Transactions of the ASME®

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

Reduced-Order Models for Nonlinear Dynamical

Journal of Applied Mechanics

Published Bimonthly by The American Society of Mechanical Engineers

VOLUME 70 • NUMBER 4 • JULY 2003

TECHNICAL PAPERS

BRIEF NOTES

(**Contents continued on inside back cover**)

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

Journal of Applied Mechanics **Volume 70, Number 4** **JULY 2003**

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

Alloys Index, Aluminum Industry Abstracts, Applied Science & Technology Index, AMR Abstracts Database, 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

Energy Considerations in Systems With Varying Stiffness

J. R. Barber
K. Grosh
S. Oh

Department of Mechanical Engineering,
University of Michigan,
Ann Arbor, MI 48109-2125

If the stiffness of an elastic system changes with time, a conventional Newtonian statement of the equations of motion will generally lead to solutions that violate the fundamental mechanics principle that the work done by the external forces be equal to the increase in total energy of the system. Timoshenko's discussion of the problem of a vehicle driven across an elastic bridge is generalized to show that energy conservation can be restored only if the local deformation of the components is taken into account in determining the direction of the contact force. This result has important consequences for the interaction of elastic systