\Delta\alpha_{\tilde{P}}, t) \tag{59}$$

is the unit vector normal to deformed neutral plane at $M_{\tilde{P}}$. Setting $\boldsymbol{p}_{\bar{P}}=\boldsymbol{p}_{\tilde{P}}$ and carrying out a Taylor expansion for small $\Delta r_{\tilde{P}}$ and $\Delta\alpha_{\tilde{P}}$ up to first-order terms, yields three linear equations for the determination of $\Delta r_{\tilde{P}}$, $\Delta\alpha_{\tilde{P}}$, and $z_{\bar{P}}$. Note that the unit vectors $\boldsymbol{e}_r$ and $\boldsymbol{e}_\varphi$ at $M_{\tilde{P}}$ differ from those at M and have to be expanded in a Taylor series as well. A second Taylor expansion of the results in $\tilde{w}(r,\alpha,t)$ and its derivatives yields

$$z_{\bar{P}} = -\frac{h}{2} + \tilde{w}(r,\alpha,t) - \frac{h}{4r^2}\left[\frac{\partial}{\partial\alpha}\tilde{w}(r,\alpha,t)\right]^2 - \frac{h}{4}\left[\frac{\partial}{\partial r}\tilde{w}(r,\alpha,t)\right]^2 + \mathcal{O}(\tilde{w}^3) \tag{60}$$

$$\Delta r_{\tilde{P}} = -\frac{h}{2}\frac{\partial}{\partial r}\tilde{w}(r,\alpha,t) + \mathcal{O}(\tilde{w}^2) \tag{61}$$

$$\Delta\alpha_{\tilde{P}} = -\frac{h}{2r^2}\frac{\partial}{\partial\alpha}\tilde{w}(r,\alpha,t) + \mathcal{O}(\tilde{w}^2) \tag{62}$$

The velocities of the points $\tilde{P}$ and $\bar{P}$ are calculated from

$$\boldsymbol{v}_{\tilde{P}} = \boldsymbol{v}_M + \Delta r_{\tilde{P}}\frac{\partial}{\partial r}\boldsymbol{v}_M + \Delta\alpha_{\tilde{P}}\frac{\partial}{\partial\alpha}\boldsymbol{v}_M - \frac{h}{2}\left[\frac{{}^N d}{dt}\boldsymbol{e}_\nabla(r,\alpha,t) + \Delta r_{\tilde{P}}\frac{\partial}{\partial r}\frac{{}^N d}{dt}\boldsymbol{e}_\nabla(r,\alpha,t) + \Delta\alpha_{\tilde{P}}\frac{\partial}{\partial\alpha}\frac{{}^N d}{dt}\boldsymbol{e}_\nabla(r,\alpha,t)\right] + \mathcal{O}(\tilde{w}^3) \tag{63}$$

and

$$\boldsymbol{v}_{\bar{P}} = \dot{z}_{\bar{P}}\boldsymbol{e}_z \tag{64}$$

respectively. The derivation for the contact at Q can be performed analogously.

3.2 Contact Forces. The directions of the friction forces follow from the relative velocities between the points $\tilde{P}$ (of the rotating plate) and the point $\bar{P}$ (of the brake pad) and correspondingly between $\tilde{Q}$ and $\bar{Q}$. The unit vector in the direction of the friction force acting on the plate is

$$\boldsymbol{e}_{R_{\tilde{P}}} = \frac{\boldsymbol{v}_{\bar{P}} - \boldsymbol{v}_{\tilde{P}}}{|\boldsymbol{v}_{\bar{P}} - \boldsymbol{v}_{\tilde{P}}|} \tag{65}$$

The normal forces are perpendicular to the surface of the plate. Hence

$$\boldsymbol{e}_{N_{\tilde{P}}} = \boldsymbol{e}_\nabla(r+\Delta r_{\tilde{P}}, \alpha+\Delta\alpha_{\tilde{P}}, t) \tag{66}$$

is the direction of the normal forces acting on the upper contact point of the plate $\tilde{P}$. A force balance in $\boldsymbol{e}_z$ direction (see Fig. 2)

$$-R_P\boldsymbol{e}_{R_{\tilde{P}}}\cdot\boldsymbol{e}_z - N_P\boldsymbol{e}_{N_{\tilde{P}}}\cdot\boldsymbol{e}_z - k\left(z_{\bar{P}} + \frac{h}{2}\right) + N_0 = 0 \tag{67}$$

and Coulomb's law $R_P=\mu N_P$ yield the contact forces at P depending on $\tilde{w}(r,\alpha,t)$ and its derivatives. Analogously to the beam, the forces $\boldsymbol{N}_P$ and $\boldsymbol{R}_P$ are defined as the normal and tangential components of the contact force in the deformed configuration. The forces at Q are calculated similarly.

As mentioned above for the beam, the linearization of the equations of motion, especially of Eq. (65), requires the exclusion of stick–slip phenomena. For the linear case thus

$$\frac{h}{2r}\left|\frac{\partial^2}{\partial\varphi\,\partial t}w(r,\varphi,t)\right| \ll \dot{q}_0 r \tag{68}$$

has to hold for all times. Note that typical vibration amplitudes for a disk squealing with frequencies between 1 and 5 kHz are in the range of micrometers. Assuming harmonic vibrations of the disk with 2 kHz the realistic parameters used in the next section yield a minimal rotational speed of approximately 0.3 min^{-1}. Therefore Eq. (68) is fulfilled down to very low speeds of the disk and stick–slip should in general not be considered as a source of brake squeal.

3.3 Approximate Solution. In the previous sections we derived all necessary ingredients for Eq. (7), so we can proceed with the approximate solution scheme described in Sec. 2.3. We substitute

$$w(r,\varphi,t) = \sum_{i=1}^{I} q_i(t)W_i(r,\varphi) \tag{69}$$

into Eq. (7). The equations of the discretized system now can be derived with Lagrange's Eqs. (27) and the generalized forces

$$f_j = M_A\frac{\partial\dot{q}_0}{\partial\dot{q}_j} + (R_P\boldsymbol{e}_{R_{\tilde{P}}} + N_P\boldsymbol{e}_{N_{\tilde{P}}})\cdot\frac{\partial\boldsymbol{v}_{\tilde{P}}}{\partial\dot{q}_j} + (R_Q\boldsymbol{e}_{R_{\tilde{Q}}} + N_Q\boldsymbol{e}_{N_{\tilde{Q}}})\cdot\frac{\partial\boldsymbol{v}_{\tilde{Q}}}{\partial\dot{q}_j} \tag{70}$$

are obtained from the virtual power of all forces and torques acting on the plate. The general form of the linearized equations of motion is quite lengthy but can be calculated easily.

In order to compare the present model with the one described in Ref. [6], we now consider brake pads at $r=r_0$, $\varphi=0$[1] and using the shape function

$$w(r,\varphi,t) = R_{m,n}(r)[q_1(t)\cos m\varphi + q_2(t)\sin m\varphi] \tag{71}$$

where $R_{m,n}(r)=R(r)$ is the radial component of an eigenfunction of the corresponding nonrotating plate. The quantities m and n denote the number of nodal diameters and nodal circles, respectively. Note that all boundary conditions of the rotating plate are satisfied by the shape functions. Assuming the nonholonomic constraint

$$\dot{q}_0(t) = \Omega = \text{const} \tag{72}$$

one obtains the necessary driving torque

$$M_A = 2\mu r_0 N_0 \tag{73}$$

and the equation of motion linearized for small $q_1, q_2, \dot{q}_1, \dot{q}_2$ as

$$\boldsymbol{M}\ddot{\boldsymbol{q}} + \boldsymbol{D}\dot{\boldsymbol{q}} + \boldsymbol{K}\boldsymbol{q} = 0, \quad \boldsymbol{q}(t) = [q_1(t) \quad q_2(t)]^T \tag{74}$$

with

$$\boldsymbol{M} = \begin{bmatrix} M & 0 \\ 0 & M \end{bmatrix} \tag{75}$$

$$\boldsymbol{D} = \begin{bmatrix} \frac{1}{2}\mu N_0\frac{h^2}{r_0\Omega}R'(r_0)^2 & 2mM\Omega \\ -2mM\Omega & 0 \end{bmatrix} \tag{76}$$

[1]Due to the symmetry of the system, the circumferential position of the pads is arbitrary, i.e., it is always possible to transform the equations of motion to the described form.

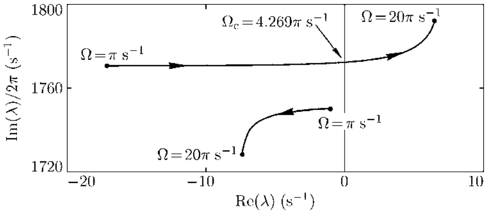


Fig. 7 Root locus for varying Ω (only eigenvalues with pos. imag. part)

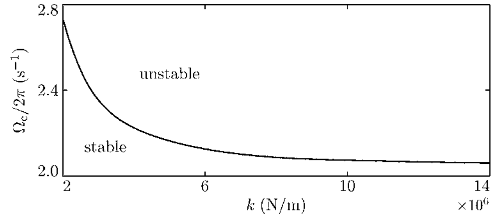


Fig. 8 Critical speed for varying *k*

$$
\boldsymbol{K} = \begin{bmatrix} k_D - m^2 M\Omega^2 + 2kR(r_0)^2 + N_0 hR'(r_0)^2 & -\mu k\dfrac{h}{r_0} mR(r_0)^2 - \dfrac{1}{2}\mu N_0 \dfrac{h^2}{r_0^3} m[r_0^2 R'(r_0)^2 - r_0 R(r_0)R'(r_0)] \\ \dfrac{1}{2}\mu N_0 \dfrac{h^2}{r_0^3} m[m^2 R(r_0)^2 + r_0^2 R'(r_0)^2 - r_0 R(r_0)R'(r_0)] & k_D - m^2 M\Omega^2 + (1+\mu^2)N_0 \dfrac{h}{r_0^2} m^2 R(r_0)^2 \end{bmatrix} \tag{77}
$$

using the abbreviations

$$
M = \rho h \pi \int_{r_i}^{r_o} rR^2(r)dr \tag{78}
$$

$$
k_D = \omega_{m,n}^2 M \tag{79}
$$

Since $R_{m,n}(r)\cos m\varphi$ and $R_{m,n}(r)\sin m\varphi$ form a pair of orthogonal eigenfunctions of the corresponding nonrotating plate, $\omega_{m,n}$ is the circular eigenfrequency of this double mode.

3.4 Stability Analysis. The equation of motion Eq. (74) deserves closer attention. The matrix $\boldsymbol{D}$ Eq. (76) not only contains the expected gyroscopic terms, but also an additional viscous damping term. The gyroscopic terms are typical for moving or rotating continua [7] and were observed in the equation of motion of the beam Eq. (34) as well. The viscous damping term results from the kinematics of the plate and the frictional contact between the disk and the brake pads. It cannot be observed in two-dimensional models, since it is due to the fact that the friction force has a component in the radial direction of the disk, even when the motion of the brake pad is restricted to the direction normal to the undeformed plate. The corresponding energy dissipation influences the stability of the trivial solution of Eq. (74) significantly and should not be neglected by an assumed circumferential direction of the friction forces. The asymmetry of the matrix $\boldsymbol{K}$ Eq. (77) makes the system susceptible to self-excited vibrations.

Since the described model is intended for the explanation of brake squeal its parameters should be chosen such that it reflects a real brake system. The corresponding realistic parameters of the present two degree of freedom model

$$
\begin{array}{lll}
M = 0.50\ \mathrm{kg}, & k_D = 6.14 \cdot 10^7\ \mathrm{N/m}, & m = 3 \\
r_0 = 0.13\ \mathrm{m}, & R(r_0) = 0.35, & R'(r_0) = 6.15\ \mathrm{m}^{-1} \\
h = 0.02\ \mathrm{m}, & \Omega = 5\pi\ \mathrm{s}^{-1}, & N_0 = 2.00\ \mathrm{kN} \\
k = 6.00 \cdot 10^6\ \mathrm{N/m}, & \mu = 0.6 &
\end{array}
$$

are based on the multi-body disk brake models described in Refs. [12,13]. They represent an equivalent annular Kirchhoff plate with parameters

$$
\begin{array}{lll}
\rho h = 126\ \mathrm{kg/m}^2, & D = 67{,}100\ \mathrm{Nm}, & \nu = 0.3 \\
r_i = 0.025\ \mathrm{m}, & r_0 = 0.162\ \mathrm{m} &
\end{array}
$$

vibrating in a mode with three nodal diameters and one nodal circle. Several of the parameters will be varied in the following, to show their influence on the stability of the trivial solution of Eq. (74). All other parameters will be held constant unless mentioned otherwise in the corresponding diagram.

Brake squeal usually occurs at low speed and relatively low braking pressure, e.g., during a roll out in front of a traffic light. Therefore the rotational speed of the brake disk is an important parameter in matters of brake squeal. Substituting

$$
\boldsymbol{q}(t) = \hat{\boldsymbol{q}} e^{\lambda t} \tag{80}
$$

into the equation of motion Eq. (74) leads to an eigenvalue problem for λ, where a positive real part of one of the eigenvalues corresponds to an instability of the trivial solution. Figure 7 shows the root locus of the eigenvalues for varying speeds of the brake disk. Above a critical speed Ω_c there exist eigenvalues with a positive real part. That is to say, the trivial solution becomes unstable and the system shows self-excited vibrations, which can be interpreted as squeal. This instability in the linear system implies that the vibration amplitudes of the linearized system tend toward infinity. However, the vibration amplitudes of a real brake system are bounded and tend toward a limit cycle. This behavior can be modeled by considering nonlinearities of the system, e.g., a hardening spring characteristic of the brake pad [14,12].

A common first step to improve the noise behavior of a brake system is the modification of the pad's back plate by adding damping layers or a change of the friction material itself. The parameter of the present model corresponding to the stiffness of the friction material is k. The influence of k on the critical speed of the system is shown in Fig. 8. The critical speed increases for decreasing stiffness of the brake pads k. In contrast to that, the critical speed does not significantly depend on the friction coefficient μ for the present model.

In addition to the braking performance, the improvement of the

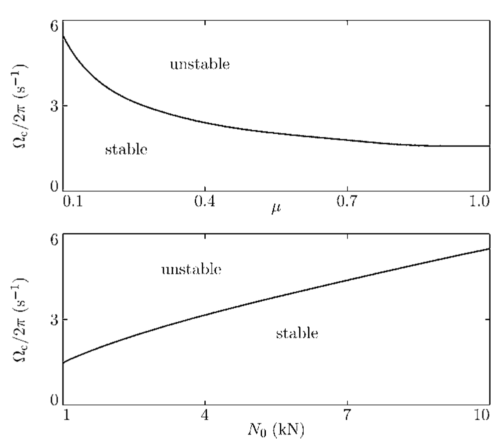


Fig. 9 Critical speed for varying μ and N_0 but constant braking torque (r_0=0.13 m)

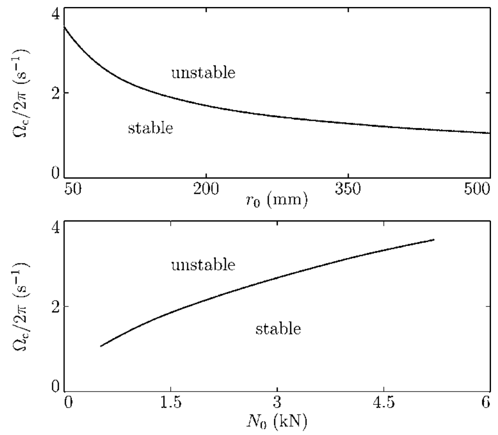


Fig. 10 Critical speed for varying r_0 and N_0 but constant braking torque (μ=0.6)

noise behavior of the brake system is an important goal of the manufacturer. Within the scope of the present model, the performance of the brake system is given by the braking torque Eq. (73) $M_A = 2\mu r_0 N_0$. An important question is how μ, r_0, and N_0 should be chosen to improve the noise behavior of the system (i.e., to increase the critical speed of the disk) without affecting the braking torque. Figure 9 shows the critical speed of the disk for varying μ and N_0, whereas r_0 and M_A are held constant. The graph indicates that with this goal in mind the braking torque should be generated by a high "brake pressure" N_0 and a low friction coefficient μ. The dependence of the critical speed on r_0 and N_0 with constant braking torque is shown in Fig. 10. As in Fig. 9, here the critical speed also increases with increasing "brake pressure" for constant brake torque. One can see that the brake disk should not be too large in order to avoid very low critical speeds.

It is obvious that all the design proposals obtained from Figs. 9 and 10 conflict with the performance of the brake system. A low friction coefficient μ and a small radius r_0 limit the capability of the brake system in particular for emergency stops. It should be mentioned that the presented stability results coincide with the ones for the minimal brake disk model in Ref. [6]. Even an extended version of the model shows a qualitatively similar behavior (see Ref. [14]), as long as the contact mechanism is modeled in the presented way. In general the stability results strongly depend on the chosen parameters. In particular the viscous damping due to the kinematics influences the stability of the system significantly. In contrast to the traveling beam problem, the presence of damping renders the system asymptotically stable for certain parameter configurations, so that a stability boundary can be found.

4 Conclusions

Two cases of continuous systems in frictional point contact with idealized brake pads are described in this paper. A discretization scheme is introduced and its good convergence was shown for the traveling beam. It is pointed out that a consistent minimal model for the explanation of disk brake squeal has to consider the three-dimensional kinematics of the brake disk. Even though an excitation mechanism can be shown with two-dimensional problems like the traveling beam or the simplified models from the literature [3–5], the more accurate formulation of the frictional contact in the three-dimensional case yields additional dissipative terms. The derived minimal disk brake model is an extension of the one presented in Ref. [6] and the stability behavior of the models coincides. The analysis of the stability of the trivial solution shows the influence of the brake's design parameters.

The model is presently extended to distributed contact over finite areas and also to include the in- and out-of-plane motions of the brake pads. The equations of motion derived with the presented discretization scheme are well applicable for a model based control of disk brake squeal. Furthermore, the minimal model can be used to verify the contact formulation of more comprehensive models, for example models based on a commercial FEM formulation.

Acknowledgment

The authors thank Dr. Oleg Kirillov from Moscow State Lomonosov University, presently a Alexander-von-Humboldt Research Fellow at TU Darmstadt, Germany, for the fruitful discussion on stability theory, which helped to improve the present paper. In particular he brought to the authors' attention Refs. [9,10].

References

[1] Kinkaid, N. M., O'Reilly, O. M., and Papadopoulos, P., 2003, "Automotive Disc Brake Squeal," J. Sound Vib., **267**, pp. 105–166.

[2] Ibrahim, R. A., 1994, "Friction-Induced Vibration, Chatter, Squeal, and Chaos, Part II: Dynamic and Modeling," Appl. Mech. Rev., **47**(7), pp. 227–253.

[3] Ouyang, H., and Mottershed, J. E., 2004, "Dynamic Instability of an Elastic Disk under the Action of a Rotating Friction Couple," J. Appl. Mech., **71**, pp. 753–758.

[4] Ono, K., Chen, J., and Bogy, D. B., 1991, "Stability Analysis for the Head-Disk Interface in a Flexible Disk Drive," J. Appl. Mech., **58**, pp. 1005–1014.

[5] Kinkaid, N. M., O'Reilly, O. M., and Papadopoulos, P., 2005, "On the Transient Dynamics of a Multi-Degree-of-Freedom Friction Oscillator: A New Mechanism for Brake Noise," J. Sound Vib., **287**(4–5), pp. 901–917.

[6] von Wagner, U., Hochlenert, D., and Hagedorn, P., "Minimal Models for the Explanation of Disk Brake Squeal," J. Sound Vib., in press.

[7] Wickert, J. A., and Mote, C. D., Jr., 1990, "Classical Vibration Analysis of Axially Moving Continua," J. Appl. Mech., **57**, pp. 738–744.

[8] Chakraborty, G., and Mallik, K., 2000, "Wave Propagation in and Vibration of a Travelling Beam With and Without Non-Nonlinear Effects," J. Sound Vib., **236**(2), pp. 277–290.

[9] Karapetjan, A. V., 1975, "The Stability of Nonconservative Systems," Vestn. Mosk. Univ., Ser. 1: Mat., Mekh., **30**(4), pp. 109–113.

[10] Lakhadanov, V. M., 1975, "On the Stability of Nonlinear Systems," J. Appl. Math. Mech., **39**(1), pp. 53–58.

[11] Leissa, A., 1969, "Vibration of Plates," NASA SP-160, Washington, D.C.

[12] Jearsiripongkul, T., 2005, "Squeal in Floating Caliper Disk Brakes: A Mathematical Model," Ph.D. thesis, Technische Universität Darmstadt, Germany.

[13] von Wagner, U., Hochlenert, D., Jearsiripongkul, T., and Hagedorn, P., 2004, "Active Control of Brake Squeal via Smart Pads," *Proceedings of the 22nd Annual Brake Colloquium*, Anaheim, CA, SAE Paper 2004-01-2773.

[14] Hochlenert, D., and Hagedorn, P., 2006, "Control of Disc Brake Squeal—Modelling and Experiments," *Structural Control and Health Monitoring*, Wiley, New York, **13**, pp. 260–276.

Wrinkling of Plane Isotropic Biological Membranes

Roberta Massabò[1]
e-mail: Roberta.massabo@unige.it

Luigi Gambarotta

Department of Civil, Environmental and Architectural Engineering, University of Genova, Genova, 16145 Italy

*The problem of the wrinkling of plane isotropic membranes characterized by a Fung type constitutive model in biaxial tension has been formulated and solved within the framework of finite strain hyperelasticity. The formulation follows the approach of Pipkin [Pipkin, A.C., 1986, IMA J. Appl. Math., **36**, pp. 85–99; 1994, ibid., **52**, pp. 297–308], and the out of plane geometric nonlinearities are treated as constitutive nonlinearities through a modification of the elastic potential. The wrinkling criteria are based on the natural contraction of a membrane in simple tension. Both the natural contraction and the modified elastic potential are defined in closed form. The model has been implemented in a finite element code and the numerical solution validated using study cases with analytical solution. Applications are presented that simulate the response of stretched membranes, where distinct regions of behavior (taut, wrinkled, and slack or inactive) develop during loading, and a simple procedure of reconstructive surgery, characterized by the excision of a circular portion of the skin and the suture of the wound edges, where the wrinkling of the skin causes the extrusion of the edges (dog-ear formation).* [DOI: 10.1115/1.2424240]

Keywords: wrinkling, biological membranes, finite strains, constitutive modeling, relaxed energy density, reconstructive surgery

1 Introduction

Thin membranes, including biological membranes like the skin, have very small and, in general, negligible flexural stiffness and can be modeled as mathematical membranes. In such structures, stress states characterized by negative principal components are not admissible. In the presence of strains that would require compressive stresses, the membrane wrinkles (if the strain field would require a state of biaxial tension-compression) or becomes slack and inactive (if the strain field would require biaxial compression). Therefore, in a membrane under a general state of stress, three regions of behavior can be identified: taut regions, where the material is in a state of biaxial stress (both principal stresses are positive); wrinkled regions, where the material is wrinkled and only one principal stress component is present in the direction of the wrinkles; slack or biaxially wrinkled regions, where the membrane is inactive.

Figures 1(*a*) and 1(*b*) show results of experimental tests on stretched polyethylene membranes where the different regions of behavior are identified by the letters *T*, *W*, and *S*. The membrane in Fig. 1(*a*), after Ref. [1], is a thin rectangular sheet laterally clamped and subject to a uniform vertical displacement applied along the central portion of its base. The membrane in Fig. 1(*b*), after Ref. [2], is a thin rectangular sheet, laterally clamped and subject to uniaxial tensile strain. In this sheet wrinkling is generated by the action of the clamped boundaries that prevent transverse contraction of the sheet and set up a local biaxial state of stress; the transverse stresses are tensile near the boundary but they would tend to be compressive away from it.

In human skin, local wrinkling is often observed after the closure and suture of the wound edges in procedures of reconstructive surgery. This phenomenon gives rise to an anesthetism known as extrusion of the edges or dog-ear formation. Figure 1(*c*), taken from a surgical publication [3], shows with simple drawings an example of dog-ear formation and the main surgical techniques used, during or after primary surgery, to remove it.

For states of biaxial tension and short term loading, the response of the skin and other anisotropic biological membranes is well described by the classical hyperelastic constitutive model of Tong and Fung [4] and by the reduced version of this model, known as Fung model [5], which is also implemented in finite element codes (FEAP [6]). The model is phenomenological and treats the tissues as homogeneous. Recently, the isotropic form of Fung model has been calibrated on in vivo human scalp skin through a combined experimental/numerical procedure proposed by the authors [7]. The in vivo experiments were on scalp flaps tested within procedures of cosmetic surgery.

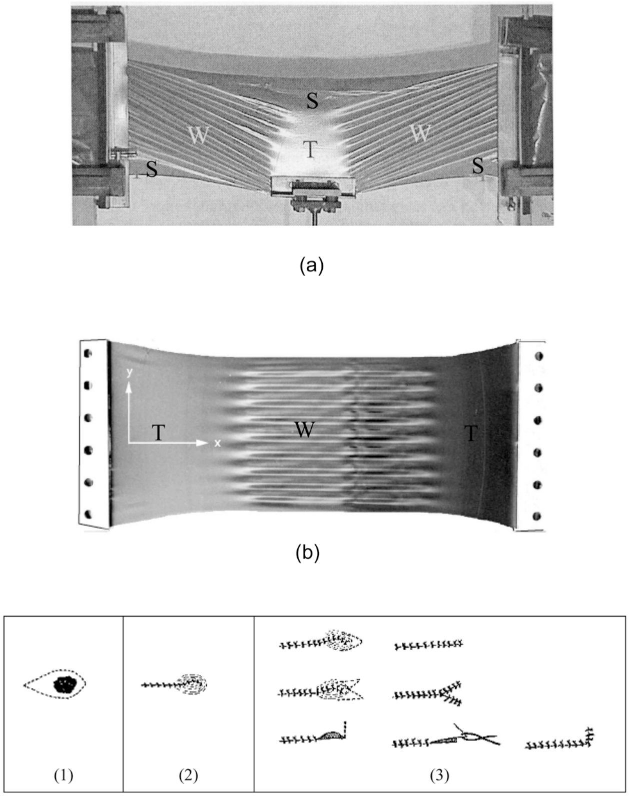
Fung model should not be used in the presence of large compressive deformations: the model leads to an unrealistic extremum in the nominal stress versus stretch curve of homogeneous uniaxial compression and does not satisfy all growth conditions [8,9]. However, in the presence of large compressive deformations, the membranes wrinkle and the above limitations are irrelevant. The problem then arises of how to treat the out of plane nonlinearities within the framework of finite strain membrane theory.

Different theories have been proposed in the literature to solve the problem of the geometric instability of finite strain membranes (Pipkin [10,11], Wu and Canfield [12], Steigmann [13], among others). All theories follow the linear theories commonly referred to as tension field theories that date back to the pioneering work of Wagner on thin metal webs [14]. Wagner first introduced the concept of Tension Field to describe the uniaxial stress state in a fully wrinkled membrane. In addition, he introduced a criterion for wrinkling based on the transverse contraction of a membrane in uniaxial tension. He also noticed that wrinkling occurs with no expense of energy and proved that the wrinkles are straight lines. Additional work on the tension field theory is due to Mansfield [15,16] who applied energy principles to define the directions of the tension rays in wrinkled regions. The main limitations of classical tension field theories are the assumptions of small deformations and of a body that is fully wrinkled at equilibrium.

Among the nonlinear theories, reference is made in this paper to the relaxed energy density theory of Pipkin [10,11]. The theory treats the out of plane geometric nonlinearities as constitutive nonlinearities through a modification of the strain energy function

[1]Corresponding author.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received June 29, 2005; final manuscript received June 26, 2006. Review conducted by Ellen M. Arruda.



(c)

Fig. 1 (*a*) Experimental test on a clamped rectangular polyethylene membrane, after Barsotti et al. [1]. (*b*) Experimental test on a clamped rectangular polyethylene membrane loaded in axial tension, after Cerda et al. See Ref. [2]. (*c*) Drawings after Ref. [3]: (1) Spindle shape excision of a skin cancer; (2) dog-ear formation following the suture of the wound edges; (3) procedures used to eliminate the extrusion (postsuture) or to avoid it (while suturing).

of the material in the wrinkled and slack regions. The new function is defined the Relaxed Energy Density function. Relaxed energy functions have been derived in Ref. [10] for a neo-Hookean material, in Ref. [17] for a Varga material and in Ref. [1] for a linear elastic material. This approach has a number of interesting features: the problem can be solved using the in plane membrane field variables, with no need to refer to out of plane variables; the theory uses a variational approach and therefore it overcomes the difficulties of other theories in the definition of the boundaries between wrinkled, slack, and taut regions in partly wrinkled membranes. In addition, the modified constitutive model can be easily implemented in any commercial finite element codes as a user defined material model so that the well established nonlinear algorithms of solution of the code can be used. For completeness it should be mentioned that other numerical techniques have been used in the literature to analyze partly wrinkled membranes with small and finite strains using finite elements and material models other than Fung model. These techniques simulate wrinkling for instance through a modification of the material stiffness matrix or the material constitutive relationships (no-compression materials) ([18–23], among others).

A hyperelastic constitutive model is derived in this paper that accounts for the wrinkling of finite strain plane biological membranes characterized by a Fung type behavior and subject to in-plane and geometry independent loading. Only isotropic membranes, such as those examined in Ref. [7], are considered at this stage. The model is implemented in the finite element code FEAP [6] and used to analyze some exemplary problems.

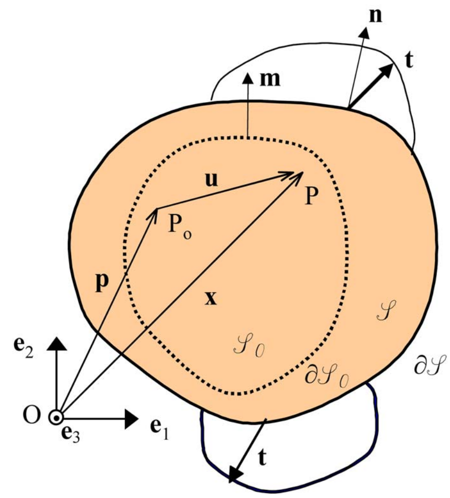


Fig. 2 Reference (dotted line) and current (solid line) configurations of a plane membrane

2 Membrane Stress and Strain Measures

The study refers to finite strain plane homogeneous membranes subject to in-plane loading. Figure 2 shows an exemplary plane membrane and a Cartesian reference basis with unit vectors $\mathbf{e}_\alpha$ $(\alpha=1,2,3)$. The membrane is in the plane $(\mathbf{e}_1,\mathbf{e}_2)$. In the reference configuration, $\mathcal{S}_0$ with boundary $\partial\mathcal{S}_0$ and unit normal $\mathbf{m}$ (dotted lines), the membrane is assumed to be undeformed. In the current configuration, the material points occupy the surface $\mathcal{S}$ with boundary $\partial_{\mathcal{S}}$ of unit normal $\mathbf{n}$. The spatial position vector, $\mathbf{x}$, defines the position of the material point in the current configuration and is related to the displacement vector $\mathbf{u}$ by $\mathbf{u}=\mathbf{x}\text{-}\mathbf{p}$, with $\mathbf{p}$ the position vector in the reference configuration. The components of $\mathbf{u}$ are $u_\alpha=\mathbf{u}\cdot\mathbf{e}_\alpha$.

The plane deformation of the membrane is described by the mapping $\mathbf{x}=\mathbf{f}(\mathbf{p})$ and by the in plane deformation gradient $\mathbf{F}$, and is measured locally by the material Green–Lagrange strain tensor $\mathbf{E}=\varepsilon_{\alpha\beta}\mathbf{e}_\alpha\otimes\mathbf{e}_\beta$, where $\varepsilon_{\alpha\beta}$ for $\alpha,\beta=1,2$ are the strain components in the Cartesian reference system and $\otimes$ is the tensor product. The polar decomposition, $\mathbf{F}=\mathbf{RU}$, defines the deformation gradient in terms of the right stretch tensor, $\mathbf{U}$, and the rotation tensor, $\mathbf{R}$. The spectral decomposition of $\mathbf{U}$ gives the principal stretches λ_α, $\mathbf{U}=\lambda_\alpha\mathbf{n}_\alpha\otimes\mathbf{n}_\alpha$, with $\mathbf{n}_\alpha$ as the unit eigenvectors. The principal components of $\mathbf{E}$ are related to the principal stretches, λ_1 and λ_2, through $\varepsilon_\alpha=1/2(\lambda_\alpha^2-1)$.

The vector $\mathbf{t}$ in Fig. 2 defines edge forces acting per unit initial length on the boundary ∂S of the surface The membrane forces are described by the spatial tensor $\mathbf{T}=t_{\alpha\beta}\mathbf{e}_\alpha\otimes\mathbf{e}_\beta$, the second Piola–Kirchhoff tensor $\mathbf{\Sigma}=\sigma_{\alpha\beta}\mathbf{e}_\alpha\otimes\mathbf{e}_\beta$ (PKII) and the first Piola-Kirchhoff tensor $\mathbf{S}=s_{\alpha\beta}\mathbf{e}_\alpha\otimes\mathbf{e}_\beta$ (PKI), where $\alpha,\beta=1,2$, and $t_{\alpha\beta}$, $\sigma_{\alpha\beta}$, and $s_{\alpha\beta}$ are the corresponding membrane force components. The tensors are related through $\mathbf{\Sigma}=J\mathbf{F}^{-1}\mathbf{T}\mathbf{F}^{-T}$ and $\mathbf{S}=J\mathbf{T}\mathbf{F}^{-T}$, where $J=\det(\mathbf{F})=\lambda_1\lambda_2$. In the case of isotropic elasticity, $\mathbf{U}$ and $\mathbf{\Sigma}$ are coaxial.

Where the membrane wrinkles, an out of plane displacement arises, $\mathbf{u}_3=u_3\mathbf{e}_3$, and the deformation gradient becomes highly discontinuous (see Sect. 4). A measure of the wrinkliness of the deformation is given by the wrinkle-strain tensor $\mathbf{W}$ introduced by Wu and Canfield [12]. The principal quantities associated with $\mathbf{W}$ are $W_1=\lambda_1-\lambda_1^*$ and $W_2=\lambda_2-\lambda_2^*$, where λ_1^* and λ_2^* are the true principal stretches experienced by the deformed (wrinkled) surface and defined by the ratio of the deformed length of the membrane in the wrinkled configuration and the undeformed length (λ_1

and λ_2 are membrane variables and refer to the projection of the wrinkled membrane onto the original tangent plane). If the 1-direction is the direction of the wrinkles, then $\lambda_1=\lambda_1^*$ and $W_1=0$ (the membrane is in tension in the 1-direction) while $W_2=\lambda_2-\lambda_2^*<0$.

3 Fung Constitutive Model for Membranes in Biaxial Tension

Following the classical model of Tong and Fung [4] for anisotropic tissues, the elastic strain energy per unit surface area in an isotropic membrane, $\hat{W}(\mathbf{E})$, is defined as a function of the Green-Lagrange membrane strain tensor, $\mathbf{E}$, by the following form:

$$\hat{W}(\mathbf{E}) = A[\varepsilon_{11}^2 + \varepsilon_{22}^2 + 2\alpha\varepsilon_{11}\varepsilon_{22} + 2(1-\alpha)(\varepsilon_{12}\varepsilon_{21})] + c\exp\{B[\varepsilon_{11}^2 + \varepsilon_{22}^2 + 2\beta\varepsilon_{11}\varepsilon_{22} + 2(1-\beta)(\varepsilon_{12}\varepsilon_{21})]\} \tag{1a}$$

where A, α, c, B, and β are constants to be determined experimentally. The parameter A and c have dimensions $[\mathrm{F}][\mathrm{L}]^{-1}$ and α, β, and B are dimensionless. The second Piola–Kirchhoff membrane force tensor, PKII, is derived from Eq. (1a) and its components are:

$$\sigma_{\alpha\beta} = \partial\hat{W}/\partial\varepsilon_{\alpha\beta} \quad (\alpha,\beta=1,2) \tag{1b}$$

The coefficients of the constitutive tangent matrix are calculated through partial derivation of the $\sigma_{\alpha\beta}$ in Eq. (1b).

The model Eq. (1a) is able to reproduce the main characteristic features of the mechanical response of soft isotropic biological tissues subject to short term biaxial tension. In particular, the model reproduces the highly nonlinear strain stiffening response of two-phase tissues, like the skin, that at low strains is mainly controlled by the elastin and at large strain by the collagen fibers, after they have become uncrimpled.

A reduced version of Eq. (1a), also known as Fung model, omits the first term leaving only the exponential term that is sufficient to accurately describe the mechanical response provided the level of stress is not too small [5]:

$$\hat{W}(\mathbf{E}) = c\exp\{B[\varepsilon_{11}^2 + \varepsilon_{22}^2 + 2\beta\varepsilon_{11}\varepsilon_{22} + 2(1-\beta)(\varepsilon_{12}\varepsilon_{21})]\} \tag{2}$$

where the argument of the exponential is a quadratic form of the strain components. In a compact form, the elastic potential (2) is written as $\hat{W}(\mathbf{E})=c\exp(\mathbf{E}\cdot\mathbb{A}\mathbf{E})$, where $\mathbb{A}$ is a fourth order isotropic tensor depending on the dimensionless constants B and β. At incipient deformation, for $\mathbf{F}=\mathbf{I}$, the constitutive parameters in Eq. (2) assume mechanical significance. In particular, β represents a generalized Poisson coefficient, given by $(\partial\sigma_{11}/\partial\varepsilon_{22})\times(\partial\sigma_{11}/\partial\varepsilon_{11})^{-1}$, and the products $2cB$ and $cB(1-\beta)$ are generalized Young's and shear moduli, given by $(\partial\sigma_{11}/\partial\varepsilon_{11})$ and $1/2(\partial\sigma_{12}/\partial\varepsilon_{12})$, respectively [4].

Constitutive inequalities, which define the range of admissible values of the model parameters c, B, and β, must be defined to ensure a strain-stiffening response and the convexity of the strain energy function. To reproduce the strain stiffening behavior of the skin, the parameter c must be positive and the argument of the exponential function must be positive definite. These conditions, along with the requirement of convexity of $\hat{W}$, yield [7]

$$c>0, \quad B>0, \quad -1<\beta<1 \tag{3}$$

These inequalities ensure the condition of convexity of $\hat{W}$ at incipient deformation ($\mathbf{F}=\mathbf{I}$), which gives the strongest limitations on the parameters, both in the plane of the Green strains and in the plane of the principal stretches λ_i (for $i=1,2$), for all states characterized by positive principal strains.

4 Constitutive Model for Wrinkling Membranes

The model (2) cannot be used to describe the behavior of tissues in compression. As already mentioned in Sec. 1 the model yields an unrealistic extremum in the nominal stress versus stretch curve of homogeneous uniaxial compression and does not satisfy the growth condition $\hat{W}(\mathbf{F})\to+\infty$ for $\det\mathbf{F}\to+0^+$ (i.e., an infinite amount of energy should be required to reduce the length of one side of a rectangle of material to zero) [8,9]. However, these limitations are unimportant when dealing with biological membranes, such as the skin. Elastic membranes have very small and in general negligible compressive and flexural rigidity and they are unable to sustain compressive stresses. In the presence of strains that would require compressive stresses, the membrane wrinkles or becomes slack.

In this formulation it is assumed that the membrane has zero compressive and bending stiffness. In such a structure, minimum energy states can involve a continuous distribution of infinitesimal wrinkles (oscillations with very high frequency characterized by magnitude and wavelength tending to zero and leading to a discontinuous deformation gradient) that cannot be described by ordinary membrane theory [15]. The problem is solved following the approach of Pipkin [10,11] and the representation is in terms of principal Green–Lagrange strains, ε_1 and ε_2. It is assumed that wrinkling occurs in the directions of the principal strains, which coincide with the directions of the principal PKII membrane forces since the material is isotropic. For vanishing rotations $\mathbf{R}=\mathbf{I}$, $\boldsymbol{\Sigma}$ and $\mathbf{T}$ are coaxial and the principal directions describe the actual distribution of the wrinkles in the current deformed configurations.

The *natural contraction* of the membrane is defined as the transverse contraction of a membrane in a state of simple tension. With a specified $\varepsilon_1>0$, the natural contraction is the value of $\varepsilon_2=\hat{w}(\varepsilon_1)$ at which the minimum with respect to ε_2 of the elastic potential $\hat{W}$ of Eq. (2) is attained. This leads to the condition:

$$2cB(\varepsilon_2+\beta\varepsilon_1)\exp\{B[\varepsilon_1^2+\varepsilon_2^2+2\beta\varepsilon_1\varepsilon_2]\}=0 \tag{4}$$

where the elastic potential of Eq. (2) has been defined in the plane of the principal strains. Consequently

$$\varepsilon_2=\hat{w}(\varepsilon_1)=-\beta\varepsilon_1 \tag{5}$$

From Eq. (1b), $\sigma_2=0$ at the minimum of $\hat{W}$ and the only nonzero membrane force component σ_1 is given by

$$\sigma_1 = \left.\frac{\partial\hat{W}(\varepsilon_1,\varepsilon_2)}{\partial\varepsilon_1}\right|_{\varepsilon_2=\hat{w}(\varepsilon_1)} = 2cB\varepsilon_1(1-\beta^2)\exp\{B\varepsilon_1^2(1-\beta^2)\} \tag{6}$$

From the natural contraction, the natural width used by Pipkin [10,11] can be derived; it represents the deformed transverse dimension of a unit square of material and is given by the principal stretch $\lambda_2=\sqrt{2\hat{w}(\varepsilon_1)+1}$.

4.1 Wrinkling Criteria. Following the original formulation of Wagner [14] and the work of Pipkin [10,11], the natural contraction is used to set up criteria for wrinkling. It is assumed in the following that $\varepsilon_1>\varepsilon_2$.

If $\varepsilon_1>0$ and $\varepsilon_2>\hat{w}(\varepsilon_1)=-\beta\varepsilon_1$, then a tensile membrane force $\sigma_2>0$ must exist in the membrane to make the membrane wider in the transverse direction than it would be in simple tension. The membrane therefore is taut and the analysis can be performed using the elastic potential (2).

If $\varepsilon_1>0$ and $\varepsilon_2=\hat{w}(\varepsilon_1)=-\beta\varepsilon_1$, then the membrane force $\sigma_2=0$ as noted above, the membrane is in simple tension and again the analysis requires no changes to the elastic potential (2).

If $\varepsilon_1>0$ and $-1/2\leqslant\varepsilon_2<\hat{w}(\varepsilon_1)=-\beta\varepsilon_1$, then the membrane is narrower in the transverse direction than it would be in simple tension. Using the elastic potential (2) and classical membrane

Table 1 Wrinkling criteria

	Criteria for wrinkling	Relaxed energy density
Taut regions	If $\varepsilon_1 \geqslant \hat{w}(\varepsilon_2) = -\beta\varepsilon_2$ and $\varepsilon_2 \geqslant \hat{w}(\varepsilon_1) = -\beta\varepsilon_1$	$\hat{W}_r(\varepsilon_1, \varepsilon_2) = \hat{W}(\mathbf{E}) = c\,\exp\{B[\varepsilon_1^2 + \varepsilon_2^2 + 2\beta\varepsilon_1\varepsilon_2]\}$
Wrinkled regions	If $\varepsilon_1 \geqslant 0$ and $-1/2 \leqslant \varepsilon_2 \leqslant \hat{w}(\varepsilon_1) = -\beta\varepsilon_1$ If $\varepsilon_2 \geqslant 0$ and $-1/2 \leqslant \varepsilon_1 \leqslant \hat{w}(\varepsilon_2) = -\beta\varepsilon_2$	$\hat{W}_r(\varepsilon_1) = c\,\exp\{B\varepsilon_1^2(1-\beta^2)\}$ $\hat{W}_r(\varepsilon_2) = c\,\exp\{B\varepsilon_2^2(1-\beta^2)\}$
Slack regions	If $\varepsilon_1 \leqslant 0$ and $\varepsilon_2 \leqslant 0$	$\hat{W}_r(\varepsilon_1, \varepsilon_2) = 0$

theory, a compressive membrane force $\sigma_2 < 0$ would be predicted. Instead, the membrane stays in simple tension and the decrease in transverse width is accomplished by wrinkling. The tension field that arises can be included in the theory through a modification of the elastic potential (2). The new energy function, which is called here relaxed energy function, $\hat{W}_r$, following [10,11], represents the average energy function per unit area in the wrinkled region and it will be defined in the next section.

If $\varepsilon_1 \leqslant 0$, then both membrane forces should be negative according to the classical membrane theory and the elastic potential (2). Instead, the membrane is inactive, namely $\sigma_1 = \sigma_2 = 0$, and the decreases in widths are accomplished by biaxial wrinkling. As for the previous case, this can be included in the model through a modification of the elastic potential (2).

4.2 Relaxed Energy Function. According to Pipkin [10,11], the relaxed energy function, $\hat{W}_r$, is the quasi-convexification of $\hat{W}$. No general algorithms exist to define $\hat{W}_r$ from $\hat{W}$. However, $\hat{W}_r$ can be defined in the plane of the principal Green–Lagrange strains considering the events that take place when the compressive strain in a membrane is progressively increased while the tensile strain is kept constant. Once the limiting condition for wrinkling has been reached, at $\varepsilon_2 = \hat{w}(\varepsilon_1) = -\beta\varepsilon_1$, where the energy function (2) is at its minimum in ε_2, the folding process that follows for $\varepsilon_2 < \hat{w}(\varepsilon_1)$ occurs with no expense of energy and the elastic potential $\hat{W}$ will retain its minimum value. Therefore $\hat{W}_r$ in the wrinkling regions takes the value $\hat{W}_r(\varepsilon_1) = \hat{W}[\varepsilon_1, \hat{w}(\varepsilon_1)]$ or:

$$\hat{W}_r = c\,\exp\{B\varepsilon_1^2(1-\beta^2)\} \tag{7}$$

The principal membrane forces can be derived from Eq. (7), using Eq. (1b) with $\hat{W}_r$ instead of $\hat{W}$, that shows that $\sigma_2 = 0$ and σ_1 is given by Eq. (6). Because of the isotropy, the directions of the principal PKII membrane forces, as well as of the Cauchy membrane forces, do not change because of wrinkling.

Once the limiting condition for biaxial wrinkling has been reached, $\varepsilon_1 \leqslant 0$ and $\varepsilon_2 < \varepsilon_1$, the membrane becomes inactive, it is not stretched and its energy function stays equal to zero in all possible deformations that follows. Therefore, in the slack regions, $\hat{W}_r(\varepsilon_1) = 0$.

The above considerations are summarized in Table 1 that also generalizes the solution to cases with $\varepsilon_2 > \varepsilon_1$.

It is easily proved that $\hat{W}_r$ is convex in all domains defined in (Table 1) provided the elastic constants satisfy the inequalities (3) (convexity in the domain $\varepsilon_1 \geqslant -\beta\varepsilon_2$ and $\varepsilon_2 \geqslant -\beta\varepsilon_1$ where $\hat{W}_r(\varepsilon_1, \varepsilon_2) = \hat{W}(\mathbf{E})$ has been proven previously in Ref. [7]). Consequently, $\hat{W}_r$ is quasiconvex and satisfies the requirements of Pipkin theory [10,11].

All basic assumptions of the tension field theory emerge as a result of the assumed relaxed energy functions once they are used in the solution of the hyperelastic membrane.

The criteria formulated here, which are based on the principal strains, could also be expressed as mixed criteria in terms of principal stresses and strains [18]. Assuming $\varepsilon_1 \geqslant \varepsilon_2$ and $\sigma_1 \geqslant \sigma_2$, the criteria become: ($T$) $\sigma_2 > 0$ in a taut region, (S) $\sigma_1 \leqslant 0$ and $\varepsilon_1 \leqslant 0$ in a slack region, and (W) $\sigma_2 \leqslant 0$ and $\varepsilon_1 > 0$ in a wrinkled region. On the other hand, the criteria cannot be posed in terms of principal stresses only: without the condition $\varepsilon_1 \leqslant 0$ in (S), the criterion could indicate as slack a region that has $\varepsilon_1 > 0$ and is therefore uniaxially wrinkled. Posing the criterion in terms of principal stresses and strains is necessary when dealing with orthotropic materials since the wrinkles align themselves with the direction of the positive principal stresses that do not necessarily coincide with the principal strain directions [20]. An extension of the constitutive model formulated here to orthotropic Fung materials would require such an approach.

In real membranes the compressive and bending stiffnesses, while negligible, are nonzero and the wrinkles have nonzero magnitude and wavelength [2]. In addition, two types of creases can occur, deep folds and wrinkles, depending on the type of structure and loading conditions [21]. In the analysis above, which has been developed within the framework of membrane theory, there has been no need to distinguish between folds and wrinkles since the actual distribution of creases cannot be described [22]. However, the relaxed energy function (Table 1) derived for fine-scale wrinkling leads to statically determinate solutions that are appropriate if a detailed knowledge of the deformation field is not required and only predictions of the sizes of the different regions are sought [17].

5 Calibration of the Model Parameters

A numerical/experimental procedure has been formulated in Ref. [7] for the characterization of the elastic constants of the constitutive model (2) for in vivo human skin. The procedure has been applied to scalp flaps [1][2] tested in vivo within procedures of cosmetic surgery (hair replacement) [24]. The scalp flaps were rectangular with dimensions 120×150 mm. In the analyses, the scalp skin has been assumed to be homogeneous and isotropic. The assumption of isotropy was made mainly to simplify the numerical solution and reduce the number of constitutive parameters to be determined from experiments and it is partially justified by the observation that the relaxed skin tension lines are concentrical on the top of the skull [25]. As a consequence of this assumption, the natural stress and strain fields of the skin are isotropic and homogeneous and are described by the principal natural stretch, λ_n.

The elastic constants of the isotropic Fung model (2) and the principal natural stretch, λ_n, which have been deduced from the inverse analysis in Ref. [7], $c = 1.13$ kPa cm, $B = 0.89$, $\beta = 0.6$, and $\lambda_n = 1.1$, will be used in the examples that follow.

[2]A flap of skin is a domain of the skin that has been detached (undermined) from the subcutaneous attachments.

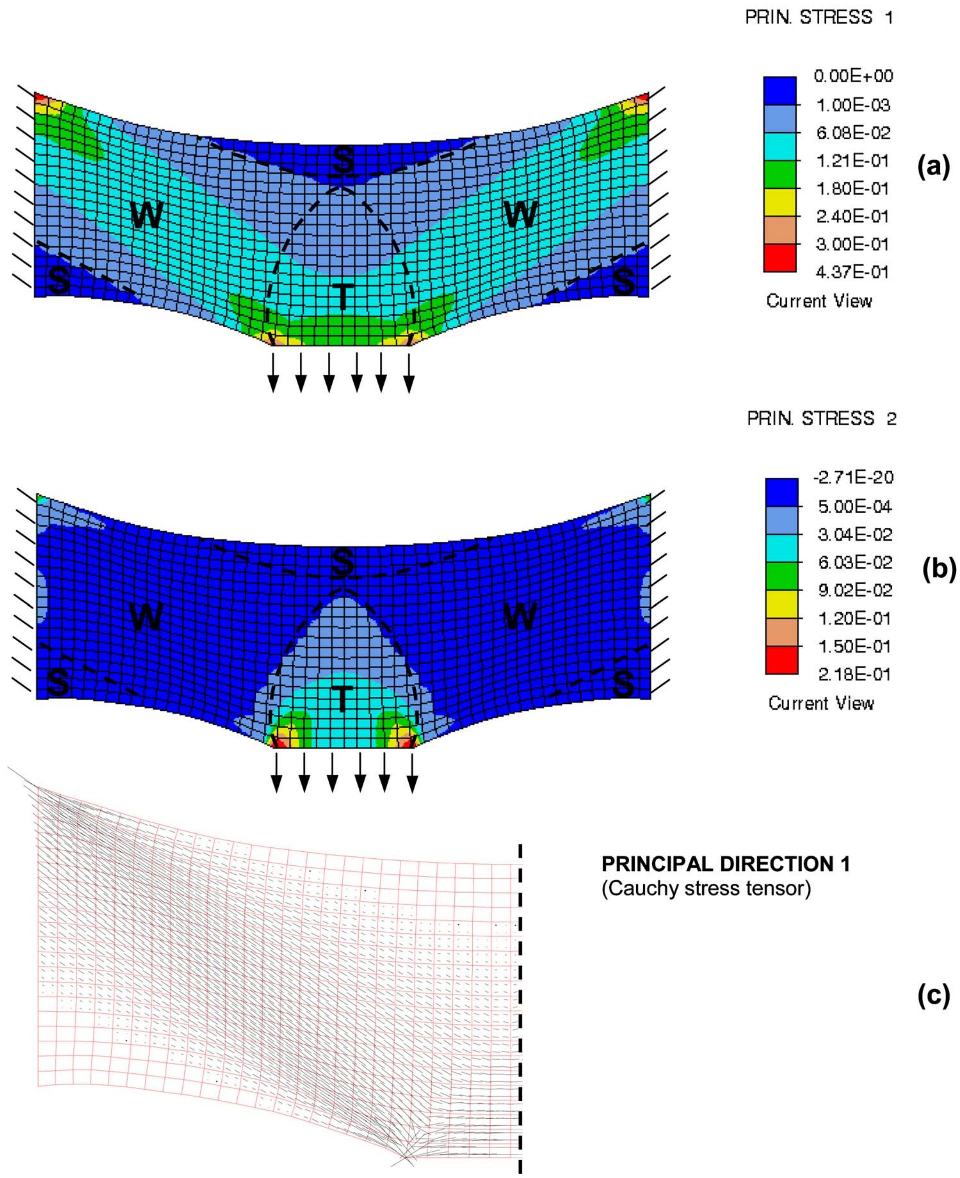


Fig. 3 Principal membrane force fields (PKII, in kPa cm) plotted onto the deformed configuration of a stretched membrane. The figure highlights regions where the membrane is in biaxial tension (*T*), where it has wrinkled (*W*) and where it is slack and inactive (*S*). The principal direction 1 of the Cauchy membrane force tensor indicates the direction of the wrinkles in the *W* regions (shown on half of the domain).

6 Implementation of the Constitutive Model Into a Finite Element Code and Applications

The constitutive model derived in Section 4 for a wrinkling membrane has been implemented in the finite element code FEAP [6]. Since the out of plane geometric nonlinearities are treated as constitutive nonlinearities, this has been simply done through the definition of a new finite deformation hyperelastic model among the material models of FEAP. The model coincides with the Fung model, which is already implemented in FEAP, for states of biaxial tension and accounts for uniaxial and biaxial wrinkling in strain states of tension-compression and compression-compression through the modified elastic potential and the criteria of (Table 1). The application of the model in Table 1 only requires the calculation of the principal strains, which is easily accomplished from the special strain components. For the solution of the nonlinear problem, a modified Newton algorithm is applied where the initial tangent matrix is properly stiffened to match the strain stiffening constitutive model so as to ensure stable convergence to the solution. Using this method, the calculation of the current tangent matrix is not necessary to find the solution, so avoiding problems of singularities associated with the presence of slack regions.

The accuracy of the finite element solution and the nonlinear solution algorithm have been checked through comparison with special problems that have analytical solution. Different homogeneous cases of membranes in tension-compression and tension and shear, already examined in Ref. [19] using a different constitutive model, have been considered. These membranes develop a tension field over the entire domain and the uniaxial stress and the direction of the wrinkles can be defined analytically using the equations in Table 1. In all cases, FEAP converged to the exact solutions within a limited number of iterations. The limiting case of a membrane subject to pure uniaxial compression has been used to test the stability of the iteration process. The numerical code obtained the exact solution, given by the inverted membrane

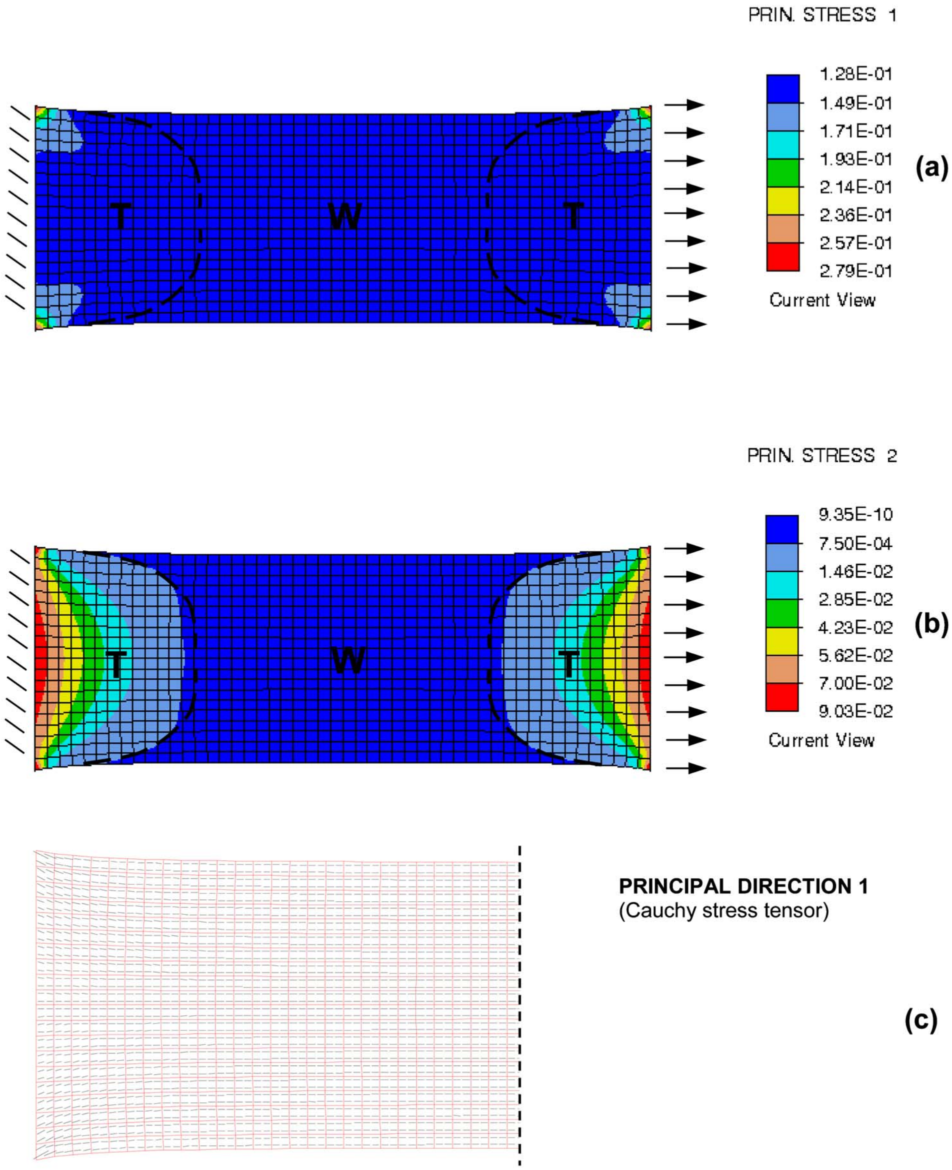


Fig. 4 Principal membrane force fields (PKII, in kPa cm) plotted onto the deformed configuration of a stretched membrane subject to axial tension. The figure highlights regions where the membrane is in biaxial tension (*T*) and where it has wrinkled (*W*). The principal direction 1 of the Cauchy membrane force tensor indicates the direction of the wrinkles in the *W* regions (shown on half of the domain).

in pure tension, even if over several iterations the membrane were completely inactive/slack.

In order to check the capability of the finite element model to predict the boundaries of the different regions, taut and wrinkled, the case of an annular membrane subject to tractions applied along the internal and external boundaries has been considered [26,27]. In the case of small deformations, the size of the wrinkled region can be derived analytically and depends only on the ratios of the internal and external tractions and the radii of the annular membrane; moreover, limiting values for these ratios exist for which the membrane is entirely wrinkled. Both the size of the wrinkled region in small deformations and the ratios that produce full wrinkling have been accurately predicted by the model implemented in FEAP. Details on the solution of this problem can be found in Ref. [27].

For the analysis of the annular membrane problem, as well as the problems that will be presented in the following, different element densities were considered to check the influence of the mesh discretization on the accuracy of the solution and on the numerical estimates of the boundaries of the wrinkled/taut/slack domains. Estimates of the levels of uncertainty are given in the following sections. It was observed that refined discretizations led to more precise estimates of the boundaries of the domains that, however, are very well predicted also by relatively coarse meshes (a similar conclusion was drawn in Ref. [22]).

6.1 Stretched Membrane-I. In order to highlight the capability of the proposed model to predict the distinct regions of behavior that may appear in a stretched membrane, the response of a thin rectangular sheet ($75 \times 25 \times 1$ mm) laterally clamped and subject to a uniform vertical displacement (u_2=6 mm) applied along the central fifth of its base has been analyzed using the finite

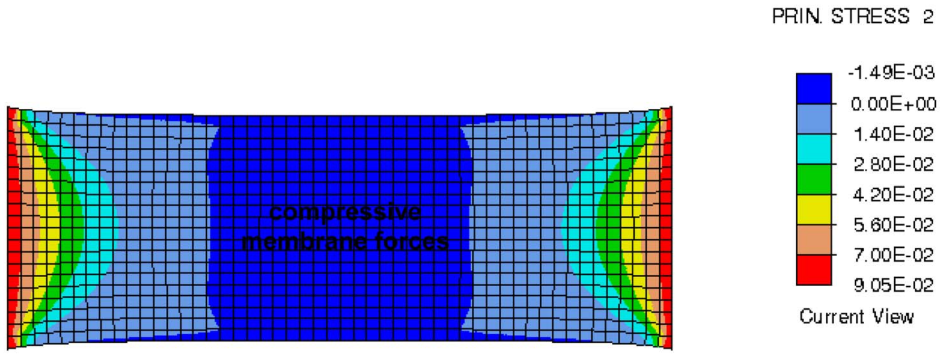


Fig. 5 Principal membrane force field (PKII, in kPa cm) obtained using the model of Eq. (2) plotted onto the deformed configuration of a stretched membrane subject to axial tension. The figure highlights the region where the membrane forces are compressive (up to –1.49E–03 kPa cm).

element model formulated in this paper. The elastic constants $c = 1.13$ kPa cm, $B = 0.89$, and $\beta = 0.6$ have been used in the simulation. The mesh consisted of 50×20, 4-node, quadrilateral membrane elements. The uncertainty on the solution for this mesh, measured by the relative rate of convergence of the reactions at loading points on increasing the element density, is less then 0.01.

Figures 3(*a*) and 3(*b*) show principal membrane force fields (PKII, in kPa cm) plotted on the actual deformed configuration of the membrane. Slack regions, *S*, are inactive and characterized by zero principal membrane forces in both diagrams; wrinkled regions, *W*, have only one nonzero principal membrane force; and taut regions, *T*, are in biaxial tension. Figure 3(*c*) shows directions of the maximum principal Cauchy membrane forces that in the wrinkled regions indicate the directions of the wrinkles. Regions of behavior similar to those of Fig. 3 have been observed in Fig. 1(*a*) in the thin polyethylene membrane tested by Barsotti et al. [1] and subject to similar boundary and loading conditions. The comparison is only qualitative since Fung model is not apt to describe the quantitative response of polyethylene.

6.2 Stretched Membrane-II. Another interesting example of wrinkling is the response to an applied uniaxial tensile strain of a thin rectangular sheet ($25 \times 10 \times 1$ mm) that is laterally clamped. The sheet has been strained up to a global displacement $u_1 = 2.5$ mm. The elastic constants $c = 1.13$ kPa cm, $B = 0.89$ and $\beta = 0.6$ have been used in the simulation. The mesh consisted in 50×20, 4-node, quadrilateral membrane elements. The uncertainty in the solution is less than 0.001.

Figures 4(*a*) and 4(*b*) show principal membrane force fields (PKII, in kPa cm) plotted on the actual deformed configuration of the membrane. Wrinkled regions, W, have only one nonzero principal membrane force; and taut regions, *T*, are in biaxial tension. Figure 4(*c*) shows directions of the maximum principal Cauchy membrane forces that in the wrinkled regions indicate the directions of the wrinkles (horizontal in this sheet). In this problem, wrinkling is generated in the sheet by the action of the clamped boundaries that prevent transverse contraction of the sheet and set up a local biaxial state of stress; the transverse stresses are tensile near the boundary but they would tend to be compressive away from it. These stresses are shown in the diagram of Fig. 5 that has been obtained using the constitutive model Eq. (2) that allows compressive forces to develop in the sheet. As for the previous example, regions of behavior similar to those of Fig. 4 have been observed in Fig. 1(*b*) in the thin polyethylene membrane tested in Ref. [2] and subject to similar boundary and loading conditions.

6.3 Simulation of Reconstructive Surgery. In [7,28] a numerical procedure has been formulated for the simulation of reconstructive surgery of the skin. The procedure refers to surgical operations characterized by the following steps: the incision of the skin; the undermining of a portion of the skin surrounding the cutaneous defect (cancerous and noncancerous growths, burn wounds, lacerations, birth defects); the excision of the defect; the closure and suture of the wound. Similar steps characterize also surgical procedures of cosmetic surgery, such as hair replacement procedures.

Figure 6 describes with simple schematics the numerical procedure proposed in Ref. [28] and refers to a spindle shape excision from a skin flap. Figure 6(*a*) shows one fourth of the skin flap in its natural configuration. The edges of the flap with the thick solid lines are fixed constraints in the simulations and represent the skin surrounding the flap that is still attached to the subcutaneous tissues. The edge of the flap along the vertical axis is a line of symmetry and the horizontal edge, which is also a line of symme-

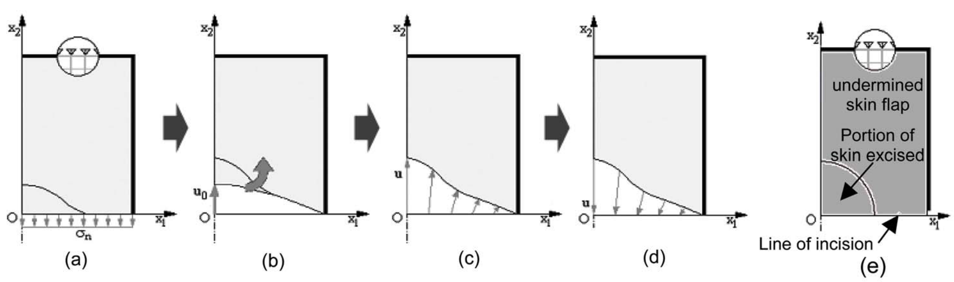


Fig. 6 (*a*)–(*d*) Steps of a procedure of reconstructive surgery on a skin flap. (*a*) Natural configuration of the skin; (*b*) incision and skin partial relaxation; (*c*) excision of a spindle shape portion of the skin; (*d*) closure and suture. (*e*) Excision of a circular portion of a skin flap (membrane forces diagrams after closure are shown in Fig. 7).

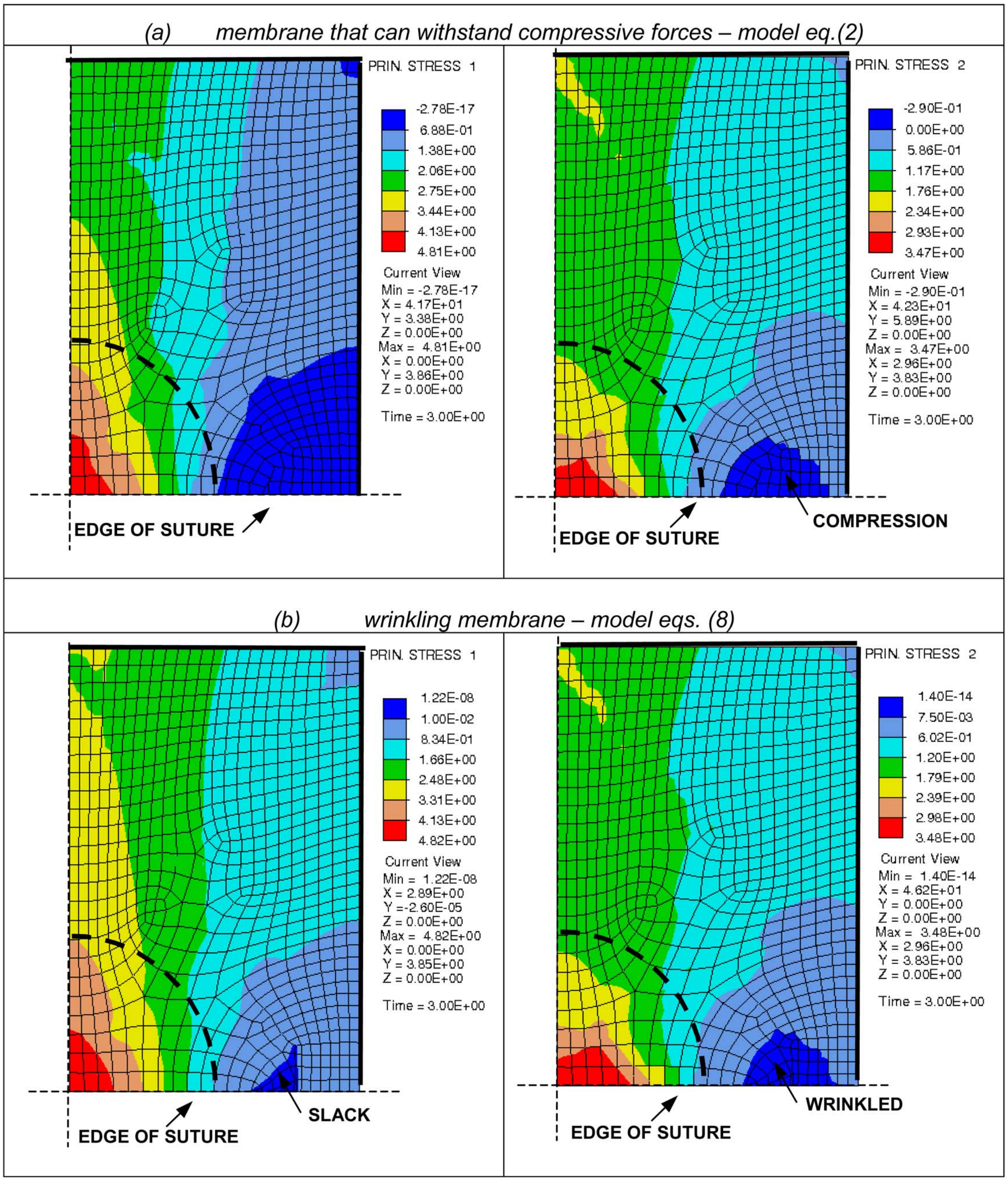


Fig. 7 Principal membrane force fields (PKII, in kPa cm) in a skin flap after the excision of a circular portion of the skin (dashed line) and the closure of the wound edges. The figure compares solutions obtained using the constitutive model (2), diagrams (*a*), which allows compressive stresses in the skin, and the constitutive model (Table 1), diagrams (*b*), which accounts for the wrinkling of the skin. The diagrams highlight the phenomenon of the extrusion of the edges of the wound after suture that occurs in the regions identified as slack and wrinkled.

try, is where the incision is to be performed. In the numerical simulation it is assumed that the phase of undermining of the skin precedes the incision so that the skin flap in its natural configuration is detached from the subcutaneous tissues. The stress and strain fields of the skin in the fictitious natural state of Fig. 6(*a*) are characterized by the isotropic natural stretch and membrane force, λ_n and σ_n. The natural configuration is obtained in the simulations starting from the skin flap in its reference undeformed configuration, with the flap completely detached from the subcutaneous tissues, and applying uniform displacements along the edges that are constrained in Fig. 6(*a*) and uniform tractions σ_n along the line of the future incision. The reference and natural configurations are homothetic because of the isotropy.

The configurations assumed by the skin flap during surgery are obtained by imposing suitable displacements, derived by the natural stretch λ_n, to the external edges of the skin domains in their reference configurations. The reference configuration of (*a*) is the intact, detached, and undeformed skin flap. The reference configuration of (*c*) and (*d*) is the partly excised, detached, and undeformed skin flap. The line of incision-excision is subject to different boundary conditions in the different steps of the operation: the incision and the excision are obtained by assuming a free edge (Fig. 6(*b*) and 6(*c*) showing relaxation of the skin flap due to the partial release of the natural stress); the final closed configuration is obtained by imposing suitable vertical displacements and allowing a free adaptation along x_1 (Fig. 6(*d*)).

Preliminary applications of the numerical procedure, which used the constitutive model (2) and the elastic constants derived in Ref. [7], highlighted the presence of compressive stresses along the wound edges after the suture [28]. Figure 7(*a*) shows principal membrane force fields (PKII, in kPa cm) after the excision of a circular portion of the skin of radius 30 mm and the closure of the wound edges in a skin flap with the same dimensions (120×150 mm) and mechanical characteristics ($c=1.13$ kPa cm, $B=0.89$, $\beta=0.6$, and $\lambda_n=1.1$) of the in vivo flaps analyzed in [7]. Because of symmetry, only one fourth of the flap is shown in the figure. The region where the compressive membrane forces develop are highlighted in the figure. These compressive stresses are inadmissible in a membrane with no flexural stiffness and they are an indication of the phenomenon of the extrusion of the edges or dog-ears formation that has been illustrated in Fig. 1(*c*).

The application of the new constitutive model formulated in this paper allows for a better simulation of the procedure of reconstructive surgery accounting for the wrinkling of the skin. This simulation leads to quantitative estimates of the stresses created in the skin during and after surgery and of the boundaries of the taut, wrinkled, and slack regions.

Figure 7(*b*) shows principal membrane force fields (PKII, in kPa cm), obtained using the constitutive model given in Table 1, after the excision of the circular portion of the skin and the closure of the wound edges. The regions where the skin has wrinkled or is inactive are highlighted in the figure. The diagrams show the large tensile membrane forces created by the closure process, which are one order of magnitude larger than the membrane forces of the skin in its natural state, $\sigma_n=0.348$ kPa cm (σ_n is obtained from the natural stretch using the constitutive model (2)).

The position, shape, and size of the wrinkled and slack regions is influenced by the shapes and sizes of the excision and the skin flap, the natural stretch of the skin, and the closure method. In the analysis performed here, the closure method prescribed purely vertical displacements to all nodal points along the lines of the incision and excision. A study is currently in progress to investigate the effects of the different variables on the response of the skin during and after suture [27].

7 Conclusions

The problem of the wrinkling of plane isotropic membranes characterized by a Fung type constitutive behavior in biaxial tension has been formulated and solved within the framework of finite strain hyperelasticity, following the approach of Pipkin [10,11] and treating the out of plane geometric nonlinearities as constitutive nonlinearities through a modification of the elastic potential.

The model has been implemented into a finite element code and used to analyze the response of partly wrinkled membranes and to simulate a simple procedure of reconstructive surgery that leads to the formation of dog ears. The applications highlight the capability of the model to predict the different regions of behavior that characterize stretched membranes (taut, wrinkled, and slack). Used along with the numerical/experimental procedures previously formulated by the authors to identify the elastic constants of the skin from in vivo tests and to simulate procedures of reconstructive surgery on skin flaps, the model allows for quantitative predictions of the mechanical response of the skin during and after surgery. These predictions account for the effects of the natural tension of the skin on skin wrinkling and therefore overcome the limitations of other models proposed in the literature [29–37]. With some supporting experimental tests, the proposed model could become an important tool for the optimization of the size and shape of the excision in order to reduce the extrusion of the edges and limit the stresses generated by the suture that are known to lessen blood flow and affect wound healing [38].

The model proposed in this paper refers only to isotropic materials, which is a strong limitation when dealing with human skin. The model could be extended to orthotropic materials by restating the wrinkling criteria in terms of principal stresses and strains. The constitutive model would then require a more complicated in vivo identification of a larger number of elastic constants. This could be done using the methodology suggested in Ref. [7]. The extension to curved membranes would also require additional studies to simulate the interactions between the skin and the subcutaneous attachments.

Acknowledgment

The authors are grateful to Vittoria Villa, who recently got her degree in civil engineering at the University of Genova, for helping in the numerical implementation of the proposed model.

References

[1] Barsotti, R., Ligaro, S. S., and Royer-Carfagni, G., 2001, "The Web Bridge," Int. J. Solids Struct., **38**, pp. 5253–5272.
[2] Cerda, K., Ravi-Chandar, K., and Mahadevan, L., 2002, "Wrinkling of an Elastic Sheet Under Tension," Nature **419**, pp. 579–580.
[3] Coleman, W. P., III, Colon, G. A., and Davis, R. S., 1983, *Outpatient Surgery of the Skin*, Medical examination, New Hyde Park, NY
[4] Tong, P., and Fung, Y. C., 1976, "The Stress-Strain Relationships for the Skin," J. Biomech., **9**, pp. 649–657.
[5] Fung, Y. C., 1984, *Biomechanics—Mechanical Properties of Living Tissues*, Springer, New York.
[6] Taylor, R. T., 2001, *Finite Element Analysis Program (FEAP)*, University of California, Berkeley, CA.
[7] Gambarotta, L., Massabò, R., Morbiducci, R., Raposio, G., and Santi, P., 2005, "In Vivo Experimental Testing and Model Identification of Human Scalp Skin," J. Biomech., **38**, pp. 2237–2247.
[8] Ciarlet, P. G., 1988, *Mathematical Elasticity*, North–Holland, Amsterdam.
[9] Holzapfel, G. A., 2001, *Nonlinear Solid Mechanics*, Wiley and Sons, West Sussex.
[10] Pipkin, A. C., 1986, "The Relaxed Energy Density for Isotropic Elastic Membranes," IMA J. Appl. Math., **36**, pp. 85–99.
[11] Pipkin, A. C., 1994, "Relaxed Energy Density for Large Deformations of Membranes," IMA J. Appl. Math., **52**, pp. 297–308.
[12] Wu, C. H., and Canfield, T. R., "Wrinkling in Finite Plane-Stress Theory," Q. Appl. Math., pp. 179–199.
[13] Steigmann, D. J., 1990, "Tension-Field Theory," Proc. R. Soc. London, Ser. A, **429**, pp. 141–173.
[14] Wagner, H., 1970, "Flat Sheet Metal Girder With Very Thin Metal Web," translations J. Vanier, Technical Memorandums of the National Advisory Committee for Aeronautics, Washington, DC, February, vol. 604.
[15] Mansfield, E. H., "Load Transfer Via a Wrinkled Membrane." Proceedings of the Royal Society, London, 316, pp. 269–289.
[16] Mansfield, E. H., "The Bending and Stretching of Plates." Macmillan, New York.
[17] Steigmann, D. J., 2005 "Puncturing a Thin Elastic Sheet," Int. J. Non-Linear Mech., **40**, pp. 255–270.
[18] Im, S., and Kang, S., 1997, "Finite Element Analysis of Wrinkling Membranes," ASME J. Appl. Mech., **64**, pp. 263–269.
[19] Rodemman, D. G., 1991, "Finite-Element Analysis of Wrinkling Membranes," Commun. Appl. Numer. Methods, **7**(4), pp. 299–307.
[20] Rodemman, D. G., Drukker, J., and Oomens, C. W. J., 1987, "The Wrinkling of Thin Membranes. 2 Numerical—Analysis," ASME J. Appl. Mech., **54**, pp. 888–892.
[21] Contri, P., and Schrefler, B. A., 1988, "A Geometrically Nonlinear Finite Element Analysis of Wrinkled Membrane Surfaces by No-Compression Material Model," Commun. Appl. Numer. Methods, **4**, pp. 5–15.
[22] Stanuszek, M., 2003, "FEM Analysis of Large Deformations of Membrane Shells With Wrinkling," Finite Elem. Anal. Design, **39**, pp. 599–618.
[23] Miller, R. K. et al., 1985, "Finite Element Analysis of Partly Wrinkled Membranes," Comput. Struct., **20**, pp. 631–639.
[24] Raposio, E., and Nordström, R. E. A., 1998, "Biomechanical Properties of Scalp Flaps and Their Correlations to Reconstructive and Aesthetic Surgery Procedures," Skin Res. Technol., **4**, pp. 94–98.
[25] Borges, A. F., 1989, "Relaxed Skin Tension Lines," Dermatol. Clin., **7**(1), pp. 169–177.
[26] Cerda, E., 2005, "Mechanics of Scars," J. Biomech., **38**, pp. 1598–1603.
[27] Cavicchi, A., Gambarotta, L., and Massabò, R., 2006, "Computational Modeling of Reconstructive Surgery: The Effects of the Natural Tension on Skin Wrinkling," unpublished.
[28] Gambarotta, L., Massabò, R., and Morbiducci, R., 2002, "A Computational

Model of Scalp Skin for in Vivo Material Characterization and Reconstructive Surgery Simulation," *Proc., of the Fifth World Congress on Computational Mechanics (WCCM V)*, July 7–12, Vienna, Austria, Mang, H. A.; Rammerstorfer, F. G.; and Eberhardsteiner, J., eds., Vienna University of Technology, Austria.

[29] Danielson, D. A., and Natarajan, S., (1975), "Tension Field Theory and Stress in Stretched Skin," J. Biomech., **8**, pp. 135–142.

[30] Chaudhry, H. R., Bukiet, B., Siegel, M., Findley, T., Ritter, A. B., and Guzelsu, N., 1998, "Optimal Patterns for Suturing Wounds," J. Biomech., **31**(7), pp. 653–662.

[31] Lott-Crumpler, D. A., and Chaudhry, H. R., 2001, "Optimal Patterns for Suturing Wounds of Complex Shapes to Foster Healing," J. Biomech., **34**, 51–58.

[32] Retel, V., Vescovo, P., Jacquet, E., Trivaudey, F., Varchon, D., and Burtheret, A., 1991, "Nonlinear Model of Skin Mechanical Behaviour Analysis With Finite Element Method," Skin Res. Technol., **7**, pp. 152–158.

[33] Chretien-Marquet, B., Caillou, V., Brasnu, D. H., Bennaceur, S., and Buisson, T., 1999, "Description of Cutaneous Excision and Suture Using a Mathematical Model," Plast. Reconstr. Surg., **103**(1), pp. 145–150.

[34] Yoshida, H., Tsutsumi, S., Mizunuma, M., and Yanai, A., 2001, "A Surgical Simulation System of Skin Sutures Using a Three-Dimensional Finite Element Method," Clin. Biomech. (Los Angel. Calif.), **16**, pp. 621–626.

[35] Yoshida, H., Tsutsumi, S., Mizunuma, M., and Yanai, A., 2000, "Three-Dimensional Finite Element Analysis of Skin Suture. Part 1: Spindle Model and S-Shaped Modified Model," Med. Eng. Phys., **22**, pp. 481–485.

[36] Kirby, S. D., Wang, B., To, C. W. S., and Lampe, H. B., 1998, "Nonlinear, Three-Dimensional Finite-Element Model of Skin Biomechanics," J. Otolaryngol., **27**, pp. 153–160.

[37] Manios, A., Katsantonis, J., Tosca, A., Skulakis, Ch., and Tsiftsis, D., 1996, "The Finite Element Method as a Research and Teaching Tool in the Analysis of Local Skin Flaps," Dermatol. Surg., **22**, pp. 1029–1033.

[38] Larrabee, W. F., Holloway, G. A., and Sutton, D., 1984, "Wound Tension and Blood Flow in Skin Flaps," Ann. Otol. Rhinol. Laryngol., pp. 93–112.

On the Shear Modulus of Two-Dimensional Liquid Foams: A Theoretical Study of the Effect of Geometrical Disorder

N. P. Kruyt
Department of Mechanical Engineering,
University of Twente,
P.O. Box 217,
7500 AE Enschede,
The Netherlands
e-mail: n.p.kruyt@utwente.nl

The shear modulus of two-dimensional liquid foams in the dry limit of low liquid content has been studied theoretically. The focus is on the effect of geometrical disorder on the shear modulus (besides the influence of surface tension). Various theoretical predictions are formulated that are all based on the assumptions of isotropic geometrical characteristics, incompressible bubbles, and negligible edge curvature. Three of these predictions are based on a transformation of Princen's theory that is strictly valid only for regular hexagonal bubbles. Another prediction takes into account variations in bubble areas by considering the foam as consisting of approximately regular hexagonal bubbles with varying areas. Two other predictions are solely based on the characteristics of the bubble edges. The first of these is based on the assumption of affine movement of bubble vertices, while the second accounts for nonaffine deformation by considering the interaction with neighboring edges. The theoretical predictions for the shear modulus are compared with the result from a single foam simulation. For the single simulation considered, all predictions, except that based on affine movement of bubble vertices, are close to the value obtained from this simulation. [DOI: 10.1115/1.2424241]

Keywords: two-dimensional liquid foams, elastic behavior, shear modulus, surface tension, geometrical disorder

1 Introduction

Liquid foams exhibit complex behaviour, ranging from elastic, to plastic and to viscous [1–3]. Two-dimensional foams (see Fig. 1 for an example), as considered here, constitute a convenient framework for theories, simulations, and experiments [1].

Here the elastic behavior of two-dimensional liquid foams is studied theoretically in the "dry" limit, i.e., foams with small liquid content. It is assumed that the edges of the bubbles are approximately straight. Hence, edge curvature is neglected. This assumption is appropriate for foams that consist mostly of hexagonal bubbles [1].

The compressibility of the gas in the bubbles is of the order of $1/P_{\text{atm}}$ for an ideal gas at pressure P_{atm}, which is generally much smaller than the effect of surface tension, characterized by the Laplace overpressure. Thus, the bubbles can be considered as incompressible, implying that their area remains constant in the two-dimensional case considered here.

Since this study deals with the shear modulus, it is assumed that the bubble and edge geometry in the initial, undeformed state is isotropic. Topological transformations are not modeled, since we are dealing with the elastic behavior.

The geometrical structure of three-dimensional foams has been investigated in Refs. [5–7], using computer simulations. The elastic behavior of specific, regular three-dimensional foam models has been studied in Ref. [8].

The shear modulus of two-dimensional foams consisting of *regular* hexagons has been studied by Ref. [9]. The objective of the present theoretical study is to extend this analysis in order to incorporate effects of geometrical disorder on the shear modulus. Theoretical models that account for these effects are, to the author's knowledge, lacking in the literature. Techniques to develop such theories are inspired by those developed for predicting the elastic properties of granular materials [10–13]. Types of disorder in the foam geometry are: (i) variations in bubble areas, (ii) variations in edge lengths, and (iii) topological disorder, i.e., in the number of edges of the bubbles.

The outline of this study is as follows. First, an overview of micromechanics is presented in Sec. 2. Then a formulation for predicting the shear modulus is given in Sec. 3 that accounts for variations in bubble areas, while edge-based formulations are given in Sec. 4. The main results from a foam simulation are described in Sec. 5. Section 6 gives the comparison of the theoretical predictions with results from the foam simulation, as well as a discussion of the findings of this study and recommendations for further research.

2 Micromechanics

The starting point of the present analysis is the investigation by Ref. [9] (see also Ref. [14]) of the shear modulus μ of foams consisting of regular hexagons, i.e., without any geometrical disorder. His result is

$$\mu_{\text{Princen}} = \gamma \frac{\bar{L}}{\bar{A}} \frac{3}{2} \tag{1}$$

where γ is the surface tension (per unit height in the two-dimensional case considered; so actually γ is a line tension and the line tension per film is 2γ), $\bar{A}$ is the area of each of the hexagons, and $\bar{L}$ is the length of their edges.

For future reference this result is rewritten in two different ways, using the relation $\bar{A} = \frac{3}{2}\sqrt{3}\bar{L}^2$ between area and edge lengths for regular hexagons

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 7, 2005; final manuscript received May 24, 2006. Review conducted by Zhigang Suo.

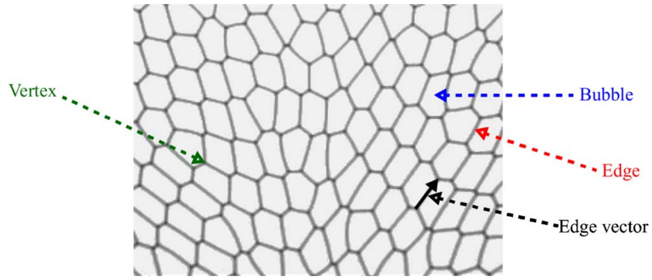


Fig. 1 Example of a two-dimensional foam: bubbles, edges, vertices, and edge vector; adapted from Ref. 4

$$\mu_{\text{Princen},A} = \frac{\gamma}{\sqrt{\bar{A}}}\sqrt[4]{\frac{3}{4}} \tag{2}$$

$$\mu_{\text{Princen},L} = \frac{\gamma}{\bar{L}}\frac{1}{\sqrt{3}} \tag{3}$$

Note that these results can also be expressed in terms of an "equivalent radius" $\bar{R}$, defined through $\bar{A}=\pi\bar{R}^2$. The equivalent radius is widely used, especially for "wet" foams at different liquid contents, but its use is not convenient in the present context.

For two-dimensional foams, each vertex corresponds to three edges, see Fig. 1. From Euler's equation for planar graphs, it follows that the number of bubbles, N_{bubble}, and the number of physical edges, N_{edge}, are related by (disregarding boundary effects) [1]

$$N_{\text{edge}} = 3N_{\text{bubble}} \tag{4}$$

Micromechanics deals with the relation between the macroscale, continuum level, and the microscale level of bubbles, see the schematic in Fig. 2 ([15,16]). Considering elastic behavior, the relevant mechanical quantities at the macroscale level are the stress tensor σ_{ij} and the strain tensor ε_{ij}. For foams, the relevant quantities at the microscale level are the forces (and their changes) due to surface tension in the edges and the edge and bubble geometry (and their changes, i.e., the deformations).

For foams with straight edges and constant surface tension the constitutive relation at the microscale level (see also Fig. 2) that relates force in the edge to the foam geometry is given by

$$f_i^e = 2\gamma\frac{l_i^e}{\sqrt{l_k^e l_k^e}} \tag{5}$$

where f_i^e is the force in the edge due to surface tension and l_i^e is the "edge vector" connecting the two vertices of edge e (see also Fig. 1). Note that the summation convention is employed, by which a summation is implied over repeated *sub*scripts (not for superscripts).

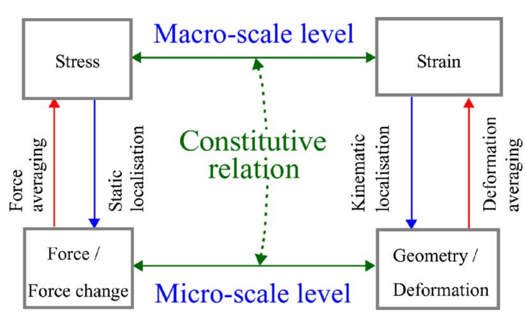


Fig. 2 Schematic of micromechanics: constitutive relations at macroscale and microscale levels, averaging and localization; adapted from Ref. 15

The associated force increment df_i^e is given by

$$df_i^e = \frac{2\gamma}{\sqrt{l_k^e l_k^e}}\left[dl_i^e - \frac{l_i^e l_j^e}{l_l^e l_l^e}dl_j^e\right] \tag{6}$$

where dl_i^e is the change of the edge vector, i.e., the edge deformation at the microscale level.

The stress tensor (excluding the isotropic contribution due to the gas inside the bubble) can be expressed in terms of the forces in the edges by averaging procedures (denoted by the arrow "force averaging" in Fig. 2). The resulting micromechanical expression for the stress tensor is ([3,10,17–20])

$$\sigma_{ij} = \frac{1}{A}\sum_{e\in E} f_i^e l_j^e \tag{7}$$

where A is the total area occupied by the bubbles and the summation is over the edges e in the set of edges E of the foam.

Analogous averaging procedures (denoted by the arrow "deformation averaging" in Fig. 2) have been proposed for the deformation, resulting in micromechanical expressions for the (elastic) strain tensor ([10,18,19,4,21,22]; for an overview see Ref. [23]).

The equilibrium conditions for the vertices, in combination with the microscale level constitutive relation (5), imply the so-called Plateau rule: the angles between the edges incident to a vertex must be equal to 120 deg. Since the edge curvature is neglected, this means that the angle between edge vectors l_i^e at a vertex must also be equal to 120 deg.

An important objective of micromechanics is to obtain constitutive relations at the macroscale level, in terms of quantities at the microscale level. To this end, additional assumptions are required. For example, by expressing the deformation at the microscale in terms of the macroscale strain (see the arrow "kinematic localisation" in Fig. 2), the deformation is obtained. Through the microscale constitutive relation, the microscale forces are found. Using the micromechanical expression for the stress tensor (see the arrow "force averaging" in Fig. 2), the stress tensor at the macroscale is obtained in terms of the strain tensor, thus yielding a micromechanically based continuum constitutive relation.

In micromechanical studies, kinematic localization assumptions are mostly used. In principle, it is also possible to obtain micromechanically based continuum constitutive relations by formulating the forces at the microscale level in terms of the macroscale stress (see the arrow "static localization" in Fig. 2). For granular materials, both procedures, kinematic and static localization, have been used successfully, see, for example, Ref. [11].

However, for foams the procedure based on static localisation fails, since relations (5) and (6) between (increment of) force and (deformation of) geometry are not bijective: given the force, the length of the edge is indeterminate. The procedures developed in the sequel will therefore be based on kinematic localization assumptions.

Since the shear modulus is investigated here, the imposed strain increment tensor $d\varepsilon_{ij}$ is isochoric

$$d\varepsilon_{ij} = d\varepsilon\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} \tag{8}$$

where $d\varepsilon$ is a magnitude of the applied strain increment.

Combining Eqs. (5) and (7), we obtain the stress tensor σ_{ij} in terms of the edge vectors l_i^e

$$\sigma_{ij} = \frac{2\gamma}{A}\sum_{e\in E}\frac{l_i^e l_j^e}{\sqrt{l_k^e l_k^e}} \tag{9}$$

From Eq. (9) it follows that in isochoric deformation, i.e., with constant A, the stress increment $d\sigma_{ij}$ is given by

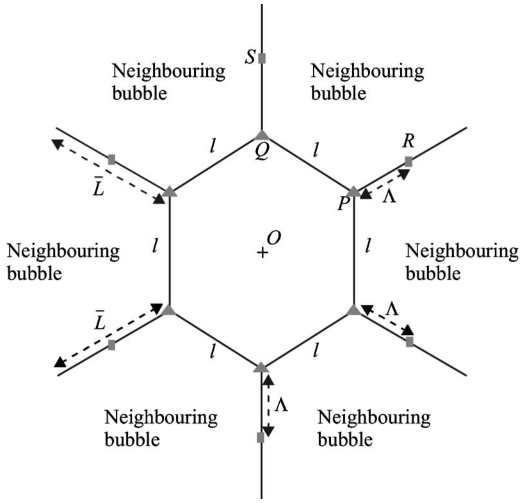


Fig. 3 Geometry of a regular hexagonal bubble

$$d\sigma_{ij} = \frac{2\gamma}{A} \sum_{e \in E} \frac{1}{\sqrt{l_l^e l_l^e}} \left[I_{ik} l_j^e + l_i^e I_{jk} - \frac{l_i^e l_j^e l_k^e}{l_m^e l_m^e} \right] dl_k^e \tag{10}$$

where I_{ij} is the identity tensor.

By definition, the shear modulus μ satisfies (using Eq. (8))

$$d\sigma_{11} - d\sigma_{22} = 2\mu[d\varepsilon_{11} - d\varepsilon_{22}] = 4\mu d\varepsilon \tag{11}$$

Although the elastic properties can also be obtained from changes with strain of the total perimeter of the foam edges [24], and hence, from the change in energy (relative to that in the base state; the energy in the base state has been studied in Ref. [25]), it is noted in Ref. [8] that it is generally easier and more accurate to determine the elastic properties directly from the changes of stress with strain, since the relative changes in energy are very small.

3 Area-Based Formulation

In this section a formulation is developed for taking into account the effect on the shear modulus of disorder in bubble areas. The formulation consists of three steps: (i) a description of the simplified geometry, (ii) an analysis of the stress response of a single bubble (with a bubble size that differs from the average bubble size) to deformation that is imposed on its neighboring bubbles, and (iii) averaging of this stress response over all bubble sizes in order to obtain the effective shear modulus of the foam.

3.1 Simplified Geometry. To enable analytical manipulations, strong simplifications of the foam geometry are required. Here, all bubbles are considered to be regular hexagons (as sketched in Fig. 3) with varying area A_0 and for each bubble its edges are straight and have equal length, although the length may vary from bubble to bubble. This length l_0 is directly related to the area A_0 of the bubble

$$l_0 = \sqrt{\frac{2A_0}{3\sqrt{3}}} \tag{12}$$

Thus, the type of disorder considered is fairly restricted: edge-length and topological disorder are not taken into account. Although a bubble with a specific orientation is shown in Fig. 3, it is assumed that the bubble orientations are randomly distributed in the isotropic state considered.

The behavior of a single bubble, with area A_0, will be considered that is immersed in a "matrix" (in the terminology of the study of heterogeneous elastic materials [26,27]) of approximately hexagonal bubbles whose area is equal to the average area, $\bar{A}$. The corresponding edge length is denoted by $\bar{L}$, i.e., $\bar{L} = \sqrt{2\bar{A}/(3\sqrt{3})}$. This geometrical model is sketched in Fig. 3. It is an idealisation that does not necessarily correspond to a compatible (constructable) foam.

Considering the symmetries present, the configuration is described by the coordinates of points $P=(x,y)^T$, $Q=(0,y+h)^T$, $R=(X,Y)^T$, and $S=(0,Z)^T$. These points are denoted by the gray triangles (P,Q) and squares (R,S) in Fig. 3. The area of the hexagonal bubble is given by $A=4xy+2xh$.

In the initial, undeformed configuration we have $x_0 = l_0 \cos 30$ deg, $y_0 = l_0/2$, $h_0 = l_0 \sin 30$ deg while points R and S are then $R=(x_0,y_0)^T + \Lambda_0(\cos 30 \text{ deg}, \sin 30 \text{ deg})^T$ and $S=(0,y_0+h_0+\Lambda_0)^T$, where Λ_0 is the value of Λ in the undeformed configuration.

The central assumption made in the current analysis is that points of the neighboring bubbles, in this case the points R and S, move affinely with the imposed strain given by Eq. (8). Thus, to first order in ε, $X=X_0[1+d\varepsilon]$, $Y=Y_0[1-d\varepsilon]$, and $Z=Z_0[1-d\varepsilon]$, where X_0, Y_0, and Z_0 are the coordinates X, Y, and Z in the undeformed configuration.

The choice of the affinely moving points determines the value of Λ_0. In general, the vertices will not move affinely, but in the elastic range the movement of the bubble centers is generally closer to affine than that of bubble vertices [28]. Thus, the midpoints of the edges of neighboring bubbles (denoted by the gray squares in Fig. 3) are assumed to move affinely. This implies the following choice for Λ_0

$$\Lambda_0 = \frac{\bar{L}}{2} \tag{13}$$

3.2 Analysis of Single Bubble Response. The response of the single bubble under consideration to the deformation imposed on the neighboring bubbles is analyzed. As explained in the previous section, this imposed deformation is defined through the affine movement of the midpoints of edges of neighboring bubbles. The vertices P and Q of the bubble under consideration will generally not move affinely. Their movement is determined from the requirement of static equilibrium of the vertices P and Q, i.e., the edges must satisfy the Plateau rule that the angle between the edges equals 120 deg. The Plateau rules for vertices P and Q require that

$$\tan 30 \text{ deg} = \frac{h}{x} \quad \tan 30 \text{ deg} = \frac{Y-y}{X-x} \tag{14}$$

An additional restriction on the movement of these vertices is that the bubble is incompressible, and hence its area is constant, $A=A_0$, or

$$4xy + 2xh = A_0 \tag{15}$$

The resulting equations for the positions of the vertices can be solved analytically. In linearized terms, the displacements du, dv, and dw are determined from

$$x = x_0 + du, \quad y = y_0 + dv, \quad h = h_0 + dw \tag{16}$$

After some algebra we obtain

$$du = \frac{\sqrt{3}}{2} l_0 d\varepsilon, \quad dv = l_0 d\varepsilon, \quad dw = \frac{1}{2} l_0 d\varepsilon \tag{17}$$

From these displacements of the vertices, the increments of the edge vectors dl_i^e can be determined.

3.3 Averaging. In this section the effective shear modulus will be determined from the response of the single bubbles to imposed deformation, as analyzed in the previous section.

First, the expression (10) for the stress increment is rewritten into an equivalent form that is based on bubble stresses

$$A d\sigma_{ij} = \sum_b A_b d\sigma_{ij}^b, \quad d\sigma_{ij}^b = \frac{\gamma}{A_b} \sum_{e \in E(b)} \frac{1}{\sqrt{l_l^e l_l^e}} \left[I_{ik} l_j^e + l_i^e I_{jk} - \frac{l_i^e l_j^e l_k^e}{l_m^e l_m^e} \right] dl_k^e \tag{18}$$

where the summation is over all bubbles b in the foam, A_b is the area of bubble b, $d\sigma_{ij}^b$ is the increment of the stress tensor of bubble b, and $E(b)$ is the set of all edges of bubble b. Note that each physical edge is counted twice in this formulation; hence the factor 2 from Eq. (10) is absent here.

In terms of the probability density function $r(A_0)$ for the bubble areas, Eq. (18) can be expressed as

$$N_{\text{bubble}} \bar{A} d\sigma_{ij} = N_{\text{bubble}} \int_0^\infty r(A_0) A_0 d\bar{\sigma}_{ij}(A_0) dA_0 \tag{19}$$

where $\bar{A}$ is the average bubble area and $d\bar{\sigma}_{ij}(A_0)$ is the average stress increment of bubbles with area corresponding to A_0.

Using the results of the analysis of Sec. 3.2 for the edge deformations dl_i^e, $d\bar{\sigma}_{ij}(A_0)$ can be determined from Eq. (18). It follows after some lengthy algebra that

$$d\bar{\sigma}_{11}(A_0) - d\bar{\sigma}_{22}(A_0) = 4\gamma \frac{[l_0(A_0) + \Lambda_0]}{A_0} d\varepsilon \tag{20}$$

Using the choice (13) for the affine points and the relation (12) between edge length and bubble area for regular hexagons, the resulting effective shear modulus μ becomes, see Eq. (11), after some algebra

$$\mu = \frac{\gamma}{\sqrt{\bar{A}}} \sqrt{\frac{2}{3\sqrt{3}}} \left[\frac{1}{2} + \int_0^\infty r(A_0) \sqrt{\frac{A_0}{\bar{A}}} dA_0 \right] \tag{21}$$

For narrow probability density functions $r(A_0)$, the result (21) can be approximated by

$$\mu \cong \frac{\gamma}{\sqrt{\bar{A}}} \sqrt[4]{\frac{3}{4}} \left[1 - \frac{1}{12} \frac{\sigma^2(A_0)}{\bar{A}^2} \right] \tag{22}$$

where $\sigma^2(A_0)$ is the variance of the bubble areas. In the case of a foam without area variations, we retrieve the result (2).

4 Edge-Based Formulations

Two edge-based formulations will be given. Both are based on first analyzing the response of a single edge to imposed deformation. The response of the set of edges that constitutes the foam is then obtained by averaging the responses of the individual edges in the foam. Since these formulations are based on edges, they account for disorder in edge lengths. No restrictions are posed on the absence of topological disorder, i.e., the number of edges of the bubbles may vary from bubble to bubble.

The first formulation is based on affine deformation of all vertices. This leads to an overprediction of the shear modulus, since this deformation does not lead to force equilibrium of the vertices, i.e., the edges do not satisfy the Plateau rules.

In the second, more refined formulation it is assumed that, for an edge under consideration, the midpoints of the neighboring edges move affinely (as was also assumed in Sec. 3). The movement of the vertices of the edge under consideration are determined from equilibrium conditions, i.e., the Plateau rules. Thus, this formulation takes into account neighboring-edge corrections.

Consider a straight edge in the undeformed configuration with length l_0 and orientation θ_0 (without loss of generality, -90 deg $\leqslant \theta_0 \leqslant 90$ deg), as shown in Fig. 4. Hence, the edge vector is given by $l_i = l_0(\cos\theta_0, \sin\theta_0)^T$. Assume that the edge length and edge orientation can be considered as independent random variables. Thus, the probability density function $q(l_0, \theta_0)$ for edge length and edge orientation is given by $q(l_0, \theta_0) = p(l_0)E(\theta_0)$, where $p(l_0)$ is the probability density function for edge lengths and $E(\theta_0)$ is the edge orientation distribution function (the contact distribution function for granular materials [29]). For the isotropic states considered

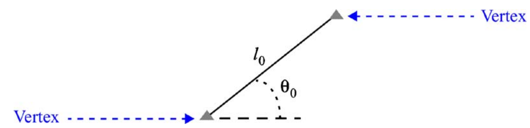


Fig. 4 Length and orientation of a straight edge

$$E(\theta_0) = 1/\pi \tag{23}$$

In terms of the probability density functions $p(l_0)$ and $E(\theta)$, the increment of the stress tensor becomes, using Eqs. (4) and (10)

$$d\sigma_{ij} = \frac{6\gamma}{\bar{A}} \int_0^\infty p(l_0) dl_0 \int_{-\pi/2}^{\pi/2} E(\theta_0) \frac{1}{l_0} \left[I_{ik} l_j + l_i I_{jk} - \frac{l_i l_j l_k}{l_0^2} \right] \overline{dl_k}(l_0, \theta_0) d\theta_0 \tag{24}$$

where $\overline{dl_i}(l_0, \theta_0)$ is the average increment of the edge vectors of edges with length l_0 and orientation θ_0 in the undeformed configuration. Note that $l_i = l_i(l_0, \theta_0)$.

A kinematic localization assumption (see Fig. 2) for $\overline{dl_i}(l_0, \theta_0)$ is therefore required to obtain the stress increment, and hence, the shear modulus, from Eq. (24). The kinematic localization assumption will be addressed in two ways in the following two subsections.

The edge deformation $dl_i(l_0, \theta_0)$ can be decomposed into components in the direction of the edge, $dl_T(l_0, \theta_0)$, and the direction perpendicular to the edge, $dl_N(l_0, \theta_0)$

$$dl_i(l_0, \theta_0) = dl_T(l_0, \theta_0) T_i(\theta_0) + dl_N(l_0, \theta_0) N_i(\theta_0) \tag{25}$$

where $T_i(\theta_0) = (\cos\theta_0, \sin\theta_0)^T$ and $N(\theta_0) = (-\sin\theta_0, \cos\theta_0)^T$.

4.1 Affine Deformation. The first kinematic localization assumption is based on the assumption that the vertices move affinely. The corresponding kinematic localization assumption for the edge deformation is

$$\overline{dl_i}(l_0, \theta_0) = d\varepsilon_{ij} l_j(l_0, \theta_0) \tag{26}$$

The equivalent components $dl_T(l_0, \theta_0)$ and $dl_N(l_0, \theta_0)$, see Eq. (25), are

$$dl_T(l_0, \theta_0) = d\varepsilon l_0 \cos 2\theta_0 \quad dl_N(l_0, \theta_0) = -d\varepsilon l_0 \sin 2\theta_0 \tag{27}$$

After some algebra it follows from Eqs. (8), (24), and (26) that

$$d\sigma_{11} - d\sigma_{22} = \frac{9\gamma}{\bar{A}} d\varepsilon \int_0^\infty p(l_0) l_0 dl_0 \tag{28}$$

The corresponding shear modulus, see Eq. (11), is given by

$$\mu = \gamma \frac{\bar{L}}{\bar{A}} \frac{9}{4} \tag{29}$$

This shear modulus is $\mu = \mu_{\text{Princen}} \frac{3}{2}$, where μ_{Princen} is the shear modulus for regular hexagonal bubbles, see Eq. (1). Hence, the affine kinematic localisation assumption leads to an overestimation of the shear modulus for regular hexagons.

4.2 Neighboring-Edge Corrections. The second kinematic localization assumption takes into account neighboring-edge corrections. It is similar to the methods developed in Refs. [12,13] for granular materials.

Consider an edge with orientation θ_0 and length l_0, see Fig. 5. Its vertices (black triangles in Fig. 5) are denoted by V_1 and V_2.

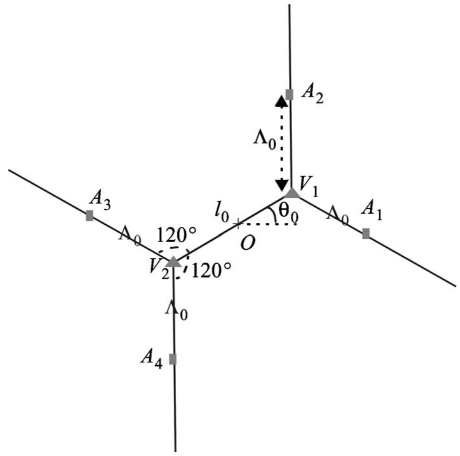


Fig. 5 Geometry of the neighborhood of an edge

Also shown are the other edges incident on vertices V_1 and V_2. On these other edges, points A_1 to A_4 are indicated in Fig. 5 by gray squares.

In the initial, undeformed configuration the coordinates of $V_1 = (l_0/2 \cos\theta_0, l_0/2 \sin\theta_0)^T$, while those of V_2 and V_3 are $V_{2,3} = V_1 + \Lambda_0(\cos(\theta_0 \pm 60 \text{ deg}), \sin(\theta_0 \pm 60 \text{ deg}))^T$, respectively, as follows from the Plateau rules. The coordinates of the other points V_2, A_3, and A_4 follow from the symmetry present.

To investigate the deformation of such an edge, it is assumed that the points A_1–A_4 move affinely according to Eq. (8). The positions of these points is fixed with Λ_0 according to Eq. (13), i.e., it is assumed again that the midpoints of edges move approximately affinely.

The displacement of vertex V_1 is defined as $(u,v)^T$. From symmetry it follows that the displacement of vertex V_2 equals $(-u, -v)^T$. These displacements are determined by Plateau's rule for vertex V_1 (equilibrium conditions). This gives, after some lengthy algebra, complicated expressions for $u(l_0,\theta_0;d\varepsilon,\Lambda_0)$ and $v(l_0,\theta_0;d\varepsilon,\Lambda_0)$. The corresponding increment $dl_i(l_0,\theta_0;d\varepsilon,\Lambda_0)$ of the edge vector is $dl_i(l_0,\theta_0;d\varepsilon,\Lambda_0)=2(u(l_0,\theta_0;d\varepsilon,\Lambda_0), v(l_0,\theta_0;d\varepsilon,\Lambda_0))^T$. This constitutes the kinematic localization assumption of the present formulation. The results for $dl_T(l_0,\theta_0)$ and $dl_N(l_0,\theta_0)$, as defined in Eq. (25), are, after substituting Eq. (13)

$$dl_T(l_0,\theta_0) = \cos 2\theta_0 [d\varepsilon l_0]\left(1+\frac{\bar{L}}{l_0}\right)$$

$$dl_N(l_0,\theta_0) = -\sin 2\theta_0 [d\varepsilon l_0]\left(\frac{1-\dfrac{\bar{L}}{l_0}}{1+2\dfrac{\bar{L}}{l_0}}\right) \tag{30}$$

where $\bar{L}$ is the average edge length.

With the kinematic localisation assumption thus formulated, the stress increment can be calculated from Eq. (24). After some lengthy algebra the effective shear modulus μ becomes, see also Eq. (11)

$$\mu = \frac{3}{4}\frac{\gamma}{\bar{A}}\int_0^\infty p(l_0)\left[\frac{2\bar{L}^2+\bar{L}l_0+3l_0^2}{2\bar{L}+l_0}\right]dl_0 \tag{31}$$

For narrow probability density functions $p(l_0)$, this result can be approximated by

$$\mu \cong \gamma\frac{\bar{L}}{\bar{A}}\frac{3}{2}\left[1+\frac{2}{9}\frac{\sigma^2(l_0)}{\bar{L}^2}\right] \tag{32}$$

where $\sigma^2(l_0)$ is the variance of the edge lengths. For zero variance, this result is identical to that of Princen [9] for regular hexagons, see Eq. (1).

5 Results From a Foam Simulation

The results of a foam simulation [30], using the Surface Evolver [31], are analyzed here. The Surface Evolver is a computer program that can minimize the energy of a surface subject to constraints. The simulated two-dimensional foam consists of 100 bubbles with equal area. Of these bubbles, 14% has five edges, 72% has six edges (hexagons), and 14% has seven edges. Periodic boundaries are used in the simulation to minimize boundary effects.

The edge orientation distribution function $E(\theta_0)$ determined from the simulation is shown in Fig. 6, together with that for an isotropic foam geometry. Considering the small number of edges present in the foam, the foam can be considered as isotropic.

For the average length of the edges we have $\bar{L}=0.6271\sqrt{\bar{A}}$; compare with $\bar{L}=0.6204\sqrt{A}$ for regular hexagons. The probability density function of edge lengths $p(l_0)$ is shown in Fig. 7. The variance in edge length equals $\sigma^2(l_0)=0.35\bar{L}^2$.

The macroscale strain was determined from the deformation of the periodic box (that corresponds to the periodic boundaries), while the stress increment was determined from Eq. (10). Thus the shear modulus could be determined from the foam simulation, see Table 1.

To further study kinematic localization assumptions (see Fig. 2), edge deformation $dl_T(l_0,\theta_0)$ and $dl_N(l_0,\theta_0)$, as defined in Eq. (25), obtained from the foam simulation are investigated. Since the number of edges is fairly small, $\overline{dl_T}(l_0,\theta_0)$ and $\overline{dl_N}(l_0,\theta_0)$ cannot be reliably estimated. Instead the edge deformation $\overline{dl_T}(\theta_0)$ and $\overline{dl_N}(\theta_0)$, i.e., $dl_T(l_0,\theta_0)$ and $dl_N(l_0,\theta_0)$ averaged over all edge lengths l_0, have been determined, see Fig. 8. These quantities represent the dependence of edge deformation on edge orientation.

Figure 8 shows that $\overline{dl_N}(\theta_0)$ is much smaller than $\overline{dl_T}(\theta_0)$, which means that deformation by changes of edge length is dominant over deformation by changes in edge orientation. The formulation based on affine edge deformation, see Sec. 4.1, predicts that $\overline{dl_T}(\theta_0)$ and $\overline{dl_N}(\theta_0)$ are of the same order of magnitude, see Eq. (27). Also shown in Fig. 8 is a fit (denoted by the solid line) of the data points dl_N^e with the form $\zeta\cos 2\theta_0[d\varepsilon\bar{L}]$. This form for $\overline{dl_T}(\theta_0)$ is predicted by both edge formulations, see Eqs. (27) and (30). The fitted value for $\zeta=1.96$, while the value predicted by Eq. (30) is $\zeta=2$, i.e., there is good agreement between simulation and prediction. For small variations in l_0, it also follows from Eq. (30) that $\overline{dl_N}(\theta_0)\cong 0$.

6 Discussion

The effect of geometrical disorder on the shear modulus of two-dimensional liquid foams has been studied theoretically. Various theoretical predictions have been formulated that are founded on the assumptions of isotropic geometrical characteristics, incompressible bubbles and negligible edge curvature.

Three of these predictions, (1)–(3), are based on Princen's theory that is strictly valid only for regular hexagons. Prediction (1) is obtained by replacing the area and edge length in Princen's result by corresponding average values. Prediction (2) is obtained by using the relation between area and edge length for regular hexagons to express Princen's result in terms of average bubble area. Similarly, prediction (3) is obtained by expressing Princen's

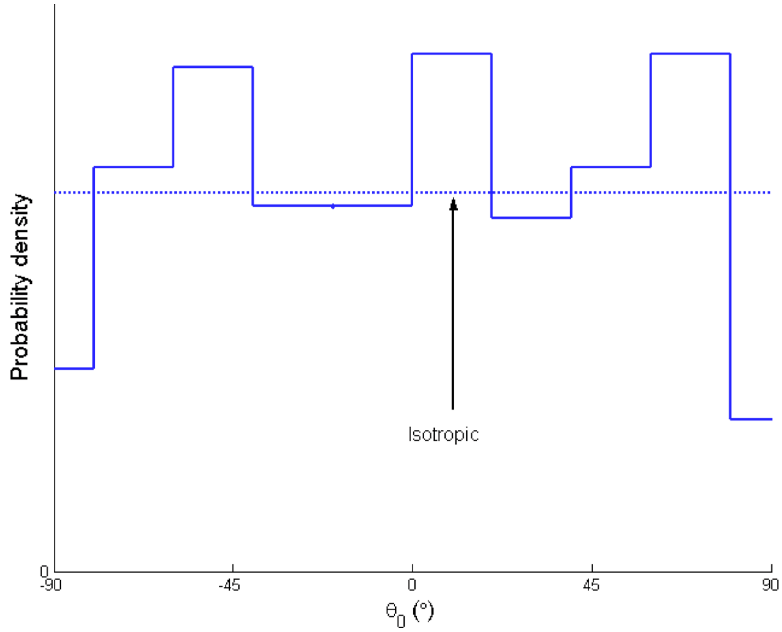


Fig. 6 Edge orientation distribution function $E(\theta_0)$ from simulation; dotted line: $E(\theta_0)$ for isotropic foam geometry

result in terms of average edge length. Note that these transformations are not based on a theoretical framework.

Prediction (21) is obtained by taking into account variations of bubble areas. It has been assumed that the bubbles can be approximated by regular hexagons with varying areas. Predictions (29) is based on affine deformation of all vertices. Prediction (32) includes corrections to affine deformation by taking into account neighboring edges.

The numerical values for the nondimensional shear moduli (scaled such that it equals 1 for prediction (1)) from the single foam simulation reported in Sec. 5 and according to the developed theories are given in Table 1, together with a short description of the underlying theory and a characterization of the type of disorder that has been modeled. Except for the prediction based on the assumption of affine deformation, Eq. (29), all predictions are within 4% of the value from this single simulation. Since the bubble areas are all equal in the simulation, predictions (2) and (21) are identical. Thus, it is not possible to verify the (range of) validity of prediction (21) from the present foam simulation.

It is somewhat surprising that predictions (1)–(3) and (21), which are all based on assumption that the foam consists exclusively of hexagonal bubbles (no topological disorder, i.e., no variation in the number of edges of the bubbles), are so close to the value from the simulation, since 28% of the bubbles in the simulation were not hexagonal. Note that no assumption regarding topological disorder (or its absence) has been made in the derivation of predictions (29) and (32).

Given the characteristics of the analyzed *single* foam simulation

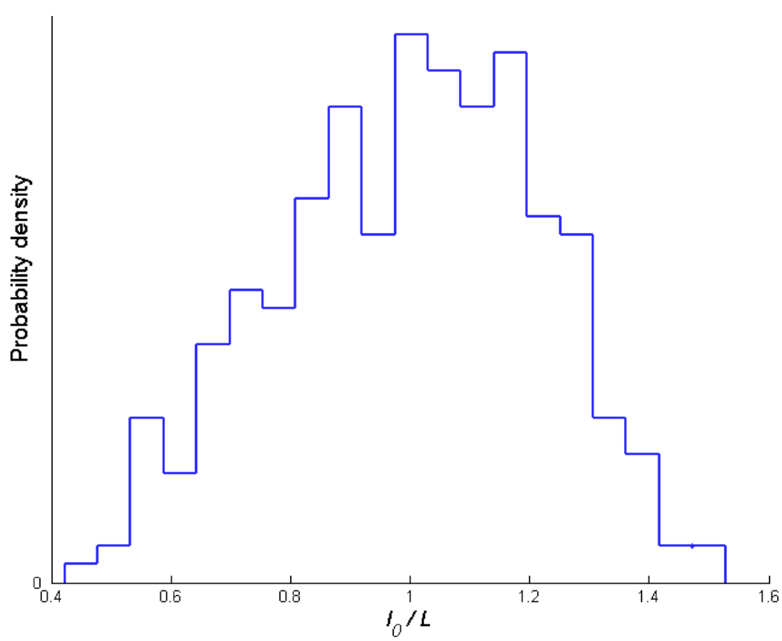


Fig. 7 Probability density function $p(l_0)$ of edge lengths from the foam simulation

Table 1 Comparison of shear moduli

Prediction	$\frac{(\mu\bar{A})}{\gamma\bar{L}}\frac{2}{3}$	Description	Type of disorder
Simulation	0.9701	Numerical	Edge length, topological
μ_{Princen}, Eq. (1)	1.0000	Based on regular hexagons	None
$\mu_{\mathrm{Princen},\ A}$, Eq. (2)	0.9893	Based on regular hexagons; result expressed in average bubble area	None
$\mu_{\mathrm{Princen},\ L}$, Eq. (3)	0.9787	Based on regular hexagons; result expressed in average edge length	None
$\mu_{\mathrm{area\ formulation}}$, Eq. (21)	0.9893	Based on regular hexagons with variation in bubble areas	Bubble area
μ_{affine}, Eq. (29)	1.5000	Based on edges, with affine deformation	Edge length (topological)
$\mu_{\mathrm{neighbor\ correction}}$, Eq. (32)	1.0107	Based on edges, with neighboring edge corrections	Edge length (topological)

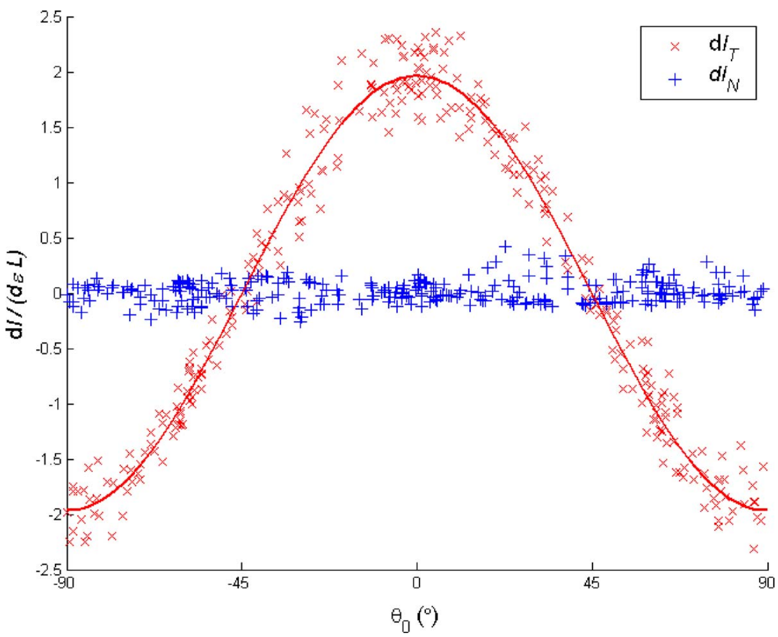


Fig. 8 Dependence of edge deformation on edge orientation; markers: data from foam simulation, solid line: fit of the form $\zeta \cos 2\theta_0[d\varepsilon L]$ to the data

(equal bubble areas, limited edge-length disorder, limited topological disorder), it is not possible to conclude decisively which theoretical prediction is most suitable for accounting for the effect of geometrical disorder on the shear modulus of two-dimensional foams. To determine the range of validity of the developed theories, more experimental data and/or foam simulations are required.

For further research it is recommended to: (1) perform experiments to check the validity of the theories developed here and to guide further theoretical developments; (2) perform detailed foam simulations with more pronounced geometrical disorder in: (i) edge lengths, (ii) bubble areas, and (iii) topology, i.e., in the number of edges per bubble; (3) perform foam simulations with larger numbers of bubbles, so detailed statistics of the edge deformations can be obtained to investigate kinematic localization assumptions (see Fig. 2), such as Eqs. (27) and (30); (4) investigate the use of "self-consistent approximations" [26,27] to derive expressions for the effective shear modulus; (5) extend the developed methods to foams with curved edges (instead of only considering straight edges, as done here); (6) investigate the use of variational principles, as employed in Ref. [11] for the prediction of rigorous upper and lower bounds to the elastic moduli of granular materials.

Acknowledgment

The author acknowledges the hospitality, tutorials and discussions on the complex physics of foams by F. Graner during the author's stay at Laboratoire de Spectrométrie Physique, Université Joseph Fourier (Grenoble, France). S. J. Cox (Institute of Mathematical and Physical Sciences, University of Wales, Aberystwyth, United Kingdom) kindly provided the results of the foam simulation. S. J. Cox and F. Graner are thanked for their comments on earlier versions of this paper.

References

[1] Weaire, D. L., and Hutzler, S., 1999, *The Physics of Foams*, Oxford University Press, Oxford.
[2] Dollet, B., Elias, F., Quilliet, C., Raufaste, C., Aubouy, M., and Graner, F.,

2005, "Two-Dimensional Flow of Foam Around Obstacles: Force Measurements," Phys. Rev. E, **71**, 031403.
[3] Janiaud, E., and Graner, F., 2005, "Foam in a Two-Dimensional Couette Shear: A Local Measurement of Bubble Deformation," J. Fluid Mech., **532**, pp. 243–267.
[4] Asipauskas, M., Aubouy, M., Glazier, J. A., Graner, F., and Jiang, Y., 2003, "A Texture Tensor to Quantify Deformations: The Example of Two-Dimensional Flowing Foams," Granular Matter, **5**, pp. 71–74.
[5] Kraynik, A. M., Reinelt, D. A., and van Swol, F., 2003, "Structure of Random Monodisperse Foams," Phys. Rev. E, **67**, p. 031403.
[6] Kraynik, A. M., Reinelt, D. A., and van Swol, F., 2004, "Structure of Random Foams," Phys. Rev. Lett., **93**, p. 208301.
[7] Kraynik, A. M., Reinelt, D. A., and van Swol, F., 2005, "Structure of Random Bidisperse Foams," Colloids Surf., A, **263**, pp. 11–17.
[8] Kraynik, A. M., and Reinelt, D. A., 1996, "Linear Elastic Behaviour of Dry Soap Films," J. Colloid Interface Sci., **181**, pp. 511–520.
[9] Princen, H. M., 1983, "Rheology of Foams and Highly Concentrated Emulsions," J. Colloid Interface Sci., **91**, pp. 160–175.
[10] Kruyt, N. P., and Rothenburg, L., 1996, "Micromechanical Definition of the Strain Tensor for Granular Materials," Trans. ASME, J. Appl. Mech., **63**, pp. 706–711.
[11] Kruyt, N. P., and Rothenburg, L., 2002, "Micromechanical Bounds for the Elastic Moduli of Granular Materials," Int. J. Solids Struct., **39**, pp. 311–324.
[12] Agnolin, I., Jenkins, J. T., and La Ragione, L., 2006, "A Continuum Theory for a Random Array of Identical, Elastic, Frictional Disks," Mech. Mater., **38**, pp. 687–701.
[13] Jenkins, J. T., Johnson, D., La Ragione, L., and Makse, H., 2005, "Fluctuations and the Effective Moduli of an Isotropic, Random Aggregate of Identical, Frictionless Spheres," J. Mech. Phys. Solids, **53**, pp. 197–225.
[14] Khan, S. A., and Armstrong, R. C., 1986, "The Rheology of Foams: Theory for Dry Foams," J. Non-Newtonian Fluid Mech., **22**, pp. 1–22.
[15] Cambou, B., 1993, "From Global to Local Variables in Granular Materials," *Powders & Grains 93*, C. Thornton, ed. Balkema, Rotterdam, The Netherlands, pp. 73–86.
[16] Kruyt, N. P., and Rothenburg, L., 2004, "Kinematic and Static Assumptions for Homogenization in Micromechanics of Granular Materials," Mech. Mater., **36**, pp. 1157–1173.
[17] Drescher, A., and Josselin de Jong, G., 1972, "Photoelastic Verification of a Mechanical Model for the Flow of a Granular Materials," J. Mech. Phys. Solids, **20**, pp. 337–351.
[18] Kruyt, N. P., 2003, "Statics and Kinematics of Discrete Cosserat-Type Granular Materials," Int. J. Solids Struct., **40**, pp. 511–534.
[19] Bagi, K., 1996, "Stress and Strain in Granular Assemblies," Mech. Mater., **22**, pp. 165–177.
[20] Bagi, K., 1999, "Microstructural Stress Tensor of Granular Assemblies With Volume Forces," Trans. ASME, J. Appl. Mech., **66**, pp. 934–936.
[21] Aubouy, M., Jiang, Y., Glazier, J. A., and Graner, F., 2003, "A Texture Tensor to Quantify Deformations," Granular Matter, **5**, pp. 67–70.
[22] Goldenberg, C., and Goldhirsch, I., 2006, "Continuum Mechanics for Small Systems and Fine Resolutions," *Handbook of Theoretical and Computational Nanotechnology*, M. Rieth and W. Schommers, eds., American Scientific, Stevenson Ranch, CA, Vol. 4, pp. 329–386.
[23] Bagi, K., 2006, "Analysis of Microstructural Strain Tensors for Granular Assemblies," Int. J. Solids Struct., **43**, pp. 3166–3184.
[24] Stamenovic, D., 1991, "The Shear Modulus of Foamlike Structures," Trans. ASME, J. Appl. Mech., **58**, pp. 288–289.
[25] Graner, F., Jiang, Y., Janiaud, E., and Flament, C., 2001, "Equilibrium States and Ground State of Two-Dimensional Fluid Foams," Phys. Rev. E, **63**, pp. 011402.
[26] Beran, M. J., 1968, *Statistical Continuum Theories*, Interscience Publishers, New York.
[27] Torquato, S., 2002, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*, Springer, Berlin.
[28] Graner, F., 2005, personal communication.
[29] Horne, M. R., 1965, "The Behaviour of an Assembly of Rotound, Rigid, Cohesionless Particles, I and II," Proc. R. Soc. London, Ser. A, **286**, pp. 62–97.
[30] Cox, S. J., 2005, personal communication.
[31] Brakke, K., 1992, "The Surface Evolver," Exp. Math., **1**, pp. 141–165; see also http://www.susqu.edu/brakke/evolver/evolver.html.

Analytical Solution for Size-Dependent Elastic Field of a Nanoscale Circular Inhomogeneity

L. Tian

R. K. N. D. Rajapakse[1]
Mem. ASME
e-mail: rajapakse@mech.ubc.ca

Department of Mechanical Engineering,
The University of British Columbia,
Vancouver, BC, Canada V6T 1Z4

Two-dimensional elastic field of a nanoscale circular hole/inhomogeneity in an infinite matrix under arbitrary remote loading and a uniform eigenstrain in the inhomogeneity is investigated. The Gurtin–Murdoch surface/interface elasticity model is applied to take into account the surface/interface stress effects. A closed-form analytical solution is obtained by using the complex potential function method of Muskhelishvili. Selected numerical results are presented to investigate the size dependency of the elastic field and the effects of surface elastic moduli and residual surface stress. Stress state is found to depend on the radius of the inhomogeneity/hole, surface elastic constants, surface residual stress, and magnitude of far-field loading. [DOI: 10.1115/1.2424242]

1 Introduction

Nanotechnology involves analysis, design, and fabrication of devices and structures with at least one of the overall dimensions in the nanometer range. Common examples are nanowires, nanotubes, nanoplates, etc. In addition, composite materials with nanoscale inclusions/tubes as reinforcing elements represent new developments in advanced materials technology. Such materials can be tailored to have unique mechanical, electronic, and optical properties [1,2]. Study of material properties at the nanoscale and elastic field of nanostructures and nanocomposites is important to further development of nanotechnology.

Atomic force microscopy is used to determine the Young's modulus, strength, and toughness of nanorods and nanotubes [3] and elastic modulus of nanowires [4]. Unlike for bulk material elements, the measured effective elastic properties are highly dependent on the size of the nanostructure. This size dependency of properties at nanoscale can be understood by incorporating the effect of surface and interfacial energies. In the classical case, bulk elastic energy controls the behavior of a material element, whereas for nanoscale elements the influence of surface effects becomes dominant due to the high surface/volume ratio. Atoms near the surface/interface have different equilibrium positions and energy from that of the bulk. Therefore, surface energy effects can be substantial for nanostructures.

Size dependency of effective properties and elastic field of nanoscale elements can be explained by accounting for the effect of surface stresses that are associated with the excess free energy of a surface. The roots of surface stress theory lies in the thermodynamics of solid surfaces as formulated by Gibbs [5]. In the traditional continuum mechanics approach the effects of surface free energy is neglected as the bulk elastic energy dominates the response. Gurtin and Murdoch [6] and Gurtin et al. [7] developed a general theoretical framework for a continuum with surface stresses. Cammarata [8] and Miller and Shenoy [9] have applied the surface stress model to examine several issues involving nanoscale elements. Sharma et al. [10] and Sharma and Ganti [11] studied the size-dependent elastic state and Eshelby's tensor of nanoinhomogeneities, respectively, by applying the Gurtin–Murdoch model. Duan et al. [12] recently examined size dependency of the effective bulk and shear moduli of a solid containing nanoinhomogeneities.

In this paper, the Gurtin–Murdoch model is applied to study the two-dimensional size-dependent elastic field of a nanoscale circular inhomogeneity in an infinite elastic plane under remote biaxial loading. This is a fundamental problem in nanomechanics and the solution is currently not available. Muskhelishvili [13] provided an analytical solution for the classical circular inclusion problem in the absence of surface stress effects by using complex potential functions. Muskhelishvili's approach is extended in the present paper to a nanoscale inhomogeneity in which the surface stresses give rise to a nonclassical boundary condition. The closed-form analytic solution derived in this paper corresponds to a uniform applied biaxial traction field and a uniform eigenstrain. It should be noted that Sharma and Ganti [11] presented the solution for radially symmetric elastic field of a circular cylindrical inclusion and spherically symmetric elastic field of a spherical inclusion subjected to uniform eigenstrain. The present solution reduces to the Sharma and Ganti solution [11] for the radially symmetric loading case. The derivation of the closed-form analytic solution is presented in the next section and is followed by numerical results for selected cases.

2 Elastic Field of Circular Inhomogeneity

Consider an infinite plane, containing a single circular inhomogeneity of nanoscale, subjected to uniform far-field tractions and a prescribed uniform eigenstrain in the inhomogeneity (Fig. 1). The materials of the matrix and inhomogeneity are assumed to be linearly elastic, homogeneous, and isotropic with Lamé constants λ_M, μ_M and λ_I, μ_I, respectively. Note that the subscripts M and I are used to identify quantities associated with the matrix and inhomogeneity, respectively. The inhomogeneity, with its center at the origin of the coordinate system, has radius R_0.

For plane problems, the displacement and stress components in Cartesian and polar coordinates, (x_1, x_2, x_3) and (r, θ, x_3) respectively, can be expressed in terms of two analytic functions $\phi(z)$ and $\psi(z)$ as [13]

$$2\mu(u_r + \mathrm{i}u_\theta) = e^{-\mathrm{i}\theta}[\kappa\phi(z) - z\overline{\phi'(z)} - \overline{\psi(z)}] \tag{1}$$

$$\sigma_{rr} + \sigma_{\theta\theta} = 2[\phi'(z) + \overline{\phi'(z)}] \tag{2}$$

$$\sigma_{rr} - \mathrm{i}\sigma_{r\theta} = \phi'(z) + \overline{\phi'(z)} - e^{\mathrm{i}2\theta}[\bar{z}\phi''(z) + \psi'(z)] \tag{3}$$

[1]Corresponding author.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 15, 2005; final manuscript received May 30, 2006. Review conducted by Zhigang Suo.

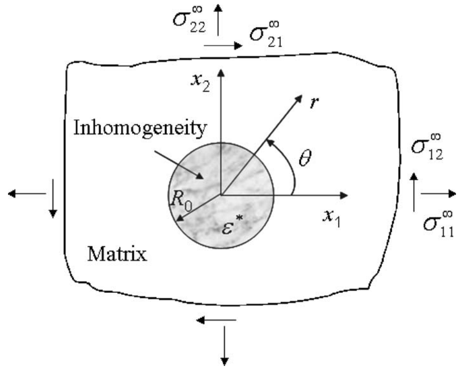


Fig. 1 Nanoscale inhomogeneity in an infinite matrix

$$2\mu(u_1 + \mathrm{i}u_2) = \kappa\phi(z) - z\overline{\phi'(z)} - \overline{\psi(z)} \tag{4}$$

$$\sigma_{11} + \sigma_{22} = 2[\phi'(z) + \overline{\phi'(z)}] \tag{5}$$

$$\sigma_{22} - \sigma_{11} + 2\mathrm{i}\sigma_{12} = 2[\bar{z}\phi''(z) + \psi'(z)] \tag{6}$$

where u_i and σ_{ij} are displacement and stress components, respectively; $z=x_1+\mathrm{i}x_2=re^{\mathrm{i}\theta}$; $\kappa=3-4\nu$ for plane strain and $(3-\nu)/(1+\nu)$ for plane stress, and μ and ν are the shear modulus and Poisson's ratio, respectively.

Assume that the inhomogeneity is perfectly bonded to the matrix. Then the displacements are continuous at the interface

$$(u_r + \mathrm{i}u_\theta)_M = (u_r + \mathrm{i}u_\theta)_I + (u_r + \mathrm{i}u_\theta)_I^*, \quad \text{at} \quad r = R_0 \tag{7}$$

where the last term is the displacement induced by the prescribed uniform dilatational eigenstrain ε^*, i.e., $\varepsilon_{11}^*=\varepsilon_{22}^*=\varepsilon^*$, and $(u_r+\mathrm{i}u_\theta)_I^*|_{r=R_0}=R_0\varepsilon^*$.

The following field equations and constitutive relations can be established for an isotropic material based on the theory proposed by Gurtin and Murdoch [6] and Gurtin et al. [7]:

in the bulk (matrix and inhomogeneity)

$$\sigma_{ij,j}^B = 0, \quad \sigma_{ij}^B = \lambda\delta_{ij}\varepsilon_{kk} + 2\mu\varepsilon_{ij} \tag{8}$$

on the interface of inhomogeneity

$$[\![\sigma_{\beta\alpha}^B n_\beta]\!] + \sigma_{\beta\alpha,\beta}^S = 0 \tag{9}$$

$$[\![\sigma_{ji}^B n_i n_j]\!] = \sigma_{\alpha\beta}^S k_{\alpha\beta} \tag{10}$$

$$\sigma_{\beta\alpha}^S = \tau^0\delta_{\beta\alpha} + 2(\mu^S - \tau^0)\varepsilon_{\beta\alpha} + (\lambda^S + \tau^0)\varepsilon_{\gamma\gamma}\delta_{\beta\alpha} \tag{11}$$

where superscripts B and S are used to denote the quantities corresponding to bulk (matrix and inhomogeneity) and surface/interface of inhomogeneity; σ_{ij} and ε_{ij} denote stress and strain, respectively; λ and μ are Lamé constants; δ_{ij} is the Kronecker delta; n_i is the normal vector on the surface/interface; λ^S and μ^S are the surface Lamé constants; τ^0 is the residual surface stress under unstrained conditions; $k_{\alpha\beta}$ is the curvature tensor of the surface or interface; and $[\![*]\!]=(*)_M-(*)_I$ denotes the jump across the matrix–inhomogeneity interface.

The first equation in Eq. (8) is the equilibrium equation of the bulk materials, whereas Eqs. (9) and (10) are the equilibrium equations of the surface/interface. Equation (11) is the surface constitutive equation. It should be noted that the surface stress tensor is a two-dimensional (2D) quantity and the strain normal to the surface is excluded in Eq. (11). Thus, the Greek indices take the value of 1 or 2, while Latin subscripts adopt values from 1 to 3.

In the (r, θ, x_3) coordinate (x_3 is the direction perpendicular to the (r, θ) plane), Eqs. (9) and (10) can be written as follows:

on the surface/interface

in θ direction

$$[\![\sigma_{r\theta}^B]\!] + \frac{\partial\sigma_{\theta\theta}^S}{R_0\partial\theta} + \frac{\partial\sigma_{3\theta}^S}{\partial x_3} = 0 \tag{12}$$

in x_3 direction

$$[\![\sigma_{r3}^B]\!] + \frac{\partial\sigma_{33}^S}{\partial x_3} + \frac{\partial\sigma_{\theta 3}^S}{R_0\partial\theta} = 0 \tag{13}$$

in r direction

$$[\![\sigma_{rr}^B]\!] = \frac{\sigma_{\theta\theta}^S}{R_0} \tag{14}$$

For plane problems, $\sigma_{r3}^B=\sigma_{3\theta}^S=\sigma_{\theta 3}^S=0$ and the derivatives with respect to x_3 are zero. Thus Eq. (13) is automatically satisfied. Equations (12) and (14) can be expressed in the following complex variable form

$$[\![\sigma_{rr}^B - \mathrm{i}\sigma_{r\theta}^B]\!] = \frac{\sigma_{\theta\theta}^S}{R_0} + \mathrm{i}\frac{\partial\sigma_{\theta\theta}^S}{R_0\partial\theta} \tag{15}$$

The left-hand side of Eq. (15) can be written in terms of potential functions by using Eq. (3). For the right-hand side, the surface stress is

$$\sigma_{\theta\theta}^S = \tau^0 + 2(\mu^S - \tau^0)\varepsilon_{\theta\theta} + (\lambda^S + \tau^0)(\varepsilon_{33} + \varepsilon_{\theta\theta}) \tag{16}$$

The elastic strain $\varepsilon_{\theta\theta}$ at the surface can be obtained from the following equations

$$\varepsilon_{\theta\theta} + \varepsilon_{rr} = \left.\frac{1}{\Gamma}(\phi'(z) + \overline{\phi'(z)})\right|_{r=R_0} \tag{17}$$

$$\varepsilon_{\theta\theta} - \varepsilon_{rr} + \mathrm{i}\varepsilon_{r\theta} = \left.\frac{1}{\mu}[\bar{z}\phi''(z) + \psi'(z)]e^{\mathrm{i}2\theta}\right|_{r=R_0} \tag{18}$$

Here, $\Gamma=\lambda+\mu$ for plane strain and $\Gamma=\mu(3\lambda+2\mu)/(\lambda+2\mu)$ for plane stress. Therefore

$$\varepsilon_{\theta\theta} = \frac{1}{2\Gamma}[\phi'(z) + \overline{\phi'(z)}] + \frac{1}{4\mu}[\bar{z}\phi''(z) + \psi'(z)]e^{2\mathrm{i}\theta} + \frac{1}{4\mu}[z\overline{\phi''(z)} + \left.\overline{\varphi'(z)}]e^{-2\mathrm{i}\theta}\right|_{r=R_0} \tag{19}$$

Note that the strain $\varepsilon_{\theta\theta}$ is continuous at the interface because of the continuous displacement at the interface. In the following derivation, the strain $\varepsilon_{\theta\theta}$ is calculated from the matrix. Note that $\varepsilon_{33}=0$ for plane strain and $\varepsilon_{33}=\nu(\varepsilon_{rr}+\varepsilon_{\theta\theta})/(\nu-1)|_{r=R_0}$ for plane stress. Because of the discontinuity of ε_{33} at the interface, the mean strain is used, i.e.

$$\varepsilon_{33} = \tfrac{1}{2}[(\varepsilon_{33})_M + (\varepsilon_{33})_I]_{r=R_0} \tag{20}$$

$(\varepsilon_{33})_M$ can be obtained by using Eq. (17). Due to the eigenstrain effect, $(\varepsilon_{33})_I$ is

$$(\varepsilon_{33})_I = \frac{\nu_I}{\nu_I - 1}[(\varepsilon_{rr} + \varepsilon_{\theta\theta})_I^{\mathrm{e}} + 2\varepsilon^*] \tag{21}$$

where $(\varepsilon_{rr}+\varepsilon_{\theta\theta})_I^{\mathrm{e}}$ is the elastic strain in the inhomogeneity which can be obtained from Eq. (17).

Introduce a complex variable ξ such that

$$z = m(\xi) = R_0\xi, \quad \xi = r_0 e^{i\theta} \tag{22}$$

Note that at the interface $r_0=1$, and $\phi(\xi)=\phi(m(\xi))$ and $\psi(\xi)=\psi(m(\xi))$. The complex potentials $\phi_M(\xi)$, $\phi_I(\xi)$, $\psi_M(\xi)$ and $\psi_I(\xi)$ corresponding to the matrix and inhomogeneity are now expanded into the following Laurent series form

$$\phi_M(\xi) = A\xi + \sum_{n=1}^{\infty} A_n\xi^{-n}, \quad \psi_M(\xi) = B\xi + \sum_{n=1}^{\infty} B_n\xi^{-n} \tag{23}$$

$$\phi_I(\xi)=\sum_{n=1}^{\infty}F_n\xi^n, \quad \psi_I(\xi)=\sum_{n=1}^{\infty}G_n\xi^n \tag{24}$$

Note that the constant terms have been omitted in Eqs. (23) and (24) since they represent the rigid body displacements and have no effect on the stress distribution. In Eq. (23), A and B are real and complex numbers, respectively, characterizing the remote stress field. In view of Eqs. (5) and (6)

$$\sigma_{11}^{\infty}+\sigma_{22}^{\infty}=\frac{4A}{R_0}, \quad \sigma_{22}^{\infty}-\sigma_{11}^{\infty}+2\mathrm{i}\sigma_{12}^{\infty}=\frac{2B}{R_0} \tag{25}$$

where σ_{11}^{∞}, σ_{22}^{∞}, and σ_{12}^{∞} are the far-field stresses which in the case of our study are assumed to be

$$\sigma_{22}^{\infty}=\sigma_0, \quad \sigma_{11}^{\infty}=a\sigma_0, \quad \sigma_{12}^{\infty}=0 \tag{26}$$

Here, a is a real number characterizing the loading ratio $\sigma_{11}^{\infty}/\sigma_{22}^{\infty}$. In the present case, both A and B are real numbers. Note that $a=1$ and $a=0$ refer to biaxial and uniaxial loadings, respectively, while $a=-1$ represents pure shear loading.

The following set of equations are obtained by substituting Eqs. (1), (23), and (24) into Eq. (7) and equating the coefficients of $e^{\mathrm{i}n\theta}$

$$0=-2\bar{F}_2, \quad \frac{\mu_{\mathrm{I}}}{\mu_M}(\kappa_M A_1-B)=-3\bar{F}_3-\bar{G}_1$$

$$\frac{\mu_I}{\mu_M}\kappa_M A_{n+1}=-(n+3)\bar{F}_{n+3}-\bar{G}_{n+1}, \quad (n=1,2,3\ldots) \tag{27}$$

$$\frac{\mu_I}{\mu_M}(\kappa_M A-A-\bar{B}_1)=\kappa_{\mathrm{I}}F_1-\bar{F}_1+2\mu_{\mathrm{I}}R_0\varepsilon^*, \quad -\frac{\mu_I}{\mu_M}\bar{B}_2=\kappa_{\mathrm{I}}F_2$$

$$\frac{\mu_I}{\mu_M}(n\bar{A}_n-\bar{B}_{n+2})=\kappa_{\mathrm{I}}F_{n+2}, \quad (n=1,2,3\ldots) \tag{28}$$

Substitution of Eqs. (3), (16), (23), and (24) into Eq. (15) yields

$$\begin{aligned}2A+B_1-(F_1+\bar{F}_1)&=\tau^0+4\Lambda_5\mu_{\mathrm{I}}R_0\varepsilon^*-\Lambda_1B_1-\Lambda_1\bar{B}_1\\&\quad+2(\Lambda_1+\Lambda_3)\Lambda_2A+\Lambda_4\Lambda_5(F_1+\bar{F}_1)\end{aligned}$$

$$2B_2-2\bar{F}_2=\Lambda_1B_2+4\Lambda_4\Lambda_5\bar{F}_2$$

$$\begin{aligned}-\bar{A}_1-B+3F_3+G_1&=-\Lambda_1B-2\Lambda_1\bar{A}_1+3\Lambda_1\bar{B}_3+(\Lambda_1+\Lambda_3)\Lambda_2\bar{A}_1\\&\quad-3\Lambda_4\Lambda_5F_3\end{aligned}$$

$$\begin{aligned}-A_1+B_3-\bar{F}_3&=\Lambda_1\bar{B}+2\Lambda_1A_1-3\Lambda_1B_3-(\Lambda_1+\Lambda_3)\Lambda_2A_1\\&\quad+3\Lambda_4\Lambda_5\bar{F}_3\end{aligned}$$

$$\begin{aligned}-\bar{A}_{n+1}+(n+3)F_{n+3}+G_{n+1}&=-\Lambda_1(n+1)(n+2)\bar{A}_{n+1}\\&\quad+\Lambda_2(\Lambda_1+\Lambda_3)(n+1)\bar{A}_{n+1}\\&\quad+\Lambda_1(n+3)\bar{B}_{n+3}\\&\quad-(n+3)\Lambda_4\Lambda_5F_{n+3}, \quad (n=1,2,3\ldots)\end{aligned}$$

$$\begin{aligned}-(n+1)A_{n+1}+B_{n+3}-\bar{F}_{n+3}&=\Lambda_1(n+1)(n+2)A_{n+1}\\&\quad-\Lambda_2(\Lambda_1+\Lambda_3)(n+1)A_{n+1}\\&\quad-\Lambda_1(n+3)B_{n+3}\\&\quad+(n+3)\Lambda_4\Lambda_5\bar{F}_{n+3}, \quad (n=1,2,3\ldots)\end{aligned} \tag{29}$$

where

$$\Lambda_1=\frac{K^S}{4\mu_MR_0}, \quad \Lambda_2=\frac{2\mu_M}{\Gamma_M}, \quad K^S=2\mu^S+\lambda^S-\tau^0, \quad \Lambda_4=\frac{2\mu_{\mathrm{I}}}{\Gamma_{\mathrm{I}}}$$

$\Lambda_3=0$ for plane strain and

$$\frac{\lambda^S+\tau^0}{4\mu_MR_0}\frac{\nu_M}{\nu_M-1}$$

for plane stress

$\Lambda_5=0$ for plane strain and

$$\frac{\lambda^S+\tau^0}{4\mu_{\mathrm{I}}R_0}\frac{\nu_{\mathrm{I}}}{\nu_{\mathrm{I}}-1} \tag{30}$$

for plane stress

The solution of Eqs. (27)–(29) yields

$$A_1=\frac{(-c_1+c_2-c_3+3\Lambda_4\Lambda_5c_6)B}{d_1+\kappa_Mc_2+d_2+3\Lambda_4\Lambda_5d_3}, \quad A_{n+1}=0, \quad (n=1,2,3\ldots) \tag{31}$$

$$B_1=\frac{\Lambda_7\tau^0+2(2\Lambda_5\Lambda_7-c_7)\mu_{\mathrm{I}}R_0\varepsilon^*+2(\Lambda_7c_4+\Lambda_6\Lambda_8c_7)A}{\Lambda_6c_7+\Lambda_7c_1}, \quad B_2=0$$

$$B_3=\frac{(c_4+c_5-c_3+3\Lambda_4\Lambda_5c_6)B}{d_1+\kappa_Mc_2+d_2+3\Lambda_4\Lambda_5d_3}, \quad B_{n+3}=0, \quad (n=1,2,3\ldots) \tag{32}$$

$$F_1=\frac{-\Lambda_6\tau^0+2\Lambda_6(c_1\Lambda_8-c_4)A-2(2\Lambda_5\Lambda_6+c_1)\mu_{\mathrm{I}}R_0\varepsilon^*}{2(\Lambda_6c_7+\Lambda_7c_1)}, \quad F_2=0$$

$$F_3=\frac{\Lambda_6}{\kappa_{\mathrm{I}}}\frac{(-c_1+c_2-c_4-c_5)B}{d_1+\kappa_Mc_2+d_2+3\Lambda_4\Lambda_5d_3}, \quad F_{n+3}=0, \quad (n=1,2,3\ldots) \tag{33}$$

$$\begin{aligned}G_1&=-\Lambda_6\left[\frac{\kappa_M(-c_1+c_2-c_3+3\Lambda_4\Lambda_5c_6)B}{d_1+\kappa_Mc_2+d_2+3\Lambda_4\Lambda_5d_3}-B\right]\\&\quad-\frac{3\Lambda_6}{\kappa_{\mathrm{I}}}\frac{(-c_1+c_2-c_4-c_5)B}{d_1+\kappa_Mc_2+d_2+3\Lambda_4\Lambda_5d_3}\\G_{n+1}&=0, \quad (n=1,2,3\ldots)\end{aligned} \tag{34}$$

where

$$\Lambda_6=\mu_{\mathrm{I}}/\mu_M, \quad \Lambda_7=(\kappa_{\mathrm{I}}-1)/2, \quad \Lambda_8=(\kappa_M-1)/2$$

$$c_1=1+2\Lambda_1, \quad c_2=\Lambda_6(1+3\Lambda_1)$$

$$c_3=\Lambda_6(1-\Lambda_1-\Lambda_6)/\kappa_{\mathrm{I}}, \quad c_4=\Lambda_2(\Lambda_1+\Lambda_3)-1$$

$$c_5=\Lambda_6[1-\Lambda_2(\Lambda_1+\Lambda_3)+2\Lambda_1+\kappa_M\Lambda_1]$$

$$c_6=\Lambda_6(\Lambda_6-1)/\kappa_{\mathrm{I}}, \quad c_7=1+\Lambda_4\Lambda_5$$

$$d_1=1+4\Lambda_1+\Lambda_2(\Lambda_1+\Lambda_3)$$

$$d_2=\Lambda_6[1+\Lambda_1+\Lambda_2(\Lambda_1+\Lambda_3)+\kappa_M\Lambda_6]/\kappa_{\mathrm{I}}$$

$$d_3=\Lambda_6(1+\kappa_M\Lambda_6)/\kappa_{\mathrm{I}}. \tag{35}$$

The complete elastic field is explicitly given by Eqs. (1)–(6) together with Eqs. (23), (24), and (30)–(35). After some manipulation, the stresses along the interface are:

in the matrix

$$\sigma_{\theta\theta}=\frac{2A-B_1}{R_0}+\frac{(B-3B_3)\cos 2\theta}{R_0}$$

$$\sigma_{rr} = \frac{2A + B_1}{R_0} + \frac{(-4A_1 - B + 3B_3)\cos 2\theta}{R_0}$$

$$\sigma_{r\theta} = \frac{(-2A_1 + B + 3B_3)\sin 2\theta}{R_0} \quad (36)$$

in the inhomogeneity

$$\sigma_{\theta\theta} = \frac{2F_1}{R_0} + \frac{(12F_3 + G_1)\cos 2\theta}{R_0}, \quad \sigma_{rr} = \frac{2F_1}{R_0} - \frac{G_1\cos 2\theta}{R_0}$$

$$\sigma_{r\theta} = \frac{(6F_3 + G_1)\sin 2\theta}{R_0} \quad (37)$$

Note that due to the surface/interface stress effects, $\Lambda_1(\Lambda_3, \Lambda_5) \neq 0$ and $\tau^0 \neq 0$, and the stress state depends on the size of the inhomogeneity R_0 as the coefficients A_1, B_1, B_3, F_1, F_3, and G_1 in Eqs. (31)–(34) are nonlinear functions of R_0. When no surface/interface stresses exist, i.e., $\Lambda_1 = \Lambda_3 = \Lambda_5 = 0$ and $\tau^0 = 0$, the above results reduce to the classical solution in which the elastic state is size independent. The shear and radial stresses are equal on either side of the interface in the classical case and

$$\sigma_{r\theta} = \frac{\Lambda_6(1 + \kappa_M)(1 - a)\sigma_0 \sin 2\theta}{2(1 + \kappa_M\Lambda_6)}$$

$$\sigma_{rr} = \frac{\Lambda_6(1 + \Lambda_8)(1 + a)\sigma_0 - 4\mu_I\varepsilon^*}{2(\Lambda_6 + \Lambda_7)} - \frac{\Lambda_6(1 + \kappa_M)(1 - a)\sigma_0\cos 2\theta}{2(1 + \kappa_M\Lambda_7)} \quad (38)$$

Hoop stresses on the interface are different and given by in the matrix

$$\sigma_{\theta\theta} = \frac{(\Lambda_6 + 2\Lambda_7 - \Lambda_6\Lambda_8)(1 + a)\sigma_0 + 4\mu_I\varepsilon^*}{2(\Lambda_6 + \Lambda_7)} + \frac{(4 - 3\Lambda_6 + \kappa_M\Lambda_6)(1 - a)\sigma_0 \cos 2\theta}{2(1 + \kappa_M\Lambda_6)} \quad (39)$$

in the inhomogeneity

$$\sigma_{\theta\theta} = \frac{(\Lambda_6 + \Lambda_6\Lambda_8)(1 + a)\sigma_0 - 4\mu_I\varepsilon^*}{2(\Lambda_6 + \Lambda_7)} + \frac{\Lambda_6(1 + \kappa_M)(1 - a)\sigma_0 \cos 2\theta}{2(1 + \kappa_M\Lambda_6)} \quad (40)$$

As can be seen from Eqs. (38)–(40), the stress state is independent of R_0 in the classical solution.

3 Numerical Results and Discussion

Selected numerical results are presented in this section and the surface elastic constants are obtained from past studies. The plane strain case in which the surface effect is represented by the parameters K^S (or Λ_1) and τ^0 is investigated in the numerical study without loss of any generality. Experiments have been performed to determine the surface stress which has an order of 1 N/m [14]. The embedded atom method was used by Miller and Shenoy [9] and Shenoy [15] to determine the surface elastic constants. Their results indicated that the surface elastic constants depend on the material type and the surface crystal orientation. For example, for Al [1 0 0] surface: $\lambda^S = 3.4939$ N/m, $\mu^S = -5.4251$ N/m, $\tau^0 = 0.5689$ N/m, $K^S = -7.9253$ N/m; while for Al [1 1 1] surface: $\lambda^S = 6.8511$ N/m, $\mu^S = -0.3760$ N/m, $\tau^0 = 0.9108$ N/m, and $K^S = 5.1882$ N/m [9]. Although the surface properties are generally anisotropic, it is assumed that the isotropic case is sufficient to illustrate the main features of the size-dependent response. In the calculations, unless specified otherwise, $\tau^0 = \pm 1$ N/m and $K^S = \pm 10$ N/m.

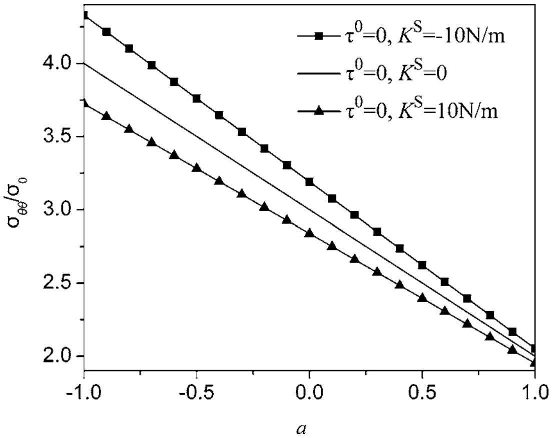


Fig. 2 Variation of stress concentration factor with loading ratio, *a*, for a hole without residual surface stress (R_0=5 nm)

3.1 Infinite Plane With a Circular Hole. An infinite plane of aluminum containing a circular hole under far-field loading is considered. The bulk elastic constants for aluminum are: $\lambda_M = 58.17$ GPa, $\mu_M = 26.13$ GPa [16]. Based on the analysis presented in the previous section, hoop stress (plane strain) at $\theta = 0$ on the hole surface is given by

$$\sigma_{\theta\theta} = (3 - a)\sigma_0 - \left[\frac{\frac{1}{2}(\Lambda_1\Lambda_2 + 2\Lambda_1)(1 + a)}{1 + 2\Lambda_1} + \frac{3(\Lambda_1\Lambda_2 + 2\Lambda_1)(1 - a)}{1 + 4\Lambda_1 + \Lambda_1\Lambda_2}\right]\sigma_0 - \frac{\tau^0/R_0}{1 + 2\Lambda_1} \quad (41)$$

Note that the first term is the classical elasticity result and the last two terms represent the surface stress effects and contain nonlinear terms of R_0. The first two terms are linear with respect to the loading magnitude while the third term is independent of external loading and linear with respect to the residual surface stress τ^0. When $a = 1$, i.e., under radially symmetric loading, the result is the same as that obtained by Sharma and Ganti [11]. Following the classical definition, a stress concentration factor can be defined for a hole by normalizing hoop stress at $\theta = 0$ by the remote loading magnitude when $\tau^0 = 0$.

The effect of the surface elastic constant, $K^S = 2\mu^S + \lambda^S - \tau^0$, on the stress state is first studied for the case $\tau^0 = 0$. In this case based on Eq. (41), the hoop stress concentration factor ($\sigma_{\theta\theta}/\sigma_0$) is independent of the loading value σ_0. Figure 2 shows the stress concentration factor for various values of far-field loading ratio a for a hole with radius $R_0 = 5$ nm. The stress concentration factor varies linearly with a and is slightly increased or decreased compared to the classical result ($\tau^0 = 0$, $K^S = 0$) depending on whether K^S is negative or positive. The influence of K^S appears to be more prominent when a is negative. Figure 3 shows the stress concentration factor for various values of the radius of the hole. The classical solution in which $\tau^0 = 0$ and $K^S = 0$ is, as expected, independent of the radius, while the surface stress effects cause the stress concentration factor to be highly size dependent, especially when the radius is less than 10 nm. The stress concentration factor increases or decreases rapidly when the cavity radius is less than 10 nm depending on whether K^S is negative or positive. The surface stress effect is negligible when the hole radius is over 15 nm and the stress concentration factor is equal to the classical elasticity solution. Similar behavior is observed for other values of a as well.

Figure 4 shows the variation of normalized hoop stress ($\sigma_{\theta\theta}(R_0, \theta)/\sigma_0$) in the circumferential direction along the surface

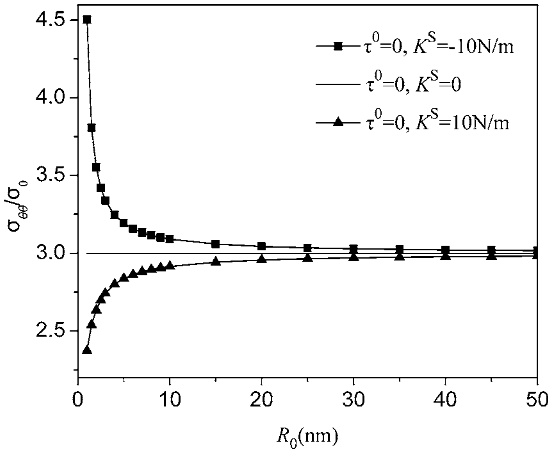


Fig. 3 Variation of stress concentration factor with radius of hole and K^S for a hole without residual surface stress ($a=0$)

of a hole with 5 nm radius under uniaxial loading ($a=0$) when $\tau^0=0$. At $\theta=0$, the normalized stress is reduced for positive K^S and increased for negative K^S when compared to the classical result; however, opposite behavior can be seen at $\theta=\pi/2$. Similar behavior is observed for pure shear loading ($a=-1$). Under biaxial loading ($a=1$), the elastic field is radially symmetric and the hoop stress does not vary along the hole. As can be seen from Fig. 2 for $a=1$, normalized hoop stress increases and decreases for negative and positive K^S values, respectively. Figure 5 shows the variation of normalized hoop stress and radial stress along the positive x_1 direction. The surface stress effect is evident near the hole surface but diminishes quite rapidly as x_1 increases, especially in the case of hoop stress. Figure 5 shows that when compared to the classical solution, the normalized radial stress is larger near the hole surface and slightly smaller far from the hole for positive values of K^S. The opposite behavior is noted for negative values of K^S.

Consider next the influence of residual surface stress, τ^0, on the stress field of a plane containing a hole. When $K^S=0$ (or $\Lambda_1=0$), the hoop stress at the hole surface is

$$\sigma_{\theta\theta}=[(1+a)+2(1-a)\cos 2\theta]\sigma_0-\tau^0/R_0 \tag{42}$$

Note that Eq. (42) applies to both plane stress and plane strain cases. As mentioned previously, the second term on the right-hand side of Eq. (42) is independent of the magnitude of remote loading and corresponds to stress due to the residual surface stress. As σ_0 would appear in the denominator of the second term on the right-hand side of Eq. (42), a stress concentration factor of the classical form is not defined when $\tau^0\neq 0$. It is clear from Eq. (42) that circumferential dependence of hoop stress along the hole surface is given by the classical elasticity solution minus a constant term that is linearly proportional to τ^0 and inversely proportional to hole radius. Similar to the effect of K^S, a negative τ^0 increases hoop stress while a positive τ^0 decreases it.

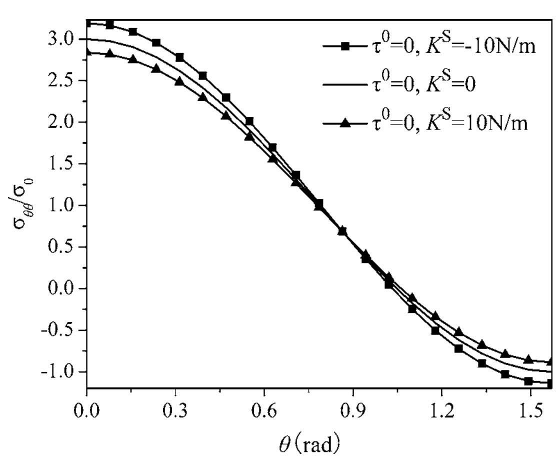


Fig. 4 Circumferential variation of normalized hoop stress along hole surface ($R_0=5$ nm, $a=0$)

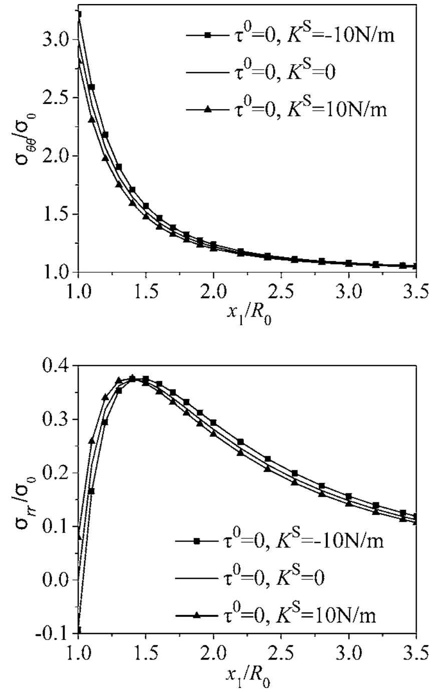


Fig. 5 Variation of normalized hoop and radial stresses along the x_1 direction for different K^S ($R_0=5$ nm, $a=0$)

Let $\sigma^C_{\theta\theta}$ and σ^C_{rr} denote hoop and radial stresses corresponding to the classical elasticity solution respectively. Let $\sigma^S_{\theta\theta}$ and σ^S_{rr} denote hoop and radial stresses due to the residual surface stress. Figure 6 shows the variation of $\sigma^C_{\theta\theta}$ and σ^C_{rr} normalized by σ_0 and $\sigma^S_{\theta\theta}$ and σ^S_{rr} normalized by τ^0/R_0 along the positive x_1 direction under uniaxial loading. Again, the residual surface stress shows a significant influence on the stress field in the vicinity of the hole surface. Its effect is negligible at a distance greater than four times the hole radius. Note that $\sigma^S_{\theta\theta}$ and σ^S_{rr} along x_1 are proportional to

$$-\frac{\tau^0}{R_0}\left(\frac{x_1}{R_0}\right)^{-2}$$

and

$$\frac{\tau^0}{R_0}\left(\frac{x_1}{R_0}\right)^{-2}$$

respectively. Hence, the normalized stress components due to τ^0 shown in Fig. 6 are independent of the radius R_0.

Yang [17] found that the effective shear modulus is size dependent only considering the strain-independent surface stress. Recently, Duan et al. [12] mentioned that such a constant surface stress should not have an influence on the effective shear modulus. This is confirmed by the present study. Equations (36) and (37) show that the shear stress in the matrix and the inhomogeneity are independent of the residual surface stress τ^0, and thus the effective in-plane shear modulus is also independent of τ^0. However, τ^0 has an effect on the effective in-plane Young's modulus.

3.2 Infinite Plane With Circular Inhomogeneity. In recent years, quantum dot and wire structures have attracted considerable attention due to their potential application in nanotechnology [18].

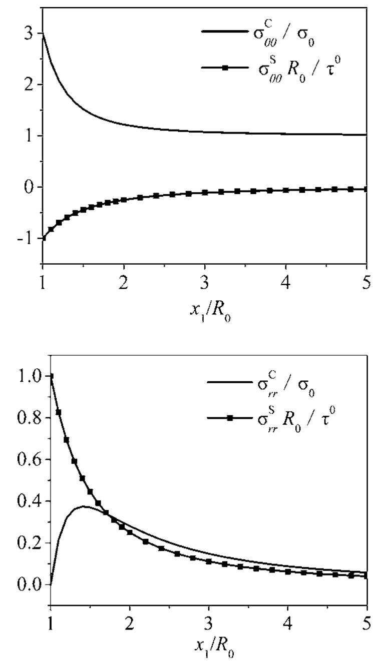


Fig. 6 Variation of normalized hoop and radial stress components along the x_1 direction for a hole with residual surface stress ($\tau^0 \neq 0$, $a=0$)

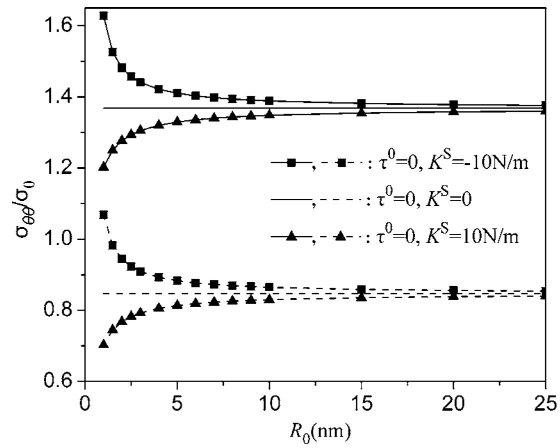


Fig. 8 Variation of normalized interfacial hoop stress at $\theta=0$ with inhomogeneity radius and K^S ($a=0$; solid line for matrix and dash line for inhomogeneity)

It is also well known that mechanical and optoelectronic properties of nanocomposites are significantly influenced by the elastic field of the inhomogeneity. It is therefore important to understand the elastic field of a nanoscale inhomogeneity in a matrix material. To show the surface/interface effect on the elastic field, a matrix-inhomogeneity system made out of InAs/GaAs is considered. The bulk Lamé constants used are: $\lambda_I=50.66$ GPa, $\mu_I=19.0$ GPa for InAs, and $\lambda_M=64.43$ GPa, $\mu_M=32.9$ GPa for GaAs [19].

Consider first the case of a GaAs plane subjected to far-field loading with no eigenstrain in the InAs inhomogeneity. Hoop stress at the point $\theta=0$ on the interface is investigated when $\tau^0=0$. Figure 7 shows the normalized hoop stress of the inhomogeneity and the matrix for various values of a. The results are similar to that of a circular hole (Fig. 2) and the effect of K^S is slightly more prominent in the inhomogeneity than in the matrix. When $a=-1$ and $K^S=10$ N/m, the differences between the present and the classical results are 7% for the inhomogeneity and 4.5% for the matrix. Figure 8 shows the normalized hoop stress for different values of the inhomogeneity radius. Similar to the case of a circular hole, the surface stress effect is significant when the radius of the inhomogeneity is less than 10 nm.

The influence of eigenstrain ε^* in the inhomogeneity is now considered in the absence of far-field loading. In this case, the stress field is radially symmetric and there is no shear stress. It is noted that the Eshelby tensor is uniform in this case but also size dependent due to surface stress effect. Figure 9 shows the normalized hoop stress $\sigma_{\theta\theta}/\mu_I\varepsilon^*$ at the inhomogeneity–matrix interface for $\tau^0=0$. The normalized hoop stress is tensile in the matrix, while it is compressive in the inhomogeneity. In the classical case, hoop stress in the inhomogeneity and matrix have the same absolute value but opposite signs. Size-dependent behavior of hoop stress is clearly evident for an inhomogeneity with a radius smaller than 15 nm.

4 Summary and Conclusion

The elastic state of a nanoscale circular hole/inhomogeneity in an infinite matrix subjected to arbitrary remote loading and a uniform eigenstrain is investigated. The Gurtin–Murdoch surface/interface elasticity model is applied to take into account the surface/interface stress effects at the nanoscale. It is shown that the complex potential function method of Muskhelishvili can be successfully extended to the present class of problems to obtain a

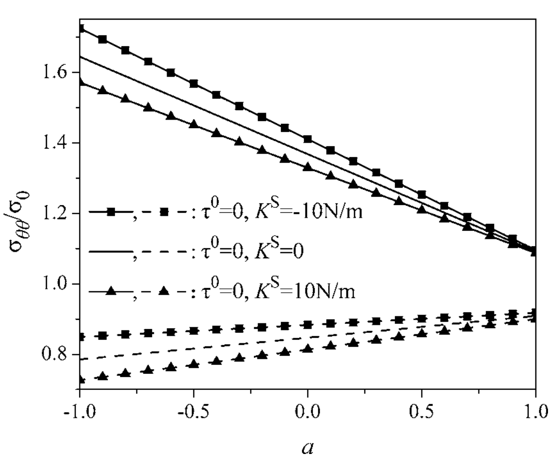


Fig. 7 Variation of normalized interfacial hoop stress at $\theta=0$ with loading ratio a for an inhomogeneity with $R_0=5$ nm (solid line for matrix and dash line for inhomogeneity)

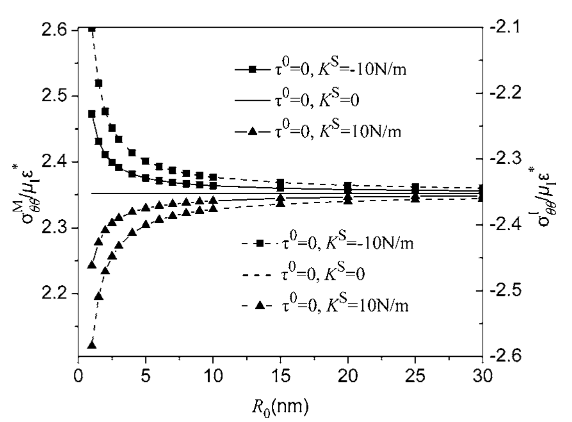


Fig. 9 Variation of normalized interfacial hoop stress due to an eigenstrain ε^* with inhomogeneity radius (solid line and left Y axis for matrix, and dash line and right Y axis for inhomogeneity)

closed-form analytical solution. The new solution presented in this paper reduces to the classical elasticity solution in the absence of surface stress effects. The stress state of a circular hole shows strong dependency on hole radius, surface elastic constants, and residual stress when the hole radius is less than 10 nm. Hoop stress around a hole can be increased or decreased due to surface elasticity effects. Similarly, the elastic field of an inhomogeneity is also affected by its radius, surface elastic constants, and residual stress when the radius of the inhomogeneity is less than 10 nm. The solution derived in this paper provides some fundamental understanding of the elastic field of materials containing nanoscale holes/inhomogeneities. The present results can also be used in the validation of numerical methods such as finite element techniques for nanomechanics problems.

Acknowledgment

The work presented in this paper was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada.

References

[1] Kuzumake, T., Miyazawa, K., Ichinose, H., and Ito, K., 1998, "Processing of Carbon Nanotube Reinforced Aluminium Composite," J. Mater. Res., **13**(9), pp. 2445–2449.

[2] Cui, Y., and Lieber, C. M., 2001, "Functional Nanoscale Electronic Devices Assembled Using Silicon Nanowire Building Blocks," Science, **291**(5505), pp. 851–853.

[3] Wong, E. W., Sheehan, P. E., and Lieber, C. M., 1997, "Nanobeam Mechanics: Elasticity, Strength, and Toughness of Nanorods and Nanotubes," Science, **277**(5334), pp. 1971–1975.

[4] Cuenot, S., Frétigny, C., Demoustier-Champagne, S., and Nysten, B., 2004, "Surface Tension Effect on the Mechanical Properties of Nanomaterials Measured by Atomic Force Microscopy," Phys. Rev. B, **69**(16), p. 165410.

[5] Gibbs, J. W., 1906, *Scientific Papers*, Longmans-Green, London, Vol. I, p. 55.

[6] Gurtin, M. E., and Murdoch, A. I., 1975, "A Continuum Theory of Elastic Material Surfaces," Arch. Ration. Mech. Anal., **57**(4), pp. 291–323.

[7] Gurtin, M. E., Weissmuller, J., and Larché, F., 1998, "A General Theory of Curved Deformable Interfaces in Solids at Equilibrium," Philos. Mag. A, **78**(5), pp. 1093–1109.

[8] Cammarata, R. C., 1997, "Surface and Interface Stress Effects on Interfacial and Nanostructured Materials," Mater. Sci. Eng., A, **237**(2), pp. 180–184.

[9] Miller, R. E., and Shenoy, V. B., 2000, "Size-Dependent Elastic Properties of Nanosized Structural Elements," Nanotechnology, **11**(3), pp. 139–147.

[10] Sharma, P., Ganti, S., and Bhate, N., 2003, "Effect of Surfaces on the Size-Dependent Elastic State of Nanoinhomogeneities," Appl. Phys. Lett., **82**(4), pp. 535–537.

[11] Sharma, P., and Ganti, S., 2004, "Size-Depedent Eshelby's Tensor for Embedded Nano-inclusions Incorporating Surface/Interface Energies," ASME J. Appl. Mech., **71**(5), pp. 663–671.

[12] Duan, H. L., Wang, J., Huang, Z. P., and Karihaloo, B. L., 2005, "Size-Dependent Effective Elastic Constants of Solids Containing Nano-inhomogeneities With Interface Stress," J. Mech. Phys. Solids, **53**(7), pp. 1574–1596.

[13] Muskhelishvili, N. I., 1963, *Some Basic Problems of the Mathematical Theory of Elasticity*, P. Noodhoff, Groningen, The Netherlands.

[14] Ruud, J. A., Witvrouw, A., and Spaepen, F., 1993, "Bulk and Interface Stresses in Silver–Nickel Multilayered Thin Films," J. Appl. Phys., **74**(4), pp. 2517–2523.

[15] Shenoy, V. B., 2005, "Atomistic Calculations of Elastic Properties of Metallic FCC Crystal Surfaces," Phys. Rev. B, **71**(9), p. 094104.

[16] Meyers, M. A., and Chawla, K. K., 1999, *Mechanical Behavior of Materials*, Prentice–Hall, New York.

[17] Yang, F., 2004, "Size-Dependent Effective Modulus of Elastic Composite Materials: Spherical Nanocavities at Dilute Concentrations," J. Appl. Phys., **95**(7), pp. 3516–3520.

[18] Maranganti, R., and Sharma, P., 2005, "A Review of Strain Field Calculations in Embedded Quantum Dots and Wires," *Handbook of Theoretical and Computational Nanotechnology*, M. Reith, and W. Schommers, eds., in press.

[19] Sharma, P., and Ganti, S., 2002, "Interfacial Elasticity Corrections to Size-Dependent Strain-State of Embedded Quantum Dots," Phys. Status Solidi B, **234**(3), pp. R10–R12.

Modeling of Threaded Joints Using Anisotropic Elastic Continua

Daniel J. Segalman
e-mail: djsegal@sandia.gov

Michael J. Starr

Sandia National Laboratories,
P.O. Box 5800,
Albuquerque, NM 87185-0847

Using fine material meshes in structural dynamics analysis is often impractical due to time step considerations. Unfortunately, fine meshes are typically required to capture the inherent physics in jointed connections. This is especially true in threaded connections which feature numerous contact interfaces and stress singularities. A systematic method is presented here for representing the threaded volume by a continuous, homogeneous, linear elastic, anisotropic equivalent material. The parameters of that equivalent material depend on thread geometry and the assumed contact condition between adjacent threads and are derived from detailed finite element simulations of a characteristic thread-pair unit cell. Numerical simulations using the equivalent material closely match the local stiffness through the load path calculated from the finely meshed thread models and also reproduce classical theoretical and experimental results from the literature.
[DOI: 10.1115/1.2424467]

Keywords: threaded connections, joints, anisotropic elasticity, equivalent material, finite element model

1 Introduction

Studies of threaded connections have become ubiquitous in recent technical literature due to the advancing capabilities of finite element tools. In most cases, these analyses are validated through comparison to the broadly accepted theoretical work of Sopwith [1] and the experimental work of Goodier and Hetenyi [2,3].

Finite element models have become accepted tools for the design of screw threads and the development of design codes for mechanical and structural applications [4]. Implicit in the appropriateness of such analyses is sufficient discretization of the model mesh to capture the physics of contact between adjacent threads and to capture adequately the singular behavior at the thread root.

Threaded joints are not only a major component of the mechanical integrity of the structure, but they are also a major path for mechanical energy flow through the system. From the perspective of structural dynamics, the energy flow through a threaded joint is generally a more important consideration than mechanical energy dissipation since it is believed that there is very little energy dissipation in tightened threaded connections. (See Ref. [5] for an example.) Very finely meshed quasistatic finite element analysis of joints can lend insight into joint mechanics, but fine meshes are impractical for direct use in structural dynamics modeling, typically requiring impractically small time steps [6].

Rather than using finely meshed individual threads, a recourse is to replace the threaded region with an equivalent medium which captures the manner in which statically indeterminate equilibrium is achieved in the joint and quantitatively represents the manner of mechanical energy transmission through the joint. Bretl and Cook [7] employed an axisymmetric technique to replace the thread zone with a layer of elements with orthotropic properties. Their numerical results agreed well with theoretical and experimental results but required an a priori assumption of zero normal stress in an assumed direction.

In this paper, a simple, low-order modeling approach that captures the general behavior that would be manifest by a very finely meshed finite element model of the threaded region is explored. The approach is similar to that of Bretl and Cook in that a narrow region including opposing thread pairs is replaced by a homogeneous, continuous material with fictitious material properties. However, rather than assuming a principal direction and normal stress condition to employ an orthotropic model, anisotropic elastic properties are deduced by performing a set of finite element simulations involving homogeneous boundary conditions on a finely meshed characteristic thread-pair unit. This systematic method of arriving at effective material properties leads to easy implementation within a finite element code. The resulting linear model provides an otherwise missing link in the linear structural dynamic analysis of systems connected by threaded assemblies.

The next section explains the motivation and theoretical development of this approach. The remainder of the paper illustrates the technique through a set of two-dimensional numerical bolt pull simulations.

2 Theoretical Construction of Equivalent Homogeneous Material

When performing failure analysis, highly discretized geometries may be inescapable, although engineering judgment may be used to concentrate investigation on local regions of expected high stresses and strains. Structural dynamics analysis, on the other hand, endeavors to capture the manner in which the local geometries and physics yield a global response. Complex interfaces, notably threaded connections, require significant numbers of elements to capture the physics of the mechanical interaction. The core problem of integrating micromechanical analysis of thread interactions with structural dynamics lies in the fundamentally different spatial (and temporal) scales associated with each. This difficulty manifests itself in two ways.

1. The myriad tiny elements needed to capture the geometry and detailed mechanics of each thread-pair define a time scale via the Courant–Friedrichs–Levy condition that is orders of magnitude smaller than those characteristic of the dynamics of the structure as a whole [8]. Regardless of the number of processors available to the analyst, time integration associated with each joint element domain will involve time steps orders of magnitude greater in number than dynamic analysis of similar structures without joints. If this difficulty is addressed through implicit inte-

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received November 15, 2005; final manuscript received May 3, 2006. Review conducted by Sanjay Govindjee.

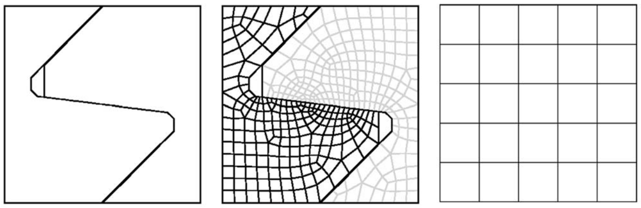


Fig. 1 Unit cell for a representative thread pair

gration, the problem manifests itself through prohibitively bad matrix conditioning. In this sense the problem has more to do with the size of the elements than with the number of elements.

2. Attempting to capture the full unilateral contact mechanics of each joint pair during the structural dynamics analysis requires solving the coupled nonlinear contact problem across each opposing element surface pair. Such problems are particularly difficult, since they converge slowly and in a manner exacerbated by the number of potential contact pairs involved.

In the method presented here, the above difficulties are eliminated by replacing the whole threaded region by a continuous equivalent material. Finite element modeling of this equivalent material could involve elements roughly on the scale of the thread pairs rather than on the scale necessary to micromodel such pairs, thus obviating the first issue raised above. If some of the nonlinearity intrinsic to thread mechanics (such as frictional energy dissipation) are to be captured, an appropriate equivalent nonlinear material would be devised. In that context, the nonlinearity would be embodied in the nonlinear properties of a small number of finite elements representing the threaded region. For example, an anisotropic elastic-plastic constitutive model might be employed for this purpose. In this manner the second of the above difficulties is circumvented. If, however, the primary interest is the elastic behavior of the threaded zone as it affects structural dynamics, an equivalent anisotropic elastic medium is sufficient. This is the approach presented here.

The theoretical construction of equivalent material properties begins with the definition of the thread-pair unit cell. Figure 1 shows a cell of the representative thread pair, the associated fine finite element mesh of the cell, and an equivalent material mesh of the same cell. The micromeshed thread pair is placed in a large mesh containing an array of similar thread pairs (Fig. 2). The boundaries of this large periodic mesh are then subjected to a number, N, of different displacements consistent with homogeneous deformations. Here, St. Venant's principle is assumed to assert that the resulting deformation field in the middle thread pair is similar to that which would result if the thread pair were part of an infinite array of such thread pairs. (Note that St. Venant's principle is applied to the far field, elastic portion of the problem, not to the inelastic near field.) From the static solutions of each of the N displacements, equivalent homogeneous stresses and strains are deduced. The mathematics of these calculations are discussed in the Appendix Secs. 2 and 3, respectively. Finally, the constitutive parameters of an equivalent anisotropic elastic material most consistent with the above ensemble of stress and strain pairs are deduced. The manner in which this optimization is achieved can be found in the Appendix Sec. 4.

3 Applications in Two Dimensions

3.1 Constitutive Formulations. The general approach introduced in this manuscript is illustrated in this section for a problem of two dimensional plane elasticity (plane stress or plane strain). In the most general case of elastic anisotropy, there are 21 material parameters to identify, but in the case of two-dimensional plane elasticity, there are only three significant stress components and three significant strain components, so the number of necessary parameters reduces to six

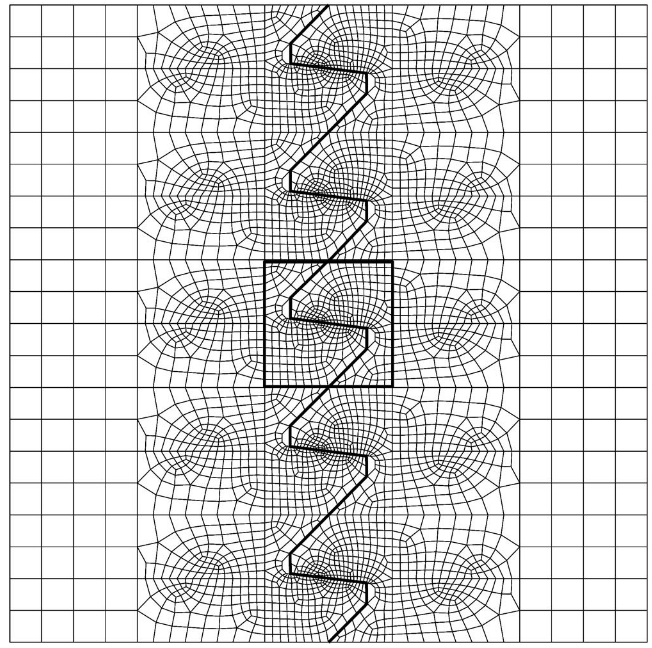


Fig. 2 The finely meshed thread test model for determining material properties in plane strain. The center thread pair is the cell on which the material parameters are calculated.

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{bmatrix} = \begin{bmatrix} C_1 & C_3 & C_5 \\ C_3 & C_2 & C_6 \\ C_5 & C_6 & C_4 \end{bmatrix} \begin{bmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{xy} \end{bmatrix} \tag{1}$$

Though it may be that the equivalent material response is nearly orthotropic, further reducing the number of parameters, the use of general anisotropy permits us to avoid having to identify the principal directions.

The quasistatic finite element code JAS3D [9] was used in the calculations discussed here, with its 3D orthotropic material model extended to accommodate full anisotropy. The choice of plane stress or plane strain elasticity is implemented through the choice of boundary conditions applied to surfaces normal to the x,y plane.

Axisymmetry is accommodated at the cost of just a little more complexity. The thread pitch is assumed small relative to the distance of the thread zone from the axis of symmetry and the circumferential direction is assumed to be a principal material direction. The equilibrium equations involve the three stresses discussed above and the stress $\sigma_{\theta\theta}$ in the circumferential direction. Given the above assumptions, the elastic constitutive model is of the following form:

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \\ \sigma_{\theta\theta} \end{bmatrix} = \begin{bmatrix} C_1 & C_3 & C_5 & \\ C_3 & C_2 & C_6 & \\ C_5 & C_6 & C_4 & \\ & & & C_7 \end{bmatrix} \begin{bmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{xy} \\ \epsilon_{\theta\theta} = u_x/r \end{bmatrix} \tag{2}$$

For simplicity of presentation, the rest of this paper focuses on problems of plane elasticity, but the methods employed apply similarly to problems of axisymmetry.

3.2 Plane Elasticity Threaded Bolt Experiments. The ability of the equivalent material to replace threaded models was explored for the two canonical threaded bolt configurations in Fig. 3. The geometry on the left, case 1, approximates the case of a bolt in an infinite substrate. The geometry on the right, case 2, is similar to the "classical" bolt/nut geometry that is frequently reported

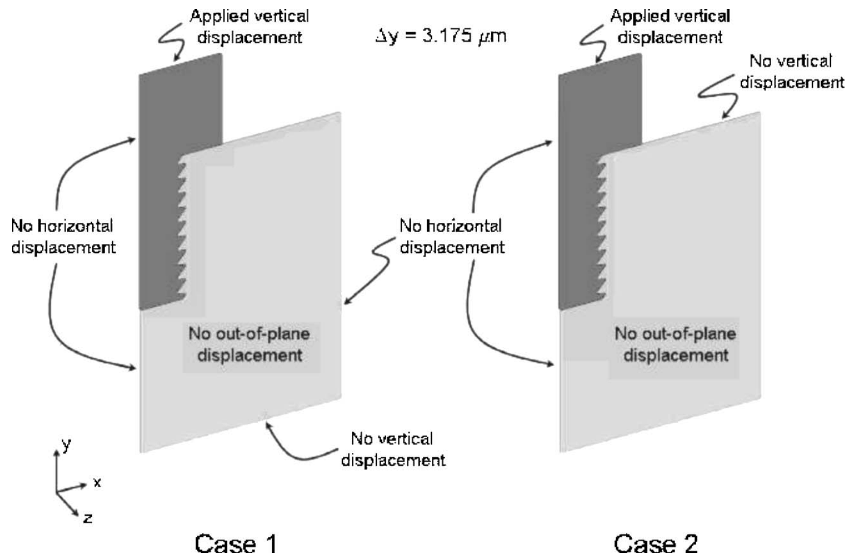


Fig. 3 Simple plane strain bolt pull test designed to exercise the equivalent material model. Case 1 approximates that of a bolt in a large block. Case 2 is consistent with the boundary conditions of a tightened bolt/nut system. The model is cut along its plane of symmetry.

in the literature. The threads are of the buttress variety with the assumed material properties of an aluminum alloy ($E=69$ GPa and $\nu=0.33$) and the boundary conditions are those of plane strain ($\epsilon_{zz}=0$).

This geometry has a large number of threads, making an approximation of periodicity plausible and the thread dimension is small compared to the distance from the axis of symmetry, supporting the plane strain geometric assumptions. In applying this method, a unit thread pair (or unit cell) is defined such that the identified thread region can be seen as an assemblage of stacked unit thread pairs. A unit thread pair is highlighted in its test matrix in Fig. 2. The associated finite element analysis of the unit thread pair requires sufficient geometric resolution to contain all of the pertinent features necessary to capture the important interactions among the teeth.

The thread pair is loaded by applying displacements upon the boundary surfaces of the matrix. Figure 4 shows the four loading cases considered for developing the material constants: two extensional cases and two shear cases. For each case, the finite element code is employed to find the stress and displacement fields over the whole structure from which the equivalent (mean) homogeneous stress and strain fields are deduced in the region of the representative thread pair. Equivalent elastic constants are then calculated in the manner discussed in the previous section, assuming perfect adherence (welded conditions) on each screw thread interface.

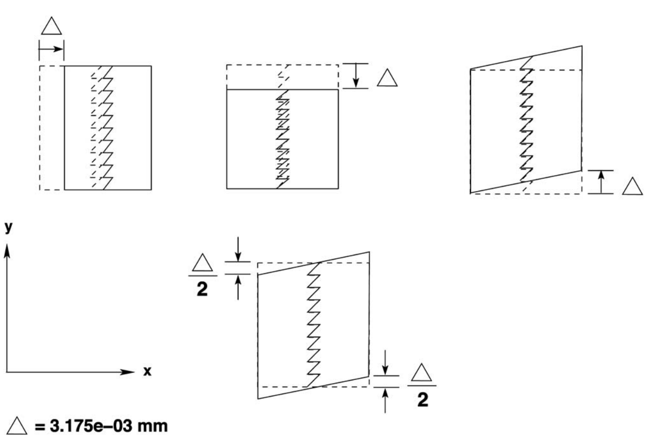


Fig. 4 The four load cases employed to develop the equivalent material parameters

With the equivalent elastic properties now defined, a displacement controlled experiment, where the top surface of the bolt head depicted in Fig. 3 was displaced upward 3.175 μm, was performed for three unique thread representations for both sets of boundary conditions. Two of the thread representations employed identical meshes (Fig. 5) and the actual screw thread geometry, but differed in the application of the contact condition across adjacent threads. The two contact conditions were: frictionless across contacting interfaces and welded across contacting interfaces. These two cases serve as bounds to the stiffness response of the bolt/block system. The number of elements employed in the mesh for these cases was 6309 hexagonal elements of which 1219 were located in the thread region. Most of the 5090 elements outside the thread area are required to accommodate the transition from the very fine mesh used in the threads.

The third thread representation was that of an equivalent, continuous, homogeneous, anisotropic elastic material. Several different meshing schemes, with increasing degrees of coarseness (decreasing degrees of fineness), were employed for comparison to the screw thread geometry results. Figure 6 shows a mesh of intermediate coarseness which has 1754 hexagonal elements with 230 of these elements located in the thread zone.

An important feature of this equivalent anisotropic material mesh is shown in the insets of Fig. 6. The equivalent material approximation is only good where the thread loads are nominally periodic. This assumption breaks down at the extreme boundaries of the threaded region and the voids between the threads must be accommodated explicitly. This is achieved by introducing “cuts” in the mesh at physical locations of noncontact that precede the first thread interface and that follow the last thread interface. (These cuts are indicated in the insets via solid black lines.) Discontinuous displacements are admitted across the cuts, approximating the free surfaces around the voids.

For the boundary conditions of case 1, a comparison of the predictions associated with each of the three described thread representations is shown in the next five figures. The mesh of Fig. 6

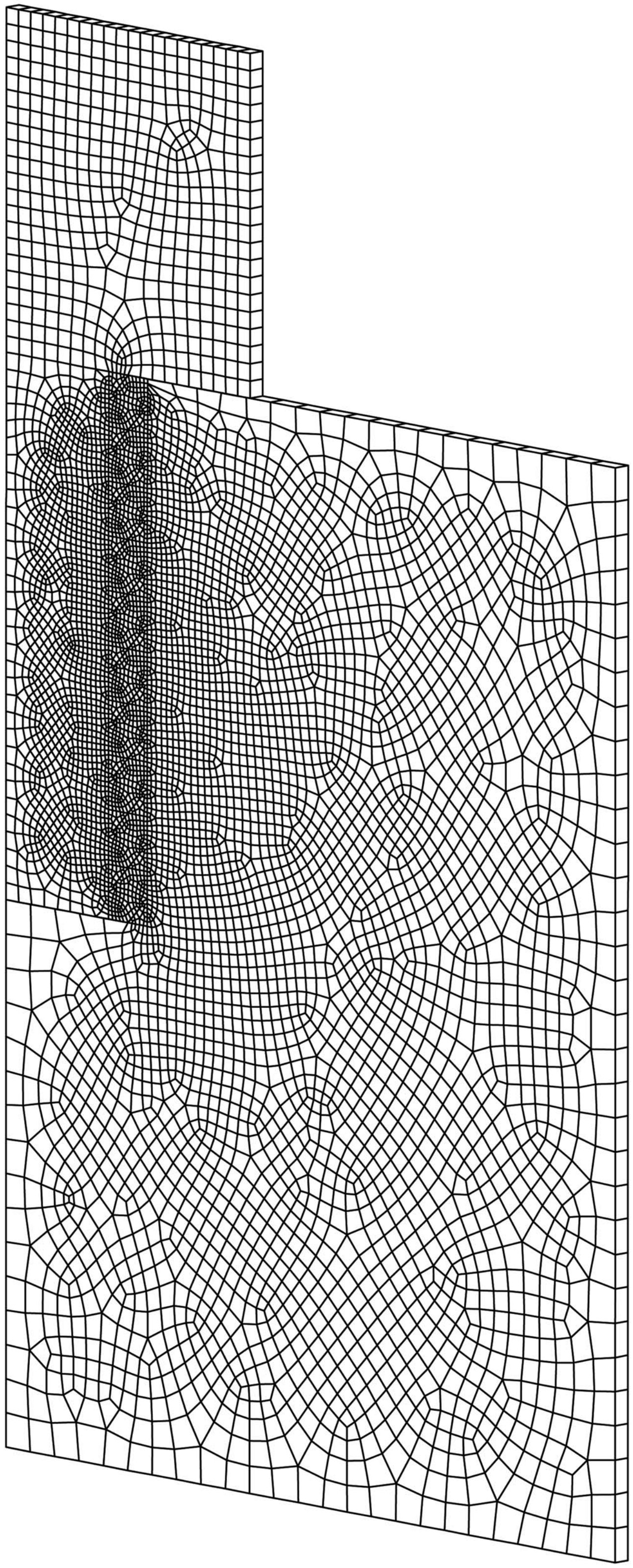


Fig. 5 The finely meshed thread model employs 6309 hex elements, with 1219 of those elements located in the thread region

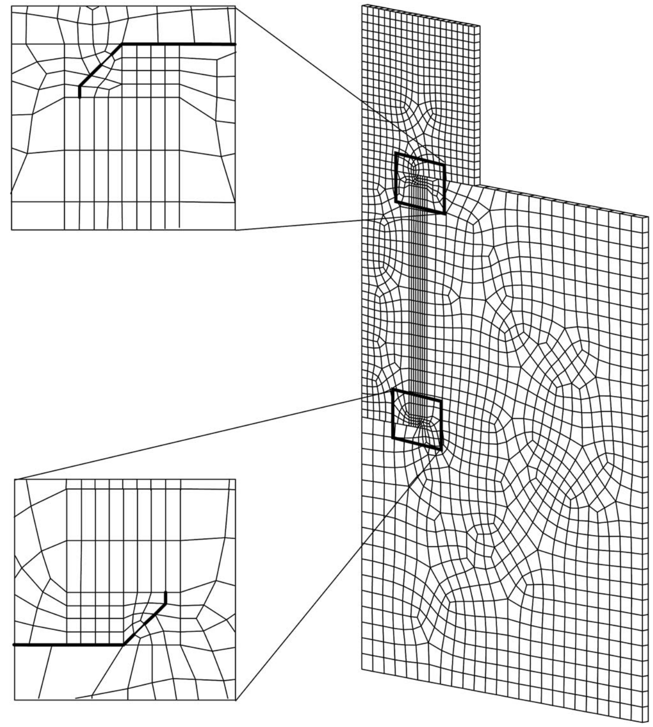


Fig. 6 The coarsely meshed equivalent material model used for comparison in the following plots employs 1754 hex elements, with 230 of those elements located in the thread region. The thick solid lines in the insets are planes of discontinuity within the equivalent material.

was used for the equivalent material model. Figure 7 shows the resultant vertical reaction force for each thread representation as the top of the bolt head is subjected to an imposed vertical displacement. For the small displacements imposed, the response in each instance appears linear with effective stiffness values of (10^6 N/m) 12.17, 12.96, and 12.90 for frictionless screw threads, welded screw threads, and the welded equivalent material model, respectively. The equivalent material model underpredicts the effective stiffness of the welded screw threads by less than half of 1%. The welded equivalent material and the frictionless interface were not expected to agree as well, and indeed these two sets of predictions differ on the order of 6%. The difference in effective stiffness values for the screw thread meshes with different contact conditions appears to be broadly consistent with those numerical results reported by Chabaan and Jutras [4].

These results raise the intriguing possibility of devising an appropriate nonlinear constitutive model to capture the behavior of the frictionless simulations. Development of such a nonlinear model could be the topic of future study.

Figure 8 shows a plot of the shear stress along the midline of the mated screw threads and through the middle of the equivalent material. The origin is positioned at the bottom of the thread stack. Of course for the meshed screw thread cases, the shear stress trace necessarily passes through an interface, whereas the equivalent material is continuous. The oscillatory period of the shear stress along the screw threads corresponds to the geometric period of thread pairs. The shear stress in the equivalent material appears to match the trends of the shear stress at the top and bottom of the thread stack. The shear stress also matches well in the mean sense along the interior thread region. Figure 9 illustrates this matching more directly as the shear stress is integrated along the length of the thread stack. Again, the origin is located at the bottom of the thread stack. The equivalent material model matches well with the welded screw thread case. The oscillations in the screw thread models reflect the periodic nature of the geometric discontinuities along the thread stack.

The next two figures show results along the midline of the bolt shaft, the plane of symmetry in the model. Figure 10 shows the vertical displacements as a function of axial coordinate. Figure 11

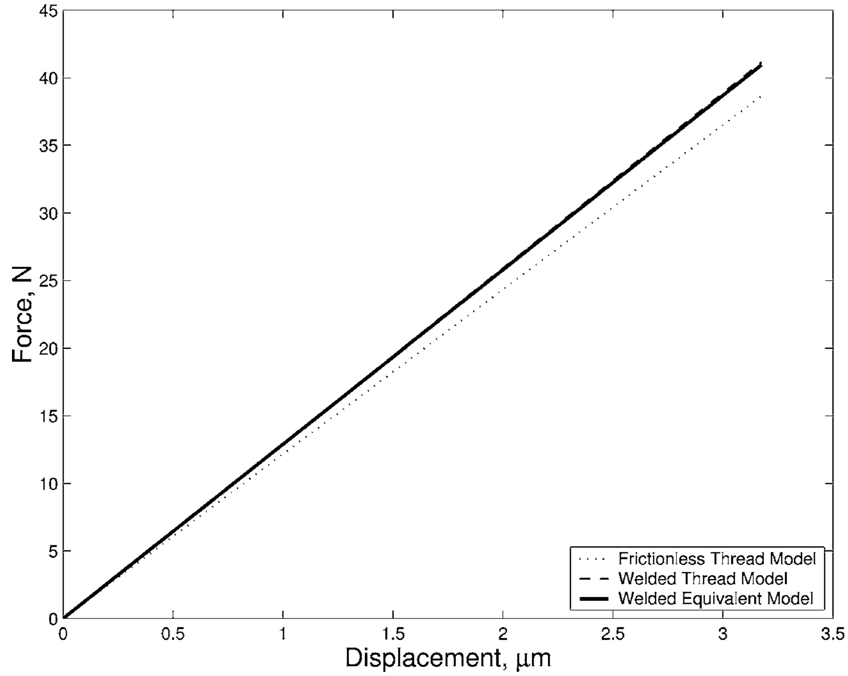


Fig. 7 Force versus displacement plot for three material cases: finely meshed welded threads, finely meshed frictionless threads, and coarsely meshed welded equivalent material. The displacement is measured at the nodes at which the displacement is imposed.

shows the normal stress along the bolt shaft as a function of axial coordinate. In both figures, the origin is located at the bottom of the bolt and the vertical dashed line at 31.5 mm indicates the position of the top of the threads. In both figures there is good matching between the welded screw thread case and the equivalent material model.

Figure 12 shows a comparison of the calculated force-displacement results between the welded thread model and three realizations of the welded equivalent material with different levels of mesh coarseness. The legend on the right hand side of the figure shows the unit cell associated with each line style used in the figure. From top to bottom, the unit cells are: the welded thread model, fine equivalent mesh (4330 total elements, 803 in the thread region), intermediate equivalent mesh, and the coarse equivalent mesh (436 total elements, 63 in the thread region). It is apparent that the force-displacement curves for all four cases are

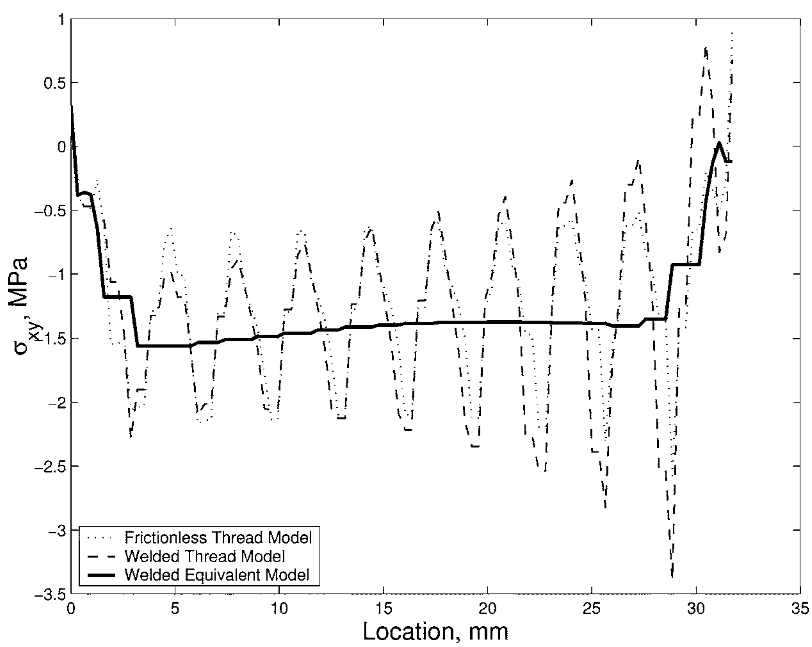


Fig. 8 Shear stress along the midline of the thread region for three material cases: finely meshed welded threads, finely meshed frictionless threads, and coarsely meshed welded equivalent material

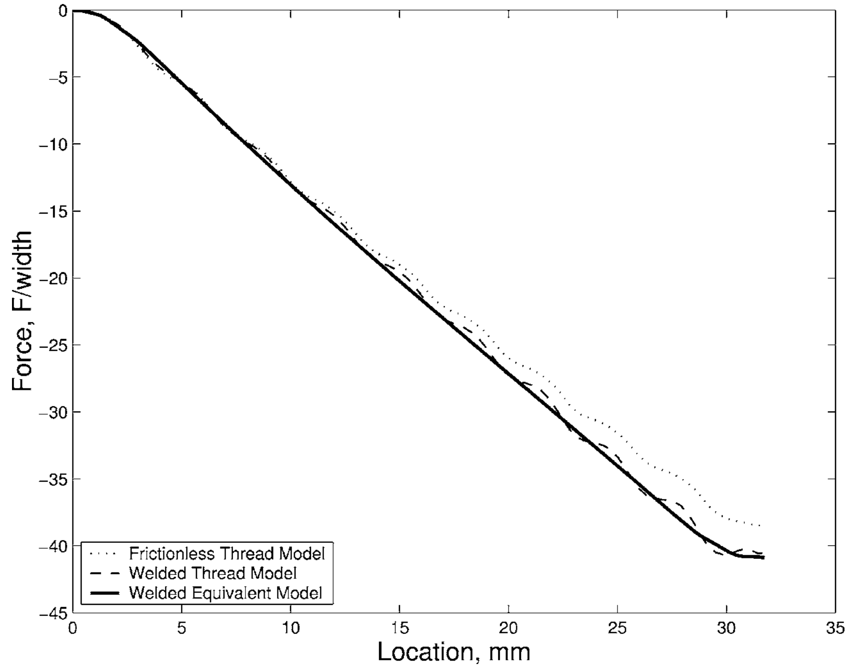


Fig. 9 Integrated shear stress along the midline of the thread region for three material cases: finely meshed welded threads, finely meshed frictionless threads, and coarsely meshed welded equivalent material

quantitatively similar. The effective stiffness (10^6 N/m) for the four meshes are, respectively, 12.96, 12.87, 12.90, and 13.00. As expected, the reported stiffness increase for the more coarsely meshed cases is due to discretization error. However, the stiffness difference is less than 0.5% between the threaded model and the coarsest equivalent model despite a reduction of the number of elements in the threaded region by a factor of nearly 20.

For the boundary conditions of case 2, the shear stress distribution through the center of the thread stack is shown in Fig. 13. This shear stress distribution, as compared to that of case 1, (Fig. 8) suggests substantially different load distributions along the

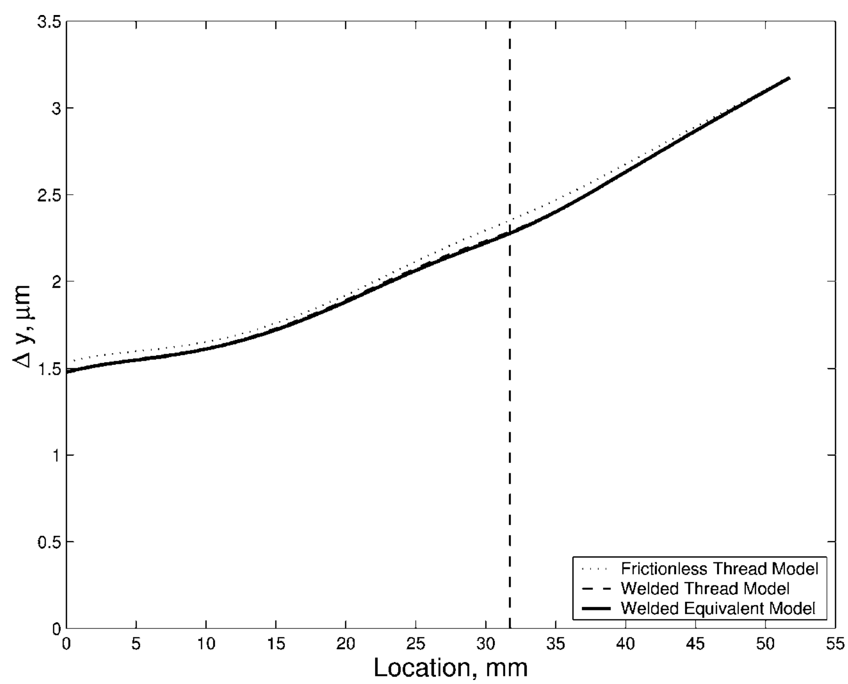


Fig. 10 Vertical displacement along the symmetry plane of the bolt model for three material cases: finely meshed welded threads, finely meshed frictionless threads, and coarsely meshed welded equivalent material. The vertical dashed line corresponds to the location of the bottom of the bolt cap.

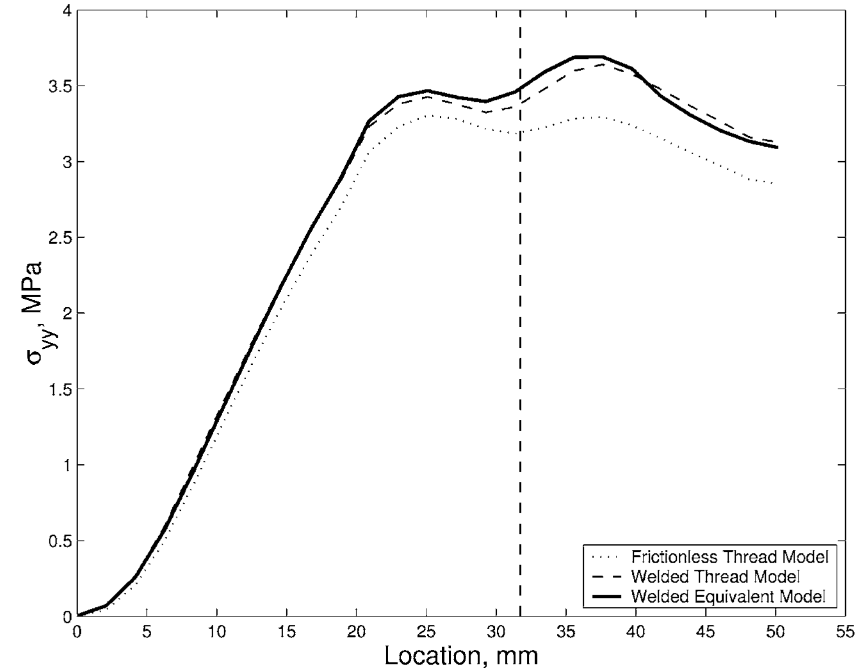


Fig. 11 Normal stress along the symmetry plane of the bolt model for three material cases: finely meshed welded threads, finely meshed frictionless threads, and coarsely meshed welded equivalent material. The vertical dashed line corresponds to the location of the bottom of the bolt cap.

threads. It is often reported as a "rule-of-thumb" that the first several threads of a screw carry the majority of the load. This is indeed true for the boundary conditions illustrated for case 2.

It was Sopwith's [1] examination of threaded joints subject to boundary conditions similar to those of case 2 which lead to the often repeated rule-of-thumb. Ironically, conditions for which the rule is in error are implicitly stated and theoretically supported in that same paper. Indeed, general applicability of the rule-of-thumb is clearly contradicted in Fig. 14. This figure plots the fraction of the total load carried by the threads for the boundary conditions of Fig. 3, where the first thread is that closest to the bolt head. The figure compares the results for the welded thread model and the welded equivalent model for both boundary condition cases. For case 1, the load is distributed essentially linearly along the length of the bolt, while for case 2, the majority of the load is carried by the first several threads. The load distribution agrees, quantitatively, very well with similar results reported by Chaaban and Jutras [4]. Figure 14 also includes an overlay of Sopwith's theo-

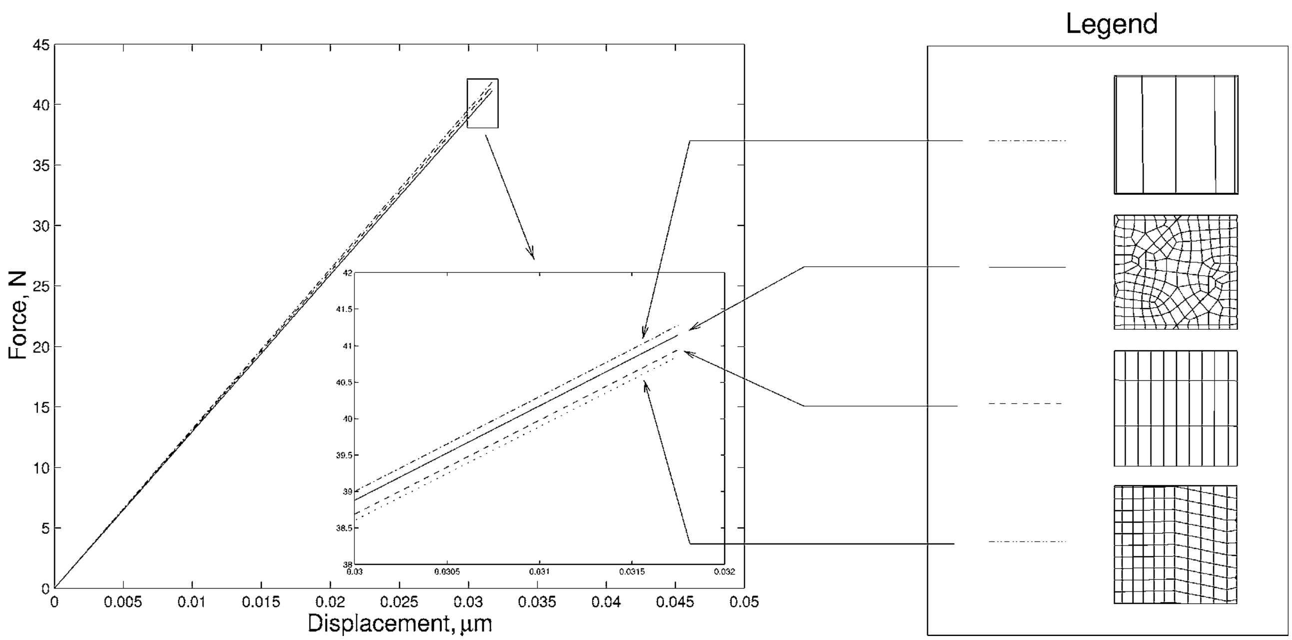


Fig. 12 Force versus displacement plot for four different mesh cases: welded thread model, fine equivalent material, intermediate equivalent material, and coarse equivalent material. The legend associates each case with a corresponding line in the force-displacement plot.

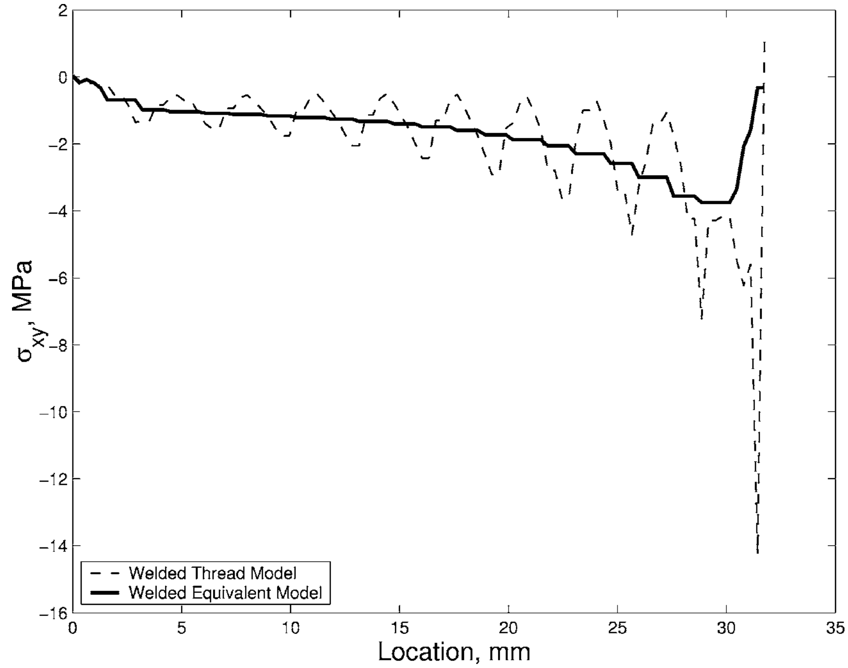


Fig. 13 Shear stress along the midline of the thread region for two material cases: finely meshed welded threads and coarsely meshed welded equivalent material

retical result. There is good agreement between Sopwith's result and case 2 provided that the appropriate triangular thread parameters are used within Sopwith's derivation.

4 Conclusions

A technique has been introduced in this paper for deriving equivalent material models to represent threaded connections. The illustrations have centered on planar elastic response of simple geometries, but the technique can be applied in direct manners to more complex materials and geometries.

It should be noted that it is not necessary to assume that the equivalent material is elastic. In particular, by accommodating the elastic/hardening-plastic nature of the metals that make up the thread cell, appropriate numerical experiments employing incre-

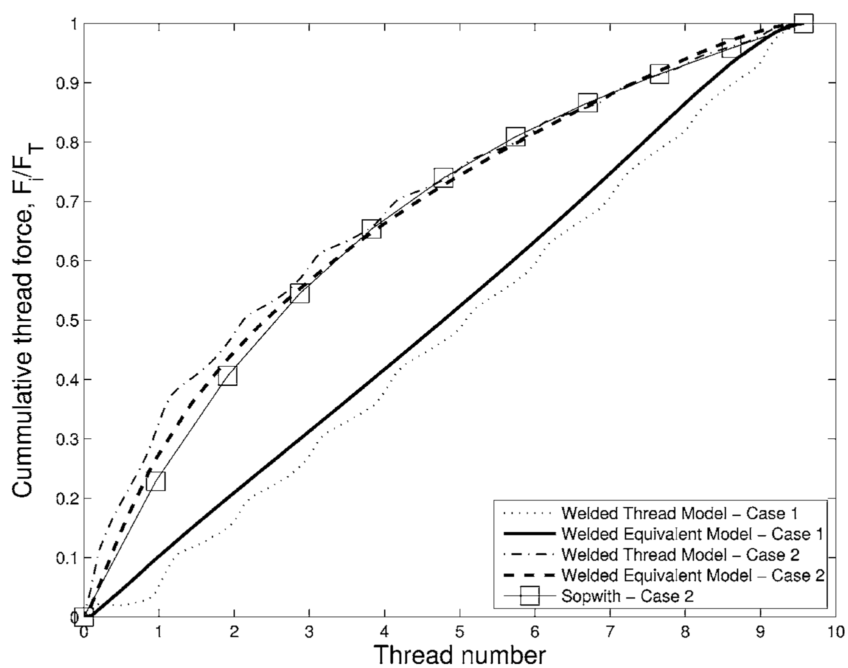


Fig. 14 Fraction of the load carried by the threads along the thread stack for the two boundary condition cases illustrated in Fig. 3. The welded thread model and welded equivalent model are shown for both cases. Sopwith's derivation overlays the results for boundary condition case 2.

mental deformations can be performed to deduce the parameters of an equivalent elastic/hardening-plastic material. This inelastic model suggests a systematic method for predicting ductile failure loads of threaded joints. Specifically, the kinematics deduced from the coarse equivalent model could be mapped onto the nonlinear finely meshed thread-pair unit cell. A quasistatic simulation employing those kinematic boundary conditions could then be compared against a set of local failure criteria.

Three-dimensional analyses similar to the two-dimensional ones performed here would be necessary to accommodate large thread depth to radius problems. Such work is planned for a future study. Because these are relatively small problems, the associated three-dimensional finite element analyses and postprocessing of the results would not be prohibitive. The resulting axisymmetric material would have one more material parameter than was the case in the two-dimensional problems discussed above.

Additionally, if experimental data indicate that significant energy dissipation can take place in a threaded joint—finite element calculations on this issue were inconclusive—dissipative models may be employed. Iwan models could be considered since they have worked well in other contexts (see Refs. [10,11]). The suitability of this approach is still an open question.

Finally, it is important to emphasize that the analysis presented here provides linear approximations for threaded connections. Though it has value in linear structural dynamics, it cannot be employed in problems of load reversals sufficient to overcome preload nor will it be helpful in problems of torsion (finite rotation).

Acknowledgment

The authors thank John F. Holland for his generous aid in meshing and finite element analysis. The authors also thank Anton Sumali of Sandia National Laboratories for his early work on this project which, though pursuing a different direction, helped lay the foundation for the approach of this paper. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract No. DE-AC04-94AL85000.

Appendix

1 Deducing Equivalent Strains, Stresses, and Parameters

The sections in this Appendix provide tools to deduce parameters of an elastic material equivalent to those of a welded unit thread pair. The key notion is to perform a number of elasticity calculations on a mesh containing an array of unit thread pairs (Fig. 2) and then to identify the single set of equivalent elastic parameters that best reproduces all those elasticity results. This effort has three elements.

1. For each detailed elasticity result, deduce an equivalent homogeneous strain for the center unit thread pair.
2. For each detailed elasticity result, deduce an equivalent homogeneous stress for the center unit thread pair.
3. From all of those strain and stress pairs, deduce a set of elastic parameters that best maps the strains onto the stresses.

The next three sections address each of these issues separately.

2 Finding Equivalent Homogeneous Strains

There are subtleties in imposing displacements on the unit cell. These subtleties originate from the fact that in the actual geometry, except at the top and bottom threads, each cell is attached to cells above and below it, as illustrated in Fig. 2. This constraint is satisfied if periodic boundary conditions are imposed explicitly, however, this is difficult to implement in most finite element codes.

Instead, constraints are imposed by embedding the unit cell of interest inside a matrix of similar cells. Displacements are imposed in a manner consistent with homogeneous deformation on the boundaries of that matrix. Labeling locations of nodes, n, on the boundary of the cell matrix as $\mathbf{x}^n$, displacements consistent with the displacement gradient $\mathbf{H}$ [12] are

$$\delta \mathbf{u}^n = \mathbf{H} \cdot \mathbf{x}^n + \mathbf{u}_0 \tag{A1}$$

where $\mathbf{u}_0$ is a rigid body translation. The displacement gradient $\mathbf{H}$ is assumed constant over the cell volume and is found by postmultiplying the above equation by the local outwardly pointing normal vector $\mathbf{n}$ and integrating over the surface of the control volume (the central thread pair)

$$\int_{\partial V} [\delta \mathbf{u}\mathbf{n}] dA = \mathbf{H} \int_{\partial V} [\mathbf{x}\mathbf{n}] dA + \left[\mathbf{u}_0 \left(\int_{\partial V} \mathbf{n}\, dA \right) \right] \tag{A2}$$

where the quantities inside brackets are dyads. If evaluated algebraically, they would be computed as

$$[\mathbf{ab}]_{ij} = a_i b_j \tag{A3}$$

Noting that

$$\int_{\partial V} \mathbf{n} dA = \mathbf{0} \tag{A4}$$

Eq. (A2) becomes

$$\mathbf{H} = \mathbf{U}\mathbf{X}^{-1} \tag{A5}$$

where

$$\mathbf{U} = \int_{\partial V} [\delta \mathbf{u}\mathbf{n}] dA \tag{A6}$$

and

$$\mathbf{X} = \int_{\partial V} [\mathbf{x}\mathbf{n}] dA \tag{A7}$$

The incremental strain corresponding to this deformation is

$$\delta \boldsymbol{\epsilon}_L = \frac{1}{2}(\mathbf{H} + \mathbf{H}^T) \tag{A8}$$

3 Finding Equivalent Homogeneous Stress

The (possibly nonlinear) equilibrium equations are solved to determine the nodal forces and displacements as well as the stresses and strains in the elements. There are two expeditious methods for deducing equivalent homogeneous approximates for the stress in the thread-pair region.

1. Average the element stresses weighted by the element volumes
$$\boldsymbol{\sigma} = \sum_k \boldsymbol{\sigma}^k V_k \Big/ \sum_k V_k \tag{A9}$$
2. Appropriately integrate the tractions applied to the boundary of the thread pair.

Both approaches are mathematically equivalent; the approach used depends on which data are most easily extracted from the finite element microanalysis of the thread pair and its surrounding material. Though the mathematics of the first approach is very simple, the mathematics of the second approach requires some explanation, as follows.

From the forces $\delta \mathbf{f}^n$ on the boundary of the unit cell at the center of the array of cells, corresponding tractions are defined as

$$\hat{\boldsymbol{\tau}}^n = \delta\mathbf{f}^n / dA^n \tag{A10}$$

where dA^n is the surface area corresponding to node n. An expression for $\boldsymbol{\sigma}_L$ is derived in terms of the $\hat{\boldsymbol{\tau}}^n$. If the tractions $\boldsymbol{\tau}$ on the surface are exactly consistent with a uniform stress field $\boldsymbol{\sigma}$, those tractions are expressed in terms of $\boldsymbol{\sigma}$

$$\boldsymbol{\tau}(s) = \boldsymbol{\sigma} \cdot \mathbf{n}(s) \tag{A11}$$

where $\boldsymbol{\tau}(s)$ is the traction at location s on the surface of the cell and $\mathbf{n}(s)$ is the unit outwardly pointing normal there. Taking the outer vector product of Eq. (A11) and integrating over the surface yields

$$\int_{\partial V} [\boldsymbol{\tau}(s)\mathbf{n}(s)]dA = \int_{\partial V} \boldsymbol{\sigma} \cdot [\mathbf{n}(s)\mathbf{n}(s)]dA \tag{A12}$$

Factoring $\boldsymbol{\sigma}$ out of the integral on the right and letting $\mathbf{Q} = \int_{\partial V}[\mathbf{n}(s)\mathbf{n}(s)]\mathrm{d}A$, obtains $\boldsymbol{\sigma}$

$$\boldsymbol{\sigma} = \left\{ \int_{\partial V} [\boldsymbol{\tau}(s)\mathbf{n}(s)]dA \right\} \cdot \mathbf{Q}^{-1} \tag{A13}$$

Equation (A13) provides a natural manner to define the mean stress $\boldsymbol{\sigma}_L$ associated with the tractions obtained via finite elements

$$\boldsymbol{\sigma}_L = \left\{ \sum_n [\hat{\boldsymbol{\tau}}^n \mathbf{n}^n] dA^n \right\} \cdot \mathbf{Q}^{-1} = \left\{ \sum_n [\delta\mathbf{f}^n \mathbf{n}^n] \right\} \cdot \mathbf{Q}^{-1} \tag{A14}$$

Since it is assumed that the stress tensor $\boldsymbol{\sigma}$ in Eq. (A11) is symmetric, Eqs. (A11)–(A14) are also derived easily in the following form:

$$\boldsymbol{\tau}(s) = \mathbf{n}(s) \cdot \boldsymbol{\sigma} \tag{A15}$$

$$\int_{\partial V} [\mathbf{n}(s)\boldsymbol{\tau}(s)]dA = \int_{\partial V} [\mathbf{n}(s)\mathbf{n}(s)]dA \cdot \boldsymbol{\sigma} \tag{A16}$$

$$\boldsymbol{\sigma} = \mathbf{Q}^{-1} \cdot \left\{ \int_{\partial V} [\mathbf{n}(s)\boldsymbol{\tau}(s)]dA \right\} \tag{A17}$$

and

$$\boldsymbol{\sigma}_R = \mathbf{Q}^{-1} \cdot \left\{ \sum_n [\mathbf{n}^n \hat{\boldsymbol{\tau}}^n] dA^n \right\} = \mathbf{Q}^{-1} \cdot \left\{ \sum_n [\mathbf{n}^n \delta\mathbf{f}^n] \right\} \tag{A18}$$

The symmetry of $\boldsymbol{\sigma}$ is guaranteed by averaging the expressions for $\boldsymbol{\sigma}_L$ and $\boldsymbol{\sigma}_R$ in Eqs. (A14) and (A18)

$$\boldsymbol{\sigma} = \frac{1}{2}\left(\left\{ \sum_n [\delta\mathbf{f}^n \mathbf{n}^n] \right\} \cdot \mathbf{Q}^{-1} + \mathbf{Q}^{-1} \cdot \left\{ \sum_n [\mathbf{n}^n \delta\mathbf{f}^n] \right\} \right) \tag{A19}$$

4 Deducing Elastic Properties

Assuming that a number of elastic finite element calculations have been performed as outlined above, a systematic method for processing those results is sought to deduce equivalent elastic parameters for the thread cell.

Each experiment should yield an array S^m, of length M of strain values and another array, T^m, of corresponding stress values where M is the number of components defining the stress or strain state. The superscript m is the index of the numerical experiment. If all the equivalent stress and strain fields deduced from the finite element calculations are consistent with the same elastic response, there is a symmetric $M \times M$ matrix, E, relating the equivalent stresses and equivalent strains

$$T^m = ES^m \tag{A20}$$

Because E is symmetric, it is fully defined by a number, K, of parameters C_k where $K \leq M(M+1)/2$. If further material assumptions, such as isotropy or simple orthotropy, are made on E, the value of K is further reduced. Corresponding to the material parameters C_k are symmetric matrices, B_k, of dimension $M \times M$ which are defined so that

$$E = \sum_{k=1}^{K} C_k B_k \tag{A21}$$

For full anisotropy in plane strain elasticity ($M=3$), there are six material properties and E is expressed

$$E = \begin{bmatrix} C_1 & C_3 & C_5 \\ C_3 & C_2 & C_6 \\ C_5 & C_6 & C_4 \end{bmatrix} \tag{A22}$$

In this case,

$$B_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad B_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$B_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad B_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$B_5 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad B_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{A23}$$

The challenge is to find the material parameters C_k. An objective function $R(C_k)$ associated with Eq. (A20) is defined

$$R = \max_m \{\mathrm{tr}[(T^m - ES^m)^T D (T^m - ES^m)]\} \tag{A24}$$

$$= \max_m \{C_j C_k \,\mathrm{tr}[(S^m)^T B_j^T D B_k S^m) - 2C_j \,\mathrm{tr}[(S^m)^T B_j^T D T^m] + \mathrm{tr}[(T^m)^T D T^m]\} \tag{A25}$$

where D is a diagonal matrix capturing the mapping between shear strain angle and the corresponding component of the strain tensor. In plane strain

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{A26}$$

In Eqs. (A24) and (A25), the stress vectors T^m and strain vectors S^m of each case are each normalized by the largest component of S^m.

The objective function $R(C_k)$ represents the maximum error obtained over all of the m numerical experiments used in the constitutive relationship of Eq. (A20) for a particular set of parameters C_k. The simplex (*fminsearch*) tool in the MATLAB Optimization Toolbox is used to minimize the objective function (which is the maximum residual error) to find the optimal values of C_k. These optimal values of C_k are then used to construct the elasticity matrix E using Eq. (A21).

References

[1] Sopwith, D. G., 1948, "The Distribution of Load in Screw Threads," Proc. Inst. Mech. Eng., **159**, pp. 373–383.
[2] Goodier, J. N., 1940, "The Distribution of Load on the Threads of Screws," ASME J. Appl. Mech., **7**(1), pp. 10–16.
[3] Hetenyi, M., 1943, "A Photoelastic Study of Bolt and Nut Fastenings," ASME J. Appl. Mech., **10**(2), pp. 93–100.
[4] Chaaban, A., and Jutras, M., 1992, "Static Analysis of Buttress Threads Using the Finite Element Method," ASME J. Pressure Vessel Technol., **114**, pp. 209–212.
[5] Hansen, N. R., Bateman, V. I., Bell, R. G., and Brown, F. A., 1996, "Shock Transmissibility of Threaded Joints," *Proceedings of the 67th Shock and Vibration Symposium*, Vol. I.
[6] Zadoks, R. I., and Kokatam, D. P. R., 2001, "Investigation of the Axial Stiff-

ness of a Bolt Using a Three-Dimensional Finite Element Model," J. Sound Vib., **246**(2), pp. 349–373.
[7] Bretl, J. L., and Cook, R. D., 1979, "Modelling the Load Transfer in Threaded Connections by the Finite Element Method," Int. J. Numer. Methods Eng., **14**, pp. 1359–1377.
[8] Belytschko, T., Liu, W. K., and Moran, B., 2000, *Nonlinear Finite Elements for Continua and Structures*, Wiley, New York, p. 315.
[9] Blanford, M. L., Heinstein, M. W., and Key, S. W., 2001, "JAS3D: A Multi-Strategy Iterative Code for Solid Mechanics Analysis," Tech. Rep., Sandia National Laboratories.
[10] Segalman, D. J., 2001, "An Initial Overview of Iwan Modeling for Mechanical Joints," Tech. Rep. SAND2001-0811, Sandia National Laboratories.
[11] Segalman, D. J., 2002, "A Four-Parameter Iwan Model for Lap-Type Joints," Tech. Rep. SAND2002-3828, Sandia National Laboratories.
[12] Truesdell, C., and Noll, W., 1965, *Handbuch der Physik, Band III/3*, Springer, Berlin.

Journal of Applied Mechanics

Technical Briefs

A Complete Perfectly Plastic Solution for the Mode I Crack Problem Under Plane Stress Loading Conditions

David J. Unger
Mem. ASME
Department of Mechanical and Civil Engineering,
University of Evansville,
1800 Lincoln Avenue,
Evansville, IN 47722
e-mail: du2@Evansville.edu

A continuous stress field for the mode I crack problem for a perfectly plastic material under plane stress loading conditions has been obtained recently. Here, a kinematically admissible velocity field is introduced, which is compatible with the continuous stress field obtained earlier. By associating these two fields together, it is shown that they constitute a complete solution for the uncontained plastic flow problem around a finite length internal crack, having a positive rate of plastic work. The yield condition employed is an alternative criterion first proposed by Richard von Mises in order to approximate the plane stress Huber-Mises yield condition, which is elliptical in shape, to one that is composed of two intersecting parabolas in the principal stress plane.
[DOI: 10.1115/1.2338055]

1 Introduction

A statically admissible perfectly plastic solution must satisfy the equations of equilibrium, the specified yield condition, and boundary conditions on traction. For a complete perfectly plastic solution to be constructed, a kinematically admissible velocity field must also be found, which is consistent with the flow rule, and which dissipates energy everywhere in the domain [1].

In [2] a statically admissible solution for a plane stress mode I perfectly plastic crack problem with a continuous stress field was found. Here, a related kinematically admissible velocity field that produces positive plastic work is proposed. Together, they comprise a complete solution of the perfectly plastic plane stress mode I crack problem.

The parabolic Mises yield condition [3] is used in this analysis as the related Huber-Mises (or J_2) yield condition does not allow a statically admissible solution of the plane stress perfectly plastic mode I crack problem without the appearance of stress discontinuities [4]. These stress discontinuities preclude a kinematically admissible velocity field for this particular problem.

The approximation to the J_2 yield condition, first proposed by Richard von Mises [3], and later elaborated upon by others in [5,6], is for plane stress loading conditions

$$\frac{\sigma_1-\sigma_2}{k}=\pm\frac{1}{1+\sqrt{2}}\left[(1+\sqrt{2})^2-\left(\frac{\sigma_1+\sigma_2}{2k}\right)^2\right],\quad \sigma_0=2k \tag{1}$$

where σ_0 is the tensile yield strength. Note that the symbol k in (1) does not represent yield stress in pure shear as it is often used in the literature.

What distinguishes this approximation (1) over the standard J_2 yield criterion is that ellipticity is limited to just two points on the yield surface.

A plot of yield condition (1) in Fig. 1 reveals a yield surface composed geometrically of two intersecting parabolas rather than the conventional ellipse, which is also displayed for comparison. The positive sign in (1) corresponds to the solid parabola shown in the figure, while the negative sign corresponds to the broken parabola. The points where the two separate parabolas intersect are the only points governed by elliptical partial differential equations on this yield surface. The remainder of the yield surface is governed by hyperbolic equations.

2 Stress Field

In Fig. 2, the continuous stress field for the mode I perfectly plastic solution obtained previously in [2] is plotted in terms of a polar coordinate system. The origin of this system is located at the crack tip O of Fig. 3, while the crack surface lies along the line $\theta=0$. This solution remains independent of the polar coordinate r locally, as measured from the left-hand crack tip.

The plastic stress field for the upper left quadrant of the plane in Fig. 3 is composed of two regions of uniform stress separated by a region $OBGC$ having a concentrated fan of mathematical characteristics. The triangular region OAB has a uniaxial state of stress σ_x of magnitude $2k$, whereas sector OCH on the far left has a uniform biaxial state of stress of magnitude $2.41k$. Because sector OCH is an elliptically plastic region, no real characteristics exist for the associated partial differential equation. Consequently, none are shown in the figure for this region.

The formulas for the stresses [2] plotted in Fig. 2 are given below for the three distinct regions represented in Fig. 3 as

$$\left.\begin{aligned}\sigma_r&=2k\cos^2\theta\\ \sigma_\theta&=2k\sin^2\theta\\ \tau_{r\theta}&=-k\sin 2\theta\end{aligned}\right\},\quad 0\leqslant\theta\leqslant\theta_{OB} \tag{2}$$

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received July 11, 2005; final manuscript received May 26, 2006. Review conducted by K. Ravi-Chandar.

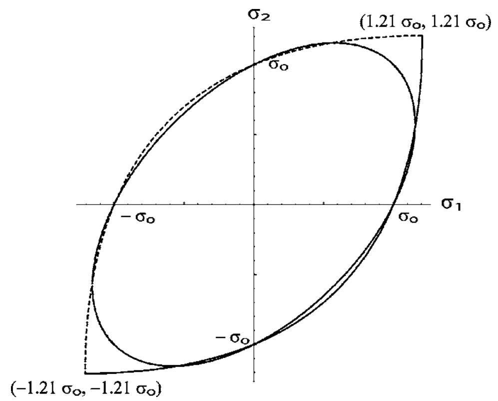


Fig. 1 A comparison of the Huber-Mises yield locus (ellipse) with the parabolic Mises yield locus (parabolas). After [3].

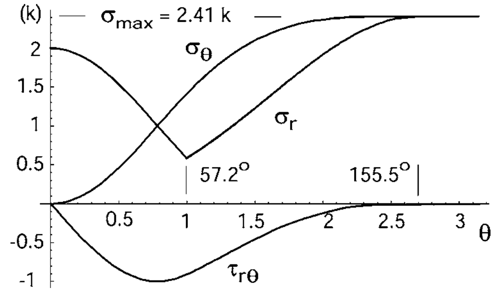


Fig. 2 Stresses of the perfectly plastic plane stress mode I crack problem under the parabolic Mises yield condition [2]

$$\left.\begin{aligned} \sigma_r &= ak\sin\frac{2}{3}(\theta+\beta)\left[1-\frac{1}{2}\cos^2\frac{2}{3}(\theta+\beta)\right] \\ \sigma_\theta &= ak\sin\frac{2}{3}(\theta+\beta)\left[1+\frac{1}{2}\cos^2\frac{2}{3}(\theta+\beta)\right] \\ \tau_{r\theta} &= -\frac{ak}{2}\cos^3\frac{2}{3}(\theta+\beta) \end{aligned}\right\}, \quad \theta_{OB} \leq \theta \leq \theta_{OC} \tag{3}$$

$$\sigma_r = \sigma_\theta = ak, \quad \tau_{r\theta} = 0, \quad \theta_{OC} \leq \theta \leq \pi \tag{4}$$

where the parameters found within (2)–(4) are defined by

$$\theta_{OB} = \sin^{-1} 2^{-1/4} \approx 57.2 \text{ deg}, \quad \theta_{OC} = 3\pi/2 - 2\theta_{OB} \approx 155.5 \text{ deg} \tag{5}$$

$$\beta = 2\theta_{OB} - 3\pi/4 \approx -20.5 \text{ deg}, \quad a = 1+\sqrt{2} \approx 2.41 \tag{6}$$

The equilibrium equations in the plane that are satisfied by the stresses in (2)–(4) are

$$\frac{\partial\sigma_r}{\partial r} + \frac{1}{r}\frac{\partial\tau_{r\theta}}{\partial\theta} + \frac{\sigma_r - \sigma_\theta}{r} = 0 \tag{7}$$

$$\frac{1}{r}\frac{\partial\sigma_\theta}{\partial\theta} + \frac{\partial\tau_{r\theta}}{\partial r} + 2\frac{\tau_{r\theta}}{r} = 0 \tag{8}$$

The following elementary transformations can be used to relate the stresses in polar form to the principal stresses

$$\{\sigma_1, \sigma_2\} = (\sigma_r + \sigma_\theta)/2 \pm \sqrt{(\sigma_r - \sigma_\theta)^2 + 4\tau_{r\theta}^2}/2 \tag{9}$$

It is easily verified that stress fields (2)–(4) satisfy both equilibrium (7) and (8) and the yield condition (1) using any commercially available symbolic mathematics software and relationships (9).

After some simplification, the results obtained by substituting (3) into (9) are

$$\{\sigma_1, \sigma_2\} = \frac{ak}{2}\left[2\sin\frac{2}{3}(\theta+\beta) \pm \cos^2\frac{2}{3}(\theta+\beta)\right], \quad \theta_{OB} \leq \theta \leq \theta_{OC} \tag{10}$$

These principal stresses (10) are related to those given implicity in [5] by substituting equations (82.11) into (72.6).

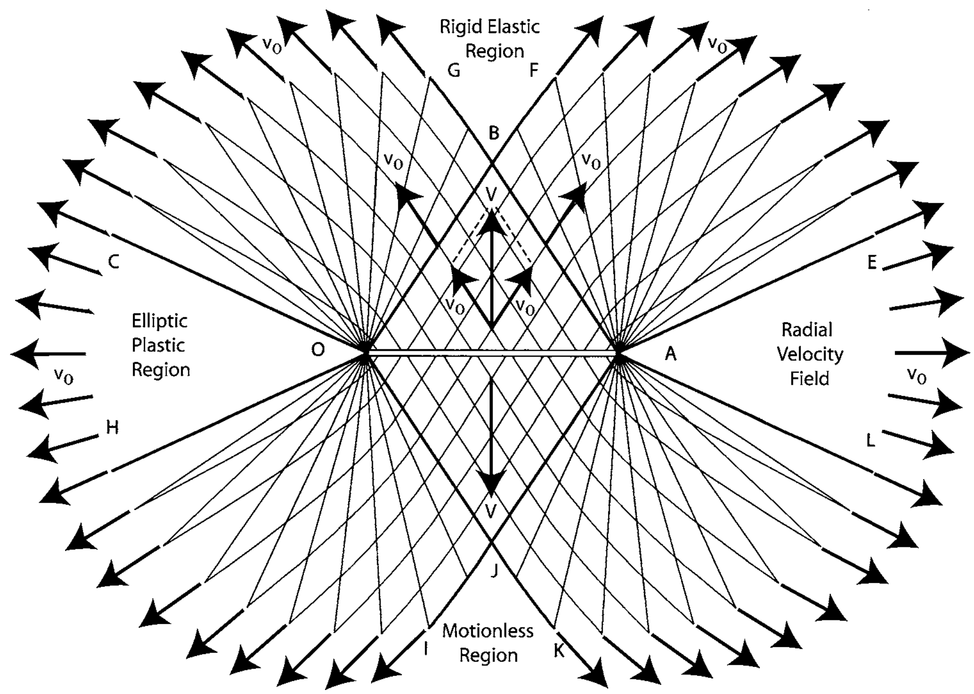


Fig. 3 Perfectly plastic velocity field for the plane stress mode I crack problem under the parabolic Mises yield condition

3 Velocity Field

A solution of a velocity field compatible with the continuous stress field (2)–(4) is presented in this section for a rigid elastic/perfectly plastic mode I crack problem under the parabolic Mises yield condition (1) assuming plane stress loading conditions. The solution presented in [2] was for a semi-infinite crack. Here, the solution is extended to the case of the finite length crack shown in Fig. 3. Notice a second crack tip is positioned at point A of Fig. 3, together with a second fan of characteristics centered at A, and a second biaxial stress region AEL to the far right. The stresses in regions $ABFE$ and AEL to the right follow immediately from symmetry considerations with corresponding regions on the left. They are given explicitly in [7]. For the finite length crack, the rigid elastic/perfectly plastic boundary follows the two curved characteristics BF and BG for the upper half-plane. Similar regions are defined for the lower half-plane. Lacking a stress discontinuity, a simple kinematically admissible velocity field can be determined for this particular stress field. The mathematical characteristics for the velocity field coincide with those of the stress field as they do for the conventional J_2 yield condition [5,6].

The associated flow rule for (1) follows immediately from the normality condition [1] that must be imposed on strain rates with respect to the yield surface. These strain rates are

$$\dot{\varepsilon}_r = \frac{\dot{\lambda}}{k}\left[\frac{\sigma_r - \sigma_\theta}{\sigma_1 - \sigma_2} + \frac{1}{2ak}(\sigma_1 + \sigma_2)\right] \quad (11)$$

$$\dot{\varepsilon}_\theta = \frac{\dot{\lambda}}{k}\left[\frac{\sigma_\theta - \sigma_r}{\sigma_1 - \sigma_2} + \frac{1}{2ak}(\sigma_1 + \sigma_2)\right] \quad (12)$$

$$\dot{\gamma}_{r\theta} = \frac{4\dot{\lambda}}{k}\frac{\tau_{r\theta}}{\sigma_1 - \sigma_2} \quad (13)$$

where $\dot{\varepsilon}_r$, $\dot{\varepsilon}_\theta$ are normal strain rates, $\dot{\gamma}_{r\theta}$ is a shear strain rate, and $\dot{\lambda}$ is a non-negative function of (r, θ).

By substituting (2)–(4) and (10) into (11)–(13) and then eliminating $\dot{\lambda}$ from (12) and (13), one obtains the following system of equations for the fan region $OBGC$

$$\frac{\dot{\varepsilon}_\theta}{(2/k)\sin\frac{2}{3}(\theta+\beta)} = -\frac{\dot{\gamma}_{r\theta}}{(2/k)\cos\frac{2}{3}(\theta+\beta)} \quad (14)$$

$$\dot{\varepsilon}_r = 0, \quad \theta_{OB} \leqslant \theta \leqslant \theta_{OC} \quad (15)$$

Equation (15) reflects the inextensibility of the material along radial characteristics in the upper left-hand fan of Fig. 3.

Now, in general, strain rates can be related to the radial velocity v_r and transverse velocity v_θ by the following equations

$$\dot{\varepsilon}_r = \frac{\partial v_r}{\partial r}, \quad \dot{\varepsilon}_\theta = \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{v_r}{r}, \quad \dot{\gamma}_{r\theta} = \frac{1}{r}\frac{\partial v_r}{\partial \theta} + \frac{\partial v_\theta}{\partial r} - \frac{v_\theta}{r} \quad (16)$$

By substituting (16) into (14), the relationship becomes

$$\frac{\partial v_\theta}{\partial \theta} + v_r = -\tan\frac{2}{3}(\theta+\beta)\left(\frac{\partial v_r}{\partial \theta} + r\frac{\partial v_\theta}{\partial r} - v_\theta\right) \quad (17)$$

Let us now assume a velocity field that is independent of r so that (15) is trivially satisfied. In addition, the following term vanishes in (17)

$$\partial v_\theta/\partial r = 0 \quad (18)$$

which reduces the partial differential equation to the ordinary differential equation below

$$v_r(\theta) + \tan\frac{2}{3}(\theta+\beta)\frac{dv_r(\theta)}{d\theta} = \tan\frac{2}{3}(\theta+\beta)v_\theta(\theta) - \frac{dv_\theta(\theta)}{d\theta} \quad (19)$$

A more general relationship in a nonorthogonal coordinate system based on the characteristic directions in the fan is given in [5,6].

An attempt will now be made to find a velocity field whose general behavior roughly resembles that of a mode I linear elastic displacement field. To this end, one notes that a simple solution of (19) exists in the form

$$v_r(\theta) = v_0 \sin\frac{2}{3}(\theta+\beta), \quad v_\theta(\theta) = v_0\cos\frac{2}{3}(\theta+\beta),$$

$$\theta_{OB} \leqslant \theta \leqslant \theta_{OC} \quad (20)$$

where v_0 is a constant. The above velocity can be shown to be tangent to the family of curved characteristics in the fan $OBGC$, which are expressible in polar coordinates by

$$r = r_c/\cos^{3/2}\frac{2}{3}(\theta+\beta), \quad \theta_{OB} \leqslant \theta \leqslant \theta_{OC} \quad (21)$$

where r_c is constant along a particular characteristic. A general discussion of the fan characteristics for (1) is given in [5,6]. There a relationship similar to (21) was derived although the orientation of the field and the constant β were different. The in-plane flow pattern associated with solution (20) along characteristics defined by (21) is illustrated in the fan region $OBGC$ of Fig. 3.

Recalling that in the linear elastic mode I crack solution the displacements along the crack are vertical and symmetrical with respect to the y axis, let us propose that region OAB of the plastic stress field moves upward as a rigid plastic block with uniform speed V, i.e.,

$$v_r = V\sin\theta, \quad v_\theta = V\cos\theta, \quad 0 \leqslant \theta \leqslant \theta_{OB} \quad (22)$$

A corresponding rigid block moves downward with a constant speed V on the opposite side of the crack. With this assumption, a velocity discontinuity must exist across line OB, which separates the fan from the uniform stress region. To have a continuous velocity field in the direction of the curved characteristics across OB, as indicated in Fig. 3, one must impose the following relationship between parameters V and v_0

$$v_0 = \frac{V}{2\sin\theta_{OB}} = \frac{V}{2^{3/4}} \approx 0.595V \quad (23)$$

A velocity discontinuity of magnitude v_0 therefore exists in the radial direction across OB.

The rigid elastic region above the fan is assumed motionless. This assumption creates another velocity discontinuity along the boundary BG as the adjacent material is moving tangentially with a speed v_0.

On the other side of the fan $OBGC$, a radial velocity field of constant speed is proposed for the elliptic plastic region, which lacks the two families of mathematical characteristics found in the hyperbolic plastic regions. This region behaves mathematically similar to plastic material associated with the square corners of the Tresca yield condition [1]. A radial velocity field is chosen for this region, as it accounts for the necessary symmetry condition of a horizontal velocity along the crack axis, while allowing for continuous velocity across line OC of Fig. 3. Mathematically, this velocity field is expressed as

$$v_r = v_0, \quad v_\theta = 0, \quad \theta_{OC} \leqslant \theta \leqslant \pi \quad (24)$$

The velocity fields for the other three quadrants of the plane are defined in a similar fashion such that obvious symmetry conditions are met.

4 Energy Dissipation

Given a statically admissible, perfectly plastic stress field, the choice of a compatible velocity field is often nonunique. A particular kinematically admissible velocity field must also be checked for energy dissipation as a physical requirement [1].

In regard to fracture mechanics, one can observe in [8] that both the Prandtl and Hill velocity analogies of the lubricated punch problem for a perfectly plastic, plane strain, mode I crack problem are kinematically admissible and satisfy the upper bound theorem of classical plasticity [1].

In this section the rates of plastic work for the stress and velocity fields presented in the previous sections are proven positive. There are two categories of energy dissipation that need to be individually considered [1], i.e.,

$$\dot{D}_1 = \sigma_r \dot{\varepsilon}_r + \sigma_\theta \dot{\varepsilon}_\theta + \tau_{r\theta}\dot{\gamma}_{r\theta}$$
$$\dot{D}_2 = -\tau \Delta v \tag{25}$$

The first $\dot{D}_1$ is the rate of energy dissipation per unit volume due to plastic deformation itself. The second $\dot{D}_2$ is the rate of energy dissipation per unit area due to traction τ between adjacent surfaces, which undergo relative motion Δv.

In the three distinct sectors comprising the solution shown in Fig. 3, one finds for the first energy contribution

$$\dot{D}_1 = \begin{cases} 0(\mathit{rigid\ plastic\ region}), & 0 \leqslant \theta \leqslant \theta_{OB} \\ \dfrac{ak}{12}\dfrac{v_0}{r}\left(3 - \cos\dfrac{4}{3}(\theta+\beta)\right), & \theta_{OB} \leqslant \theta < \theta_{OC} \\ ak\dfrac{v_0}{r}, & \theta_{OC} \leqslant \theta \leqslant \pi \end{cases} \tag{26}$$

The second components of energy dissipation are along the characteristics OB and BG, respectively, where velocity discontinuities occur

$$\dot{D}_2 = \begin{cases} OB: \quad 0.911kv_0, & \theta = \theta_{OB} \\ BG: \quad \dfrac{akv_0}{2}\cos^3\dfrac{2}{3}(\theta+\beta), & \theta_{OB} \leqslant \theta \leqslant \theta_{OC} \end{cases} \tag{27}$$

As all possible energy contributions due to plastic deformation and plastic flow dissipate energy (are non-negative), one may conclude that the two fields are physically acceptable from an energy standpoint.

5 Discussion

Alternative boundary conditions to those proposed in Sec. 3 have also been investigated. However, the nonorthogonality of plane stress mathematical characteristics place severe restrictions on the types of flow that can be realized.

One might consider allowing the rigid elastic regions BFG and IJK of Fig. 3 to move vertically upward and vertically downward respectively with finite speed U. Symmetry arguments combined with the rigidity of the elastic material rule out other possibilities regarding motion of these rigid elastic regions. However, a finite velocity solution with continuous normal components across BG cannot be constructed that does not violate material integrity across the elastic-plastic boundary.

To resolve this issue of gaps and interpenetration of material along BG, one might consider next that localized necking occurs along that characteristic curve. Unlike plane strain problems, where only discontinuous tangential components of velocity are allowed, plane stress problems do permit discontinuous normal components of velocity across elastic-plastic boundaries as well, provided the concept of localized necking is introduced. However, localized necking itself places certain restrictions on the relative velocity of the surfaces along the characteristic interface. Physically, the relative velocity must be perpendicular to the second family of characteristics, which meet the primary characteristic along which necking is assumed to occur, see [1], pp. 272–274, for the Huber-Mises yield condition. In the case of necking along the characteristic BG of Fig. 3, this condition requires that the relative velocity (or velocity discontinuity) of the surfaces be in the transverse θ direction. Consequently, the velocity in the radial direction across the elastic-plastic boundary must remain continuous. This restriction on velocity across BG indicates that the only possible solution is a rigid body motion or, in other words, a situation where no necking occurs. Instead plastic region $COBG$ moves upward at the same speed U as the rigid elastic region BFG.

Nevertheless, a rigid body motion U for region $COBG$ (and consequentially rigid plastic region OAB) might be superposed over the solution already proposed in Sec. 3, as the partial differential equations governing velocity are linear, to meet this additional boundary condition. This assumption would necessitate additional deformation in plastic region OCH and the corresponding section on the right AEL. For example, a velocity v_y might be proposed in region OCH, which varies linearly with θ while assuming a value U along OC and zero along the crack axis to maintain symmetry. The choice of possible velocity fields for OCH is naturally unlimited as this region has indeterminate principal directions for stress in the plane and consequently a much greater degree of freedom than those plastic regions possessing two families of mathematical characteristics. A solution of this type would produce a two-parameter model for velocity $\{U,V\}$.

The use of the velocity field given here was applied in [7]. There a minor modification of the velocity field of Sec. 3 in regions OCH and AEL of Fig. 3 was made to define continuous and smooth perfectly plastic caustics across lines OC, OH, AE, and AL under the parabolic Mises yield condition. This result was compared to caustics obtained for the Huber-Mises and the Tresca yield conditions.

References

[1] Kachanov, L. M., 2004, *Fundamentals of the Theory of Plasticity*, Dover, Mineola, NY.
[2] Unger, D. J., 2005, "A Plane Stress Perfectly Plastic Mode I Crack Solution With Continuous Stress Field," ASME J. Appl. Mech., **72**, pp. 62–67.
[3] von Mises, R., 1949, "Three Remarks on the Theory of the Ideal Plastic Body," *Reisner Anniversary Volume, Contributions to Applied Mechanics*, Edwards, Ann Arbor, MI, pp. 415–429.
[4] Hutchinson, J. W., 1968, "Plane Stress and Strain Fields at the Crack Tip," J. Mech. Phys. Solids, **16**, pp. 337–347.
[5] Freudenthal, A. M., and Geiringer, H., 1958, "The Mathematical Theories of the Inelastic Continuum," Elasticity and Plasticity, *Handbuch der Physik*, Vol. VI, S. Flügge, ed., Springer-Verlag, Berlin, pp. 229–433.
[6] Geiringer, H., 1973, "Ideal Plasticity," Mechanics of Solids III, *Handuch der Physik*, Vol. VIa/3, C. Truesdell, ed., Springer-Verlag, Berlin, pp. 403–533.
[7] Unger, D. J., 2005, "Perfectly Plastic Caustics for the Opening Mode of Fracture," Theor. Appl. Fract. Mech., **44**, pp. 82–94.
[8] Unger, D. J., 2001, *Analytical Fracture Mechanics*, Dover, Mineola, NY, pp. 75–90.

Improved Procedures for Static and Dynamic Analyses of Wrinkled Membranes

Amit Shaw

D Roy[1]
e-mail: royd@civil.iisc.ernet.in

Structures Lab,
Department of Civil Engineering,
Indian Institute of Science,
Bangalore 560012,
India

An analysis of large deformations of flexible membrane structures within the tension field theory is considered. A modification of the finite element procedure by Roddeman et al. (Roddeman, D. G., Drukker, J., Oomens, C. W. J., Janssen, J. D., 1987, ASME J. Appl. Mech. ***54****, pp. 884–892) is proposed to study the wrinkling behavior of a membrane element. The state of stress in the element is determined through a modified deformation gradient corresponding to a fictive nonwrinkled surface. The new model uses a continuously modified deformation gradient to capture the location orientation of wrinkles more precisely. It is argued that the fictive nonwrinkled surface may be looked upon as an everywhere-taut surface in the limit as the minor (tensile) principal stresses over the wrinkled portions go to zero. Accordingly, the modified deformation gradient is thought of as the limit of a sequence of everywhere-differentiable tensors. Under dynamic excitations, the governing equations are weakly projected to arrive at a system of nonlinear ordinary differential equations that is solved using different integration schemes. It is concluded that implicit integrators work much better than explicit ones in the present context.* [DOI: 10.1115/1.2338057]

1 Introduction

Space missions often use membrane structures of different shapes and sizes. These deployable structures consist of thin polymer films. Because of their negligibly small flexural stiffness, membranes cannot sustain compressive stresses beyond a certain level. Accordingly, regions wherein compressive stresses tend to breach this minimal level buckle locally and wrinkles form. The presence of wrinkles in a membrane structure may significantly influence the static and dynamic behavior of a space system containing membranes. It is therefore important to develop numerically accurate and computationally efficient methods for the prediction of wrinkled zones in a membrane. A few approaches, especially within the tension field theory [1], have been developed over the years to analyze wrinkled membranes. Mansfield [2] reformulated Wagner's theory by replacing the strain energy by a suitably relaxed energy density and with the work of Pipkin [3] the theory was incorporated to handle membrane wrinkling. It has been postulated that compressive stresses are eliminated when wrinkles form and the major principal stress is non-negative (tensile) everywhere in the membrane [4,5]. Based on the approach by Wu [6], Roddeman et al. [7] have accounted for wrinkling by replacing a given deformation gradient tensor, which would result in negative Cauchy stresses in the membrane, by a modified deformation gradient tensor using a parameter variable (β) measuring physical wrinkliness. Finite element analyses of wrinkling are given in Ref. [8]. The theory have also been developed for anisotropic [9] and orthotropic [10] materials.

The main aim of this study is to modify and improve upon the wrinkled element model by Roddeman et al. [7] and thus treat the wrinkliness parameter, β, to be a spatially varying smooth (continuous and differentiable almost everywhere) function, which may be discretized and interpolated. Use of this parameter function modifies the deformation gradient so the actually deformed and wrinkled surface is projected onto a fictive nonwrinkled surface. It is argued that the modified deformation gradient may be thought of as the limit of a sequence of differentiable tensor functions and this allows one to treat the parameter function $\beta(X)$ as a differentiable function. The proposed strategy, implemented via the finite element method, is demonstrated with a few numerical examples. A convergence study is also undertaken to validate the functional representation of $\beta(X)$. The model is then applied for dynamic analysis of wrinkled membranes and to temporally integrate the discretized equations of motion using a few implicit and explicit numerical schemes.

2 Basic Equations for the Analysis of Wrinkles

Wrinkles on a thin membrane are the results of local buckling due to compressive stress fields. However, within the tension field theory, deformations due to compressive stresses cannot be accounted for. Thus, once wrinkles are present, the basic governing equations must be modified [7] and it is accordingly necessary to have the following assumptions: (i) plane stress theory is applicable; (ii) flexure does not introduce stresses in the membrane; and (iii) the membrane is not able to support any compressive (negative) stresses (i.e., if a negative stress tends to appear, the membrane will wrinkle at once). For the so-called "Cauchy-elastic" materials, the constitutive equation may be written as

$$\sigma = (1/J)FH(E)F^T \tag{1}$$

with F being the deformation gradient; $J=\det(F)$ the Jacobian of F; E is the Green-Lagrange strain; H a tensor function of E (describing the material constitutive relation) and σ the Cauchy stress tensor. The superscript "T" denotes transpositions of second order tensors. F corresponds to the "real" deformation gradient and this could result in negative Cauchy stresses thereby resulting in the corresponding portions of the membrane getting wrinkled. Thus the deformation gradient of the fictive, nonwrinkled membrane segment, corresponding to non-negative principal stresses, is given by [7]

$$F' = [I + \beta(n_1 \otimes n_1)] \cdot F \tag{2}$$

where I is the unit second order tensor and the scalar parameter $\beta \geqslant 0$ is a measure of the stretching required along the direction of the negative stress. Using F' in the constitutive equation (1), the stress state may be determined by

$$\sigma(F') = 1/J' F' \cdot H(E') \cdot F'^T \tag{3}$$

where,

$$E' = 1/2(F'^T \cdot F' - I); \quad J' = \det(F') \tag{4}$$

2.1 A Continuous Representation of β. Let $X=(x,y,z)$ be an inertially fixed Cartesian coordinate system and let the membrane surface in initial (reference) configuration be denoted as M_0 and let it be sufficiently smooth so that it may be characterized by a collection of tangent planes (bundle) $TM_0(P_0)\ \forall P_0$ where P_0 is a material point on M_0 with coordinate X_0. Let the deformed membrane surface, denoted by M_D and characterized through tangent planes $TM_D(P_{D0})$ with P_{D0} being the image of P_0 on M_D, be partially wrinkled and partially taut. If P_{D0} is in a wrinkled zone, an infinitesimally small open neighborhood around P_{D0} must un-

[1]Author to whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received May 18, 2005; final manuscript received April 10, 2006. Review conducted by E. M. Arruda.

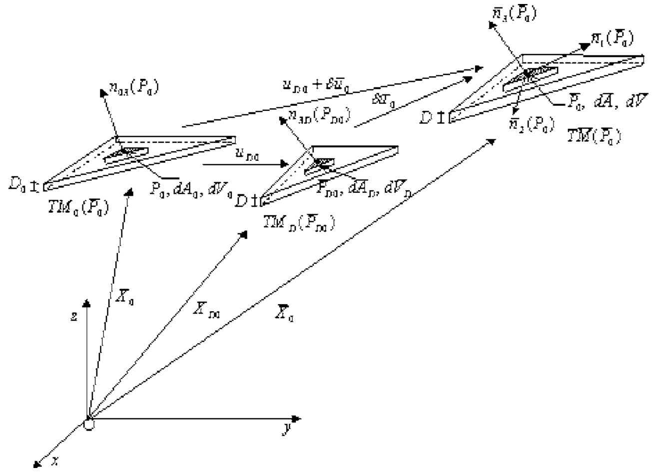


Fig. 1 Actual geometry of membrane in space

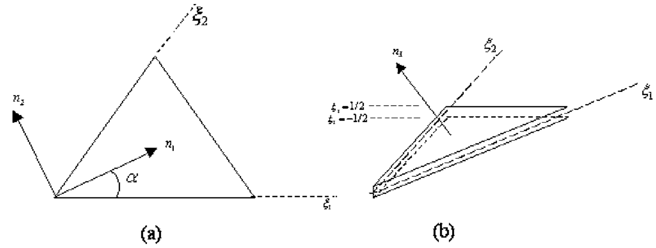


Fig. 2 (*a*) Direction of principal Cauchy frame is indicated by the angle α and (*b*) triangular element, ξ_1, ξ_2, and ξ_3 material coordinates

dergo stretching until the wrinkles just vanish whilst the tensile forces remain unaffected. On the other hand, if P_{D0} belongs to the taut zone, no changes are effected in the open neighborhood. In this way, one can define a fictive, smooth, nonwrinkled surface $\bar{M}$, characterized by tangent planes $T\bar{M}(\bar{P})$, which would provide a gross description of the surface M_D such that the strain energy density remains the same. Thus, unlike the theory by Roddeman et al. [7], the surface $\bar{M}$ is presently so formed that the wrinkles on M_D are now stretched out by a spatially continuous kinematical process. Let $\bar{P}_0$ be the image of P_{D0} on $\bar{M}$. Let dA_0, dA_{D0}, and $d\bar{A}_0$ be infinitesimal areas respectively around the points P_0, P_{D0}, $\bar{P}_0$ on $TM_0(P_0)$, $TM_D(P_{D0})$, and $T\bar{M}(\bar{P}_0)$ (see Fig. 1). If X_0, X_{D0}, and $\bar{X}_0$ are, respectively, the coordinate vectors of P_0, P_{D0}, and $\bar{P}_0$, then the actual deformation vector of P_0 may be defined as $u_{D0} = X_{D0} - X_0$. Assuming that P_{D0} is within a wrinkled zone and that $\delta\bar{u}_0$ is the stretching vector needed to move P_{D0} to $\bar{P}_0$, one readily has the identity

$$\bar{X}_0 = X_0 + u_{D0} + \delta\bar{u}_0 \quad \forall\, P_0 \in M_0 \tag{5}$$

Note that $T\bar{M}(\bar{P}_0)$ has a two-dimensional vector space structure spanned by the unit orthogonal vectors $\bar{n}_1(\bar{P}_0)$ (denoting the direction of the zero principal Cauchy stress component and hence that of stretching) and $\bar{n}_2(\bar{P}_0)$.

$$\text{Thus one has, } \vec{DB} = (\bar{n}_1 \otimes \bar{n}_1) \cdot X_{D0} \tag{6}$$

Let, $\delta\bar{u}_0 = \beta(X_0)\vec{DB} = \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1) \cdot X_{D0}$, for some $\beta(X_0) \in R^+$. Then

$$\bar{X}_0 = X_{D0} + \delta\bar{u}_0 = X_{D0} + \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1) \cdot X_{D0} \tag{7}$$

Thus, $\partial\bar{X}_0/\partial X_0 = \partial X_{D0}/\partial X_0 + \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1) \cdot \partial X_{D0}/\partial X_0 = (I + \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1)) \cdot \partial X_{D0}/\partial X_0$.

In other words

$$\bar{F}(X_0) = [I + \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1)] \cdot F_D(X_0) \tag{8}$$

where $\bar{F}(X_0)$ and $F_D(X_0)$ are the deformation gradients for the reference point P_0 corresponding to $T\bar{M}(\bar{P}_0)$ and $TM_D(P_{D0})$, respectively. From Fig. 1, one may also write down the following identities for infinitesimal volume and area elements:

$$dV_{D0} = |F_D(X_0)| dV_0 \tag{9}$$

$$\begin{aligned} d\bar{V}_0 &= |\bar{F}(X_0)| dV_0 \\ &= |[I + \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1)] \cdot F_D(X_0)| dV_0 \end{aligned}$$

$$\text{i.e., } d\bar{V}_0 = |[I + \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1)]| |F_D(X_0)| dV_0 = |[I + \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1)]| dV_{D0} \tag{10}$$

Moreover

$$d\bar{A}_0 D = |[I + \beta(X_0)(\bar{n}_1 \otimes \bar{n}_1)]| dA_{D0} D \tag{11}$$

or

$$\delta d\bar{A}_0 \triangleq d\bar{A}_0 - dA_{D0} = |\beta(X_0)(\bar{n}_1 \otimes \bar{n}_1)| dA_{D0} \tag{12}$$

where D is the membrane thickness after deformation. Equation (12), when written in an integral form, provides a qualitative de-

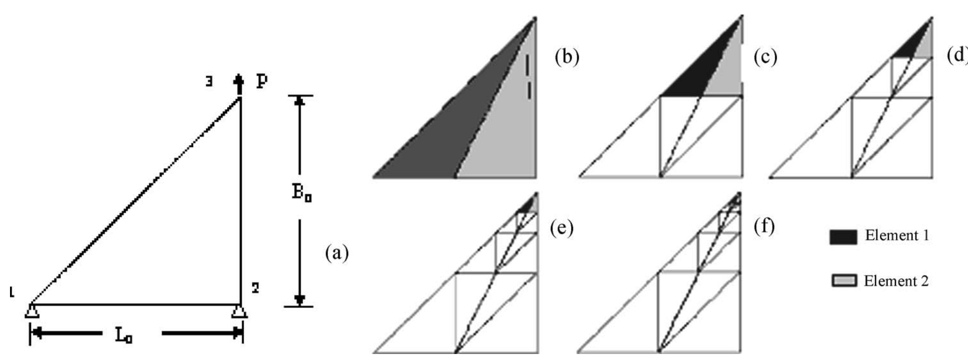


Fig. 3 (*a*) Geometry of the triangular membrane discretized with (*b*) mesh 0 (2 elements), (*c*) mesh 1 (6 elements), (*d*) mesh 2 (10 elements), (*e*) mesh 3 (14 elements), and (*f*) Mesh 4 (18 elements)

Table 1 Convergence of β at node 3 for increasing mesh density

	Uniform β			β continuously varies		
DOF	Element1	Element2	Absolute difference	Element1	Element2	Absolute difference
27	2.5×10^{-3}	2.1×10^{-3}	0.4×10^{-3}	3.9×10^{-3}	3.1×10^{-3}	0.7×10^{-3}
57	2.3×10^{-3}	1.9×10^{-3}	0.4×10^{-3}	3.2×10^{-3}	2.7×10^{-3}	0.5×10^{-3}
87	2.0×10^{-3}	1.7×10^{-3}	0.3×10^{-3}	2.6×10^{-3}	2.4×10^{-3}	0.2×10^{-3}
117	1.9×10^{-3}	1.7×10^{-3}	0.2×10^{-3}	2.2×10^{-3}	2.1×10^{-3}	0.1×10^{-3}
147	1.8×10^{-3}	1.6×10^{-3}	0.2×10^{-3}	2.1×10^{-3}	2.1×10^{-3}	0

scription of the fictitious stretching needed to be applied to an area element on M_D so as to be projected on $\bar{M}$. Since the point $X_0 \in M_0$ is arbitrarily chosen, it follows that the parameter $\beta(X)$ is a continuous function of the reference coordinate. Indeed, the continuity of $\beta(X)$ is needed so that the strain tensor

$$\bar{E} = 1/2(\bar{F}^T \cdot \bar{F} - I) \tag{13}$$

and the strain energy density $\bar{W}(\bar{F})$ defined with respect to $\bar{M}$ remain continuous. For any point in the taut region, one has $\beta(X)=0$ (implying that $\delta d\bar{A}=0$) and for any point in the wrinkled region, one has $\beta(X)>0$ (implying that $\delta d\bar{A}>0$). While $\bar{W}(\bar{F})$ is rank one convex over the taut region (guaranteeing the uniqueness of the solution in the strong sense), it is nonconvex with rank zero over the wrinkled portion [11]. When rank-one convexity of $\bar{W}(\bar{F})$ fails to hold, deformations cannot be expressed in terms of classically known smooth functions [12]. Deformations of this nature may be expressed through generalized curves, which are limits of sequences of ordinary functions with rapidly oscillating derivatives. When this happens (as in the case of the wrinkled portion of the surface M_D), the deformation gradient may generally be taken to be everywhere continuous, but nowhere differentiable over the wrinkled region. However, the fictive surface $\bar{M}$ may be thought of as a surface in the limit as the minor principal stress over the wrinkled portion goes to zero from the positive (tension) side. Consequently, over this portion, the Cauchy-Green strain tensor in the limit may be written as

$$G = \bar{F}^T\bar{F} = \bar{\lambda}^2\bar{n}_2 \otimes \bar{n}_2 + \bar{\mu}^2\bar{n}_1 \otimes \bar{n}_1 \text{ with } \bar{\mu} \to 1^+,\ \bar{\lambda} > 1 \tag{14}$$

where $\bar{\lambda}$ and $\bar{\mu}$ are the principal stretches. Since $\beta(X)$ is a differentiable function for the taut region ($\bar{\lambda} > \bar{\mu} > 1$), it may be considered as a limit of a sequence of continuous and differentiable (almost everywhere) functions for portions where Eq. (14) holds. In other words, for finite precision computations, one may actually consider $\bar{F}$ to be an almost-everywhere differentiable tensor function on $\bar{M}$, a premise that is not valid for M_D. It is also worth noting $\bar{\lambda}$ and $\bar{\mu}$ on $\bar{M}$ are associated with the following two invariants under the assumption of incompressibility:

$$I_1 = \lambda\mu = (\det G)^{1/2} \tag{15}$$

$$I_2 = \lambda^2 + \mu^2 = (\mathrm{tr}\, G + 2I_1)^{1/2} \tag{16}$$

It is interesting to note that a similar modification for the fictive deformation gradient as in Eq. (8) is possible, in principle, for analyzing slack regions as well. In particular, one may introduce two parameter functions $\beta_1(X)$ and $\beta_2(X)$ so that the expression for the modified deformation gradient becomes

$$\bar{F}(X) = [I + \beta_1(X)(\bar{n}_1 \otimes \bar{n}_1) + \beta_2(X)(\bar{n}_2 \otimes \bar{n}_2)] \cdot F_D(X) \tag{17}$$

3 The Finite Element (FE) Formulation

The strong form of equilibrium equations may be written as:

$$\nabla \cdot \sigma = 0 \tag{18}$$

The integral (weak) form of the equilibrium equation (18) is [7]

$$\int_{V_0} \pi{:}(\nabla_0\phi)^T dV_0 = \int_{A_0} f_0 \cdot \phi dA_0 \tag{19}$$

where ϕ is a kinematically consistent vector function, $\pi = S^{-1}(J\sigma)$ is the first Piola Kirchhoff (nonsymmetric) stress tensor; S is the deformation gradient (of TM_D with respect to TM_0) without considering variations in membrane thickness and f_0 is the force on the deformed surface transformed to the undeformed (initial) surface and ∇_0 is the gradient operator with respect to the initial configuration. Equation (19) is solved by the finite element method taking $\beta(X)=\beta(x,y)$ as a spatially varying, differentiable function, which may be discretized and interpolated through a usual FE procedure like the other dependent variables. Obviously, element sizes become infinitesimally small, β in different elements also approach constant values and in that sense the proposed FE model apparently approaches the one proposed by Roddeman et al. [7]. If the element size is finite, $\bar{F}$ via Roddeman's theory would be discontinuous and this appears to be inconsistent with the physics of the problem. Also, Roddeman's theory does not admit a solution through a mesh-free or an element-free approach.

The generic form of the position vector of a material point in a triangular isoparametric element is given by

$$\tilde{X}(\xi_1,\xi_2) = \psi_k(\xi_1,\xi_2)\tilde{X}_k + \xi_3\tilde{D}(\xi_1,\xi_2)\tilde{n}_3(\xi_1,\xi_2) \tag{20}$$

where ψ_k are shape functions to determine $\tilde{X}$, ξ_1, ξ_2, and ξ_3 (Fig. 2) are isoparametric material coordinates, $\tilde{D}(\xi_1,\xi_2)$ is the generic thickness at the material position (ξ_1,ξ_2), $\tilde{X}_k$ are position vectors of the nodal points, $\tilde{n}_3$ is the outward normal vector at (ξ_1,ξ_2).

Now as in Ref. [7] the real stresses in wrinkled membrane are determined by modifying the deformation tensor in the constitutive equation as

$$(J\sigma)_{\xi_1\xi_2} = \bar{F}_{\xi_1\xi_2}H(\bar{E}_{\xi_1\xi_2})\bar{F}^T_{\xi_1\xi_2} \tag{21}$$

$$\pi_{\xi_1\xi_2} = \bar{S}^{-1}_{\xi_1\xi_2}(J\sigma_{\xi_1\xi_2}) \tag{22}$$

with

$$\bar{F}_{\xi_1\xi_2} = \{I + \beta(\xi_1,\xi_1)[\bar{n}_1(\xi_1,\xi_2) \otimes \bar{n}_1(\xi_1,\xi_2)]\}F_{\xi_1\xi_2} \tag{23}$$

$$\bar{S}_{\xi_1\xi_2} = \{I + \beta(\xi_1,\xi_1)[\bar{n}_1(\xi_1,\xi_2) \otimes \bar{n}_1(\xi_1,\xi_2)]\}S_{\xi_1\xi_2} \tag{24}$$

The direction of the principal Cauchy frame is determined by the angle $\alpha(\xi_1,\xi_2)$. α, β, and D are determined by using the coupled conditions

Table 2 Convergence of β at node 3 with increasing node numbers

6-noded		9-noded		15-noded	
Total DOF	Absolute difference in β	Total DOF	Absolute difference in β	Total DOF	Absolute difference in β
27	0.7×10^{-3}	42	0.6×10^{-3}	75	0.5×10^{-3}
57	0.5×10^{-3}	93	0.2×10^{-3}	129	0.05×10^{-3}
87	0.2×10^{-3}	117	0.1×10^{-3}	207	0
117	0.1×10^{-3}	144	0.0025×10^{-3}	–	–
147	0	201	0	–	–

(a) there is no stress in wrinkling direction, i.e.,

$$\bar{n}_1 \cdot (J\sigma) \cdot \bar{n}_1 = 0 \tag{25}$$

(b) the frame $\bar{n}_1$, $\bar{n}_2$, $\bar{n}_3$ is a principal frame, i.e.,

$$\bar{n}_1 \cdot (J\sigma) \cdot \bar{n}_2 = 0 \tag{26}$$

(c) the membrane is in a state of plane stress, i.e.,

$$\bar{n}_3 \cdot (J\sigma) \cdot \bar{n}_3 = 0 \tag{27}$$

Once the modified stress tensor is calculated one obtains the finally projected form of the response equations as

$$K_{lk}(U)u_k = f_l \tag{28}$$

where K_{lk} is the (l,k)th 3×3 submatrix of the nonlinear stiffness matrix, $U=\{u_k\}$ the discretized displacement vector and $f_l \in R^3$ the prescribed nodal force on node l.

Numerical Implementation and Examples. A limited numerical exploration of the proposed algorithm for static and dynamic wrinkling analysis of a linear and isotropic membrane is conducted with the following properties: $E=270$ MPa, $\upsilon=0.4$, $D_0=1$ mm. First consider a plane triangular membrane subjected to a concentrated force as shown in Fig. 3. The membrane is modeled through six-noded triangular elements. Values of β at apex point 3 with increasing mesh density towards node 3 are shown in Table 1 and checked against a constant approximation for β across a given element. Next, a p-convergence study is performed using triangular elements with increasing number of nodes, i.e., 6-, 9-, and 15-noded elements. The convergence of β at point 3 for different element types is shown in Table 2. It is seen that a superior rate of convergence is achievable through a continuous representation of β as compared with the original model by Roddeman et al. [1].

4 FE Analysis of Dynamically Wrinkled Membranes

The strong form of the equations of motion is

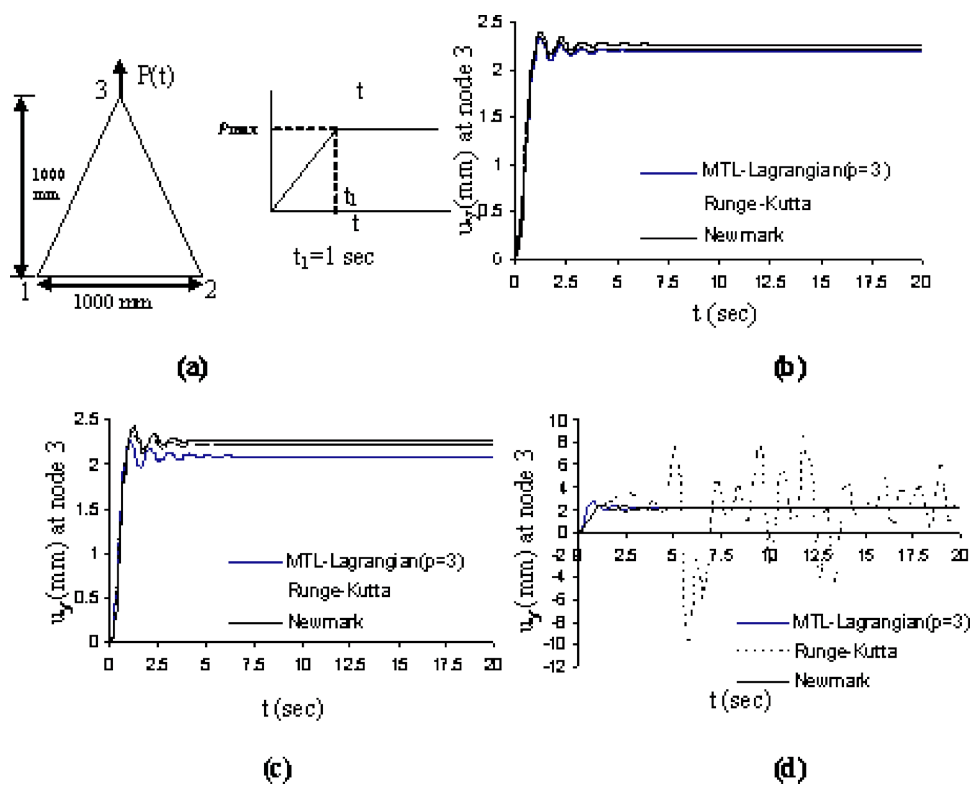


Fig. 4 (*a*) Geometry and loading pattern of the triangular membrane; comparison of displacement history at node 3 via MTL-Lagrangian, Newmark (β=0.25, γ=0.5), and Runge-Kutta method for time-varying load, (*b*) h=0.01 s; (*c*) h=0.05 s; (*d*) h=0.2 s

$$\nabla \cdot \sigma + c\frac{\partial u}{\partial t} + \rho\frac{\partial^2 u}{\partial t^2} = 0 \qquad (29)$$

where ρ and c are, respectively, the mass density and viscous damping coefficient. For undamped systems, one has $c=0$. ∇ is the gradient operator on the tangent space $T\bar{M}$ associated deformed (fictive) configuration $\bar{M}$ and σ is the real Cauchy stress tensor. The integral (weak) form of Eq. (29) is

$$\int_V (\nabla \cdot \sigma) \cdot \phi dV + c\int_V \frac{\partial u}{\partial t} \cdot \phi dV + \rho\int_V \frac{\partial^2 u}{\partial t^2} \cdot \phi dV = 0 \qquad (30)$$

where the vector function ϕ work as both the test and the shape functions. Considering the midsurface ($\xi_3=0$), solution u of Eq. (30) is approximated as $u(\xi_1,\xi_2,t)=\phi_k(\xi_1,\xi_2)u_k(t)$ so that the discretized equations of motion, after assembly over elements and with respect to a fixed reference frame, are

$$[M]\{\ddot{U}\} + [C]\{\dot{U}\} + [K(\{U\},t)]\{U\} = \{F(t)\} \qquad (31)$$

Numerical Examples. A plane triangular membrane shown in Fig. 4(*a*) is considered. The membrane is subjected to a linearly time-varying load. Material properties and the membrane thickness are taken to be the same as those in Sec. 3. Numerical integration has been performed using two second order implicit methods (i.e., the multistep transversal linearization (MTL) [13] and Newmark methods and a fourth order explicit Runge-Kutta method. In Figs. 4(*b*)–4(*d*), displacement histories at node 3, obtained via the MTL, Newmark, and Runge-Kutta methods, are plotted for different choices of the temporal step size h. It is observed that both the implicit methods work in a stable and accurate manner even for much higher values of h (such as $h=0.2$) while the explicit Runge-Kutta method yields inaccurate results for such values of h (even though it has a higher formal order of accuracy than MTL and Newmark). The implicit methods are computationally a little costlier. However, numerical experiments indicate that if one uses the maximum permissible h for each of these methods (without compromising the numerical accuracy), the computing costs in all these methods are well comparable. Nevertheless, given a rather sensitive dependence of explicit methods to h, implicit methods remain the preferred option.

5 Concluding Remarks

A weak, finite element-based formulation for static and dynamic analyses of wrinkled membranes is presented and numerically explored to an extent. A spatially continuous and differentiable (almost everywhere) parameter function $\beta(X)$ is used to modify the deformation gradient so the computed solution may be projected on a fictive nonwrinkled surface. This is the main point of departure of the proposed method from the one by Roddeman et al. [1]. Through limited numerical illustrations, it is verified that the present formulation allows for a spatially accurate determination of wrinkling locations and orientations. Since one has $\beta(X) > 0$ only over the wrinkled regions, it follows that $\beta(X)$ is localized and compactly supported. Thus the use of wavelet-based interpolating functions may help in a more accurate prediction of $\beta(X)$ with significantly lesser computational costs. Under dynamic loading, the weakly projected membrane equations are numerically integrated via second-order MTL, Newmark (both implicit), and a fourth-order Runge-Kutta (explicit) schemes. It is observed that both the implicit schemes offer consistently better results than the Runge-Kutta scheme, especially for higher step sizes.

References

[1] Wagner, H., 1929, "Flat Sheet Metal Girder With Very Thin Metal Web," In Zeitschriftfür Flugtechnik und Motorluftschiffahrt **20**, pp. 200–207.
[2] Mansfield, E. H., 1970, "Load Transfer via a Wrinkled Membrane," Proc. R. Soc. London, Ser. A **316**, pp. 269–289.
[3] Pipkin, A. C., 1986, "The Relaxed Energy Density for Isotropic Elastic Membrane," IMA J. Appl. Math. **36**, pp. 85–99.
[4] Miller, R. K., Hedgepeth, J. M., Weingarten, V. I., Das, P., and Kahyai, S., 1985, "Finite Element Analysis of Partly Wrinkled Membranes," Comput. Struct. **20**, pp. 631–639.
[5] Stein, M., and Hedgepeth, J. M., 1961, "Analysis of Partly Wrinkled Membranes," Tech. Rep. NASA TN D-813.
[6] Wu, C. H., 1978, "Nonlinear Wrinkling of Nonlinear Membranes of Revolution," ASME J. Appl. Mech. **45**, pp. 533–538.
[7] Roddeman, D. G., Drukker, J., Oomens, C. W. J., and Janssen, J. D., 1987, "The Wrinkling of Thin Membranes: Part I—Theory; Part II—Numerical Analysis," ASME J. Appl. Mech. **54**, pp. 884–892.
[8] Lu, K., Accorsi, M., and Leonard, J., 2001, "Finite Element Analysis of Membrane Wrinkling," Int. J. Numer. Methods Eng. **50**, pp. 1017–1038.
[9] Epstein, M., and Forcinito, M. A., 2001, "Anisotropic Membrane Wrinkling: Theory and Analysis," Int. J. Solids Struct. **38**, pp. 5253–5272.
[10] Raible, T., Tegeler, K., Löhnert, S., and Wriggers, P., 2005, "Development of a Wrinkling Algorithm for Orthotropic Membrane Materials," Comput. Methods Appl. Mech. Eng. **194**, pp. 2550–2568.
[11] Graves, L M., 1939, "The Weierstrass Condition for Multiple Integral Variation Problems," Duke Math. J. **5**, pp. 556–591.
[12] Young, L. C., 1969, "Lectures on Calculus of Variation and Optimal Control Theory, W. B. Saunders, Philadelphia.
[13] Roy, D., and Kumar, R., 2005, "A Multi-Step Transversal Linearization (MTL) Method in Non-Linear Structural Dynamics," J. Sound Vib. **287**, pp. 203–226.

Nonlinear Stability of Thermoelastic Sliding Contact

Jason D. Miller

D. Dane Quinn

Department of Mechanical Engineering,
The University of Akron,
Akron, OH 44325-3903

A model for sliding contact of a thermoelastic rod is considered and is subjected to a multiple scales analysis to uncover its nonlinear behavior near a neutrally stable state. The analysis reveals a combination of the contact resistance and frictional intensity that describes the generic unfolding of this critical state and its associated bifurcations. In particular, the system can describe how two equilibria coalesce in a saddle-node bifurcation and generalizes stability criteria that have been presented previously in the literature for this model. Moreover, this analysis describes the role of the initial deformation of the rod on its long-term dynamical behavior. [DOI: 10.1115/1.2423034]

1 Introduction and Formulation

The frictional heating that accompanies relative motion between two surfaces can give rise to thermoelastic deformation, leading to qualitative changes in the dynamical behavior of the system, including frictionally excited thermoelastic instabilities (TEI) (see Ref. [1], and references therein). Such situations arise in engineering applications incorporating brakes and clutches, as well as bearings and other mechanical components in which the possibility for fictional sliding exists. This response can be modeled whereby both heat generation and the thermal contact resistance depend on the contact pressure [2]. The latter can be assumed to arise from surface roughness between contacting bodies and alone can lead to instability and nonuniqueness of equilibria in thermoelastic contact problems [3]. Ciavarella et al. [4] studied the interaction of the thermal contact resistance and frictional heating in a thermoelastic rod contacting a rigid wall. In this work the frictional heating q per unit area was assumed to be

$$q = fVp$$

where f is the coefficient of friction, V is the contact velocity, and p is the contact pressure. Above some critical sliding speed, they found that for some initial conditions the contact pressure grows without bound (seizure) while below the critical speed the system always tends to a steady state. Likewise, Afferrante and Ciavarella [5] considered TEI in the presence of sliding friction within the "Aldo model" in which multiple rods are in contact of the sliding surface, with the total normal force is prescribed. While seizure is no longer possible in this model, for certain parameter values the uniform-pressure solution is no longer unique and can be unstable.

In the absence of frictional sliding, Quinn and Pelesko [6] investigated the nonlinear stability of equilibrium states in a one dimensional model of a thermoelastic rod using a multiple scales analysis. When the linear stability of the steady state is marginal, the equilibrium solution was shown to exhibit a fold type bifurcation. In the present paper, a multiple scales analysis is again applied to a one-dimensional thermoelastic rod with pressure-dependent effects arising from both the thermal contact resistance and frictional sliding.

Consider a one-dimensional thermoelastic rod suspended between a rigid wall and a moving surface. The interface between the moving surface and the stationary rod is assumed to be rough, so that frictional sliding occurs, which generates heat dependent on the surface characteristics and contact pressure. The nondimensional temperature is defined as $\theta(x,t)$, and at the wall $\theta(0,t)=0$, while the temperature of the surface is assumed to be constant, equal to 1. The rod is assumed to be homogeneous and isotropic, possessing constant thermal and linear elastic material properties, and the nondimensional equations describing the thermal field can be written as

$$\frac{\partial \theta}{\partial t} = \frac{\partial^2 \theta}{\partial x^2} \tag{1a}$$

$$\theta(0,t) = 0 \tag{1b}$$

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

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

In these equations, η describes the (nondimensional) stress-free length of the rod and is proportional to the contact pressure p of Ciavarella et al. [4]. The parameter γ is a nondimensional coefficient of thermal expansion and is proportional to the temperature difference between the wall and the sliding surface.

$R(\eta)$ is the contact resistance function while $F(\eta)$ describes the heat generated by the frictional sliding at $x=1$. The contact resistance function $R(\eta)$ and the frictional intensity function $F(\eta)$ are central to the analysis of the behavior of this thermoelastic system. In this formulation $F(\eta)$ can be identified with q as described by Ciavarella et al. [4], although in what follows no specific constitutive relationship will be assumed. Intuitively one can imagine that there would be neither conductive heat transfer nor frictional heat generation between the thermoelastic rod and the sliding belt if there is a gap, while under contact as the normal pressure is increased both the thermal resistance and frictional intensity vary. Both of these constitutive functions depend on the η. In the absence of $F(\eta)$, these equations are identical to those considered by Quinn and Pelesko [6].

The contact resistance function $R(\eta)$ and the reciprocal contact resistance function $K(\eta)=1/[1+R(\eta)]$, as defined by Quinn and Pelesko [6], are used for this model. The frictional intensity function is of the form

$$F(\eta) \geqslant 0, \ \forall\ \eta \quad \text{and} \quad \lim_{\eta \to -\infty} F(\eta) = 0$$

For this study it will be convenient to define the contact function $H(\eta)$ as

$$H(\eta) = F(\eta)\left(\frac{R(\eta)}{1+R(\eta)}\right)$$

which has the following properties:

$$H(\eta) \geqslant 0, \ \forall\ \eta \quad \text{and} \quad \lim_{\eta \to -\infty} H(\eta) = 0$$

2 Nonlinear Theory

The method of multiple scales is used to account for the nonlinearities present in the governing equations. We begin the nonlinear analysis by expanding the bifurcation parameter γ as a series in ϵ

$$\gamma = \gamma_0 + \epsilon\gamma_1 + \epsilon^2\gamma_2 + \epsilon^3\gamma_3 + \cdots$$

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 19, 2004; final manuscript received October 5, 2005. Review conducted by N. Sri Namachchivaya.

In the following analysis the system is assumed to be neutrally stable when $\gamma=\gamma_0$, that is, when $\epsilon=0$. Hence, ϵ characterizes the distance, in terms of the bifurcation parameter γ, from neutral stability. The conditions on K and H corresponding to neutral stability will arise in the analysis below. The method of multiple scales stretches time as

$$t_0=t, \quad t_1=\epsilon t, \quad t_2=\epsilon^2 t, \ldots$$

and both θ and η are expanded in ϵ

$$\theta(x,t)=\theta_0(x)+\epsilon\theta_1(x,t)+\epsilon^2\theta_2(x,t)+\cdots$$

$$\eta=\eta_0+\epsilon\eta_1+\epsilon^2\eta_2+\cdots$$

Finally, the reciprocal contact resistance function $K(\eta)$ and the reciprocal frictional intensity function $H(\eta)$ are expanded in ϵ about η_0. Following Quinn and Pelesko [6], a marginally stable equilibrium point is assumed to exist at $\eta=\eta_0$, so that $K(\eta_0)=K_0$ and $H(\eta_0)=H_0$. The behavior of the solution near this state is investigated through the introduction of the following scalings:

$$\begin{aligned} K_0' &= K_0'+\epsilon^2 k_0', \quad H_0'=H_0'+\epsilon^2 h_0' \\ K_0'' &= \epsilon k_0'', \quad H_0''=\epsilon h_0'' \\ K_0''' &= k_0''', \quad H_0'''=h_0''' \end{aligned} \tag{2}$$

so that $H(\eta)$ and $K(\eta)$ are written as

$$K(\eta)=K_0+(K_0'\eta_1)\epsilon+(K_0''\eta_2)\epsilon^2+\left[k_0'\eta_1+\frac{k_0''}{2}\eta_1^2+\frac{k_0'''}{6}\eta_1^3+K_0'\eta_3\right]\epsilon^3+\cdots$$

$$H(\eta)=H_0+(H_0'\eta_1)\epsilon+(H_0''\eta_2)\epsilon^2+\left[h_0'\eta_1+\frac{h_0''}{2}\eta_1^2+\frac{h_0'''}{6}\eta_1^3+H_0'\eta_3\right]\epsilon^3+\cdots$$

so that h_0', k_0', h_0'', k_0'', h_0''', and k_0''' characterize the boundary condition Eq. (1c) at $x=1$ near the marginally stable state, recovered for $\epsilon=0$. These expansions are introduced into Eqs. (1) and terms are collected at each order of ϵ

$$\mathcal{O}(\epsilon^0): \frac{\partial\theta_0}{\partial t_0}=\frac{\partial^2\theta_0}{\partial x^2}, \quad \theta_0(0,t)=0$$

$$(1-K_0)\cdot\frac{\partial\theta_0}{\partial x}(1,t)=K_0\cdot(1-\theta(1,t))+H_0$$

$$\eta_0=-\gamma_0\int_0^1\theta_0(x,t)dx \tag{3}$$

$$\mathcal{O}(\epsilon^1): \frac{\partial\theta_0}{\partial t_1}+\frac{\partial\theta_1}{\partial t_0}=\frac{\partial^2\theta_1}{\partial x^2}, \quad \theta_1(0,t)=0$$

$$\begin{aligned}(1-K_0)\frac{\partial\theta_1}{\partial x}(1,t)-(K_0'\eta_1)\frac{\partial\theta_0}{\partial x}(1,t) \\ =-K_0\theta_1(1,t)+(K_0'\eta_1)[1-\theta_0(1,t)]+(H_0'\eta_1)\end{aligned}$$

$$\eta_1=-\gamma_0\int_0^1\theta_1(x,t)dx-\gamma_1\int_0^1\theta_0(x,t)dx \tag{4}$$

The $\mathcal{O}(\epsilon^2)$ and $\mathcal{O}(\epsilon^3)$ equations, while straightforward, are not shown for succinctness of presentation.

The multiple scales analysis proceeds by solving the above system of equations at each order in ϵ, which results in a series of linear variational equations. To ensure that the solution to these equations is bounded, we remove all secular terms by appropriate choices which ultimately determine the nonlinear stability of the system.

At lowest order, the dominant solution is given as $\theta_0(x,t)=A_0x$, where

$$A_0=K(\eta_0)+H(\eta_0)=\text{constant}, \quad \eta_0=\frac{-\gamma_0}{2}A_0 \tag{5}$$

This expression is equivalent to the condition for the steady-state solution identified by Ciavarella et al. [4]. From the lowest order solution one can identify A_0, the thermal gradient, with η_0, the nondimensional variable related to the thermal expansion.

Moving on to $\mathcal{O}(\epsilon)$, we again assume a separable solution. We find that the eigenvalues of the characteristic equation are all non-positive, and decay exponentially, with the exception of the solution of the form

$$\theta_1(x,t)=A_1(t_1,t_2,t_3,\ldots)x$$

The coefficient A_1 is allowed to vary on timescales $t=\epsilon t$ and slower. This will be denoted by $A_1(t_1)$, and will explicitly list only the fastest time scale. The remaining have been suppressed for clarity. This solution corresponds to the eigenvalue at the origin and leads to the solvability equation

$$\left[1+\frac{\gamma_0}{2}(K_0'+H_0')\right]A_1+\left[\frac{\gamma_1}{2}(K_0'+H_0')\right]A_0=1$$

This equation is satisfied for all values of A_1 provided

$$\gamma_0=-\frac{2}{(K_0'+H_0')}, \quad \gamma_1=0$$

This condition is necessary for the neutral stability of the system for $\epsilon=0$.

At order ϵ^2 the solution has the general form:

$$\theta_2(x,t)=A_2(t_1)x+\frac{\partial A_1}{\partial t_1}\frac{x^3}{6}$$

whose solvability equation is

$$\frac{\partial A_1}{\partial t_1}\left(\frac{5}{12}-\frac{K_0}{3}\right)-\left(\frac{\gamma_2}{\gamma_0}\right)A_0=0$$

which is satisfied for

$$\frac{\partial A_1}{\partial t_1}=0, \quad \gamma_2=0$$

With this, A_1 does not vary on the time scale t_1, and can be written as $A_1=A_1(t_2)$.

Finally at order ϵ^3 the solution has a general form

$$\theta_3(x,t)=A_3(t_1)x+\left(\frac{\partial A_1}{\partial t_2}+\frac{\partial A_2}{\partial t_1}\right)\frac{x^3}{6}$$

and the solvability equation is

$$\begin{aligned}0=\left(\frac{5-4K_0}{12}\right)\left(\frac{\partial A_2}{\partial t_1}+\frac{\partial A_1}{\partial t_2}\right)+\frac{\gamma_0}{2}(k_0'+h_0')A_1-\left(\frac{\gamma_0}{2}\right)^2\frac{(k_0''+h_0'')}{2}A_1^2 \\ +\left(\frac{\gamma_0}{2}\right)^3\frac{(k_0'''+h_0''')}{6}A_1^3-\frac{\gamma_3}{\gamma_0}A_0\end{aligned}$$

which is satisfied for $\partial A_2/\partial t_1=0$ and

$$\begin{aligned}\frac{\partial A_1}{\partial t_2}=\frac{12}{5-4K_0}\Bigg[\left(\frac{\gamma_3}{\gamma_0}A_0\right)-\frac{\gamma_0}{2}(k_0'+h_0')A_1+\left(\frac{\gamma_0}{2}\right)^2\frac{(k_0''+h_0'')}{2}A_1^2 \\ -\left(\frac{\gamma_0}{2}\right)^3\frac{(k_0'''+h_0''')}{6}A_1^3\Bigg]\end{aligned}$$

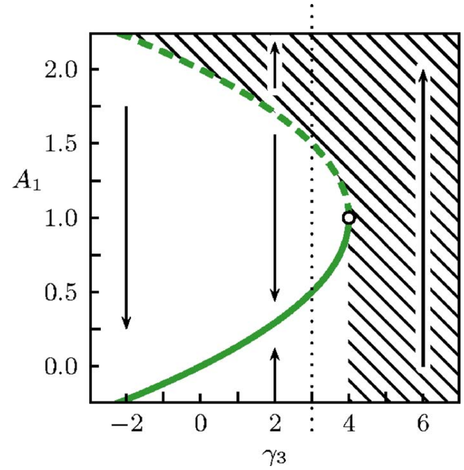


Fig. 1 Bifurcation diagram ($\beta_1/(2\beta_2)=1.00$, $\beta_0/\beta_2=0.25$, $\beta_2>0$). Stable equilibrium solutions are denoted by the solid line, while the dashed branch is unstable. The arrows denote the general evolution of initial conditions and in the shaded region initial conditions grow unbounded. The location of the saddle-node bifurcation is marked with the open circle. Finally, the response shown in Fig. 2 occurs for the dotted line at $\gamma_3=3.00$.

This is the order ϵ^1 correction to the dominant solution and determines the long term behavior of A_1, the amplitude of $\theta_1(x,t)$. For $H(\eta)\equiv 0$ this evolution equation reduces to that obtained by Quinn and Pelesko [6].

This equation may be represented as

$$\frac{\partial A_1}{\partial t_2} = c\left[\beta_0\gamma_3 - \beta_1 A_1 + \beta_2 A_1^2 - \beta_3 A_1^3\right] \tag{6}$$

where

$$\beta_0 = \frac{A_0}{\gamma_0}, \quad \beta_1 = \frac{\gamma_0}{2}(k_0' + h_0')$$

$$\beta_2 = \left(\frac{\gamma_0}{2}\right)^2 \frac{(k_0'' + h_0'')}{2}, \quad \beta_3 = \left(\frac{\gamma_0}{2}\right)^3 \frac{(k_0''' + h_0''')}{6}$$

and the quantity $c=12/(5-4K_0)$ may be scaled to unity by an appropriate time transformation. Physically the response of A_1 determined from Eq. (6) represents the $\mathcal{O}(\epsilon)$ correction to the thermal gradient of the rod at the neutrally stable stationary state, as defined by A_0 (cf. Eq. (5)). At the wall ($x=0$) the nondimensional temperature is held fixed so that as A_1 increases the temperature at the sliding surface increases as well. Equilibrium solutions for A_1 correspond to stationary thermal gradients in the rod.

For $\beta_3\neq 0$, the analysis of Eq. (6) is identical to that contained in Quinn and Pelesko [6] and was shown to possess a generic fold type bifurcation near the neutrally stable state. In contrast, for $\beta_3=0$, the system undergoes a saddle-node bifurcation as the bifurcation parameter γ_3 is varied, representing the coalescence of equilibria described by Ciavarella et al. [4]. The equilibrium locations are given as $A_1\equiv\lambda$ with

$$\lambda_{1,2} = \frac{\beta_1}{2\beta_2} \pm \sqrt{\left(\frac{\beta_1}{2\beta_2}\right)^2 - \frac{\beta_0\gamma_3}{\beta_2}}$$

which only exist for

$$\frac{\beta_0\gamma_3}{\beta_2} \leqslant \left(\frac{\beta_1}{2\beta_2}\right)^2$$

As illustrated in Fig. 1, when this inequality is satisfied a pair of equilibria exist with one stable and the other unstable (for $\gamma_3<4$ in the figure). In the contrary case, no equilibrium solutions exist and for $\beta_2>0$ ($\beta_2<0$) the thermal gradient increases (decreases) without bound. Recall that unbounded growth in the thermal gradient corresponds to unbounded growth in the contact pressure. As γ_3 varies near this neutrally stable state, the pair of equilibria undergoes a saddle-node bifurcation, occurring at $\gamma_3=4$ in the figure.

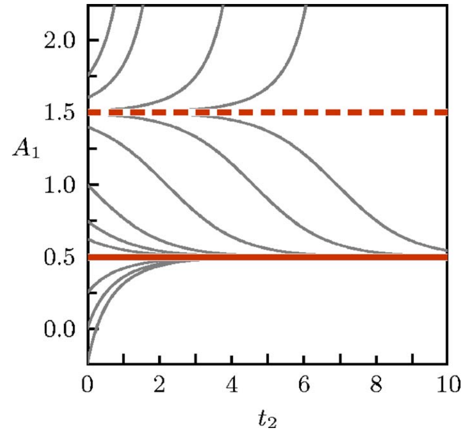


Fig. 2 Solutions for $\lambda_{1,2}=\{1.5,0.5\}$ ($\beta_0=0.25$, $\beta_1=2.00$, $\beta_2=1$, $\gamma_3=3.00$). The equilibrium solutions are shown as thick lines (solid—stable, dashed—unstable).

Near the saddle-node bifurcation ($\beta_3=0$), provided the equilibrium points exist, Eq. (6) is separable and may be solved as

$$A_1(t_2) = \frac{\lambda_1(A_1(0)-\lambda_2)e^{(\lambda_2-\lambda_1)\beta_2 t_2} - \lambda_2[A_1(0)-\lambda_1]}{(A_1(0)-\lambda_2)e^{(\lambda_2-\lambda_1)\beta_2 t_2} - [A_1(0)-\lambda_1]}$$

As illustrated in Fig. 2, for a pair of real roots ($\lambda_1>\lambda_2$) with $\beta_2>0$, if $A(0)<\lambda_1$ the solution exponentially approaches λ_2 as t_2 increases. For $A(0)>\lambda_1$ the solution to this equation is singular and increases without bound in finite time $t_{2,b}$, given as

$$t_{2,b} = \frac{1}{(\lambda_2-\lambda_1)\beta_2}\ln\left(\frac{A_1(0)-\lambda_1}{A_1(0)-\lambda_2}\right)$$

In Fig. 1, initial conditions that grow unbounded lie in the hatched region while all other initial states approach the stable equilibrium branch. The arrows denote the evolution of initial conditions. Recall however that Eq. (6) is only valid for $\mathcal{O}(1)$ values of A_1, so that the predictions of this equation cannot be trusted to reflect the behavior of the original thermoelastic system for $A_1=\mathcal{O}(1/\epsilon)$, though the solution to the original model is expected to move away from the unstable state and as the thermal gradient increases can lead to seizure [4]. The solution for $\beta_2<0$ is equivalent to the former through a time reversal, so that the bounded solution approaches λ_1 and the unbounded solution occurs for $A_1(0)<\lambda_2$.

3 Conclusions

A model for thermoelastic sliding contact was subjected to a multiple scales analysis to uncover its nonlinear behavior near a neutrally stable state. The analysis reveals a combination of the contact resistance and frictional intensity that describes the generic unfolding of this critical state and its associated bifurcations. Specifically, in a neighborhood of the saddle-node bifurcation identified through this analysis, the system contains a pair of equilibrium states, one of which is stable while the remaining is unstable. One consequence of the topology of this system is that when this pair of equilibria exists a large region of initial conditions grow and leave this neighborhood. In addition, as this pair coalesces in the saddle-node bifurcation, all initial conditions in this neighborhood evolve away from this region. This generalizes the stability criteria presented by Ciavarella et al. [4] for frictional heating that is proportional to contact pressure. In particular, this analysis identifies the quantity

$$K(\eta) + H(\eta) = \frac{1 + F(\eta)R(\eta)}{1 + R(\eta)}$$

which describes the bifurcation of equilibria, analagous to the role of the reciprocal contact resistance in the absence of frictional sliding [6].

It must be emphasized that the present analysis is concerned with the behavior of the system in a region localized around the aforementioned neutrally stable state. Thus solutions that grow unbounded in this model correspond to solutions that evolve away from their initial state. As described by Ciavarella et al. [4], under general conditions stable and unstable steady-state solutions alternate. In their analysis, below some critical sliding speed the outlying states were stable and all initial conditions lead to steady-state sliding. However, above the identified critical speed one of the outlying equilibria was unstable. In this case, initial conditions that are predicted to evolve away from the neighborhood of the neutrally stable state can possibly lead to seizure.

References

[1] Barber, J. R., and Ciavarella, M., 2000, "Contact Mechanics," Int. J. Solids Struct., **37**(1–2), pp. 29–43.

[2] Johansson, L., and Klarbing, A., 1993, "Thermoelastic Frictional Contact Problems: Modelling, Finite Element Approximation and Numerical Realization," Comput. Methods Appl. Mech. Eng., **105**, pp. 181–220.

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

[4] Ciavarella, M., Johansson, L., Afferrante, L., Klarbring, A., and Barber, J. R., 2003, "Interaction of Thermal Contact Resistance and Frictional Heating in Thermoelastic Instability," Int. J. Solids Struct., **40**(21), pp. 5583–5597.

[5] Afferrante, L., and Ciavarella, M., 2004, "The Thermoelastic Aldo Contact Model With Frictional Heating," J. Mech. Phys. Solids, **52**, pp. 617–640.

[6] Quinn, D. D., and Pelesko, J. A., 2002, "Generic Unfolding of the Thermoelastic Contact Instability," Int. J. Solids Struct., **39**(1), pp. 145–157.

Journal of Applied Mechanics

Erratum

Erratum: "A Geometrically Nonlinear Shear Deformation Theory for Composite Shells" (J. Appl. Mech., 2004, 71(1), pp. 1–9)

Wenbin Yu and Dewey H. Hodges

The wrong figure was used for Fig. 1 of the subject paper [1]. Instead, this Fig. 1 should have been used, as was pointed out in the galley proof corrections sent to ASME on December 23, 2003.

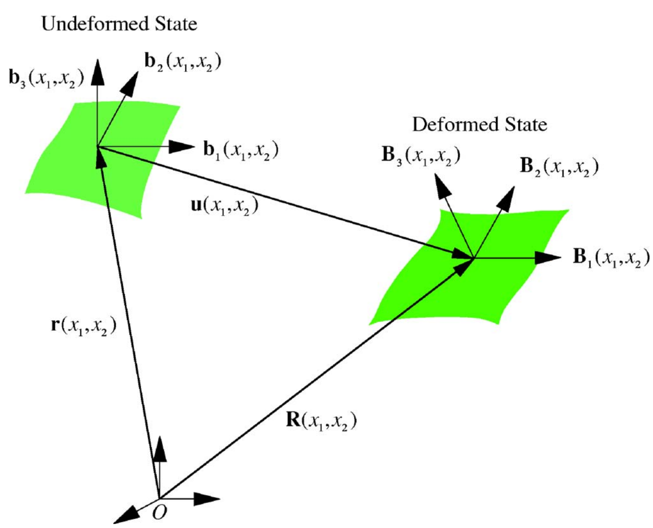


Fig. 1 Schematic of shell deformation

References

[1] Yu, W., and Hodges, D. H., 2004, "A Geometrically Nonlinear Shear Deformation Theory for Composite Shells," J. Appl. Mech. **71**(1), pp. 1–9.

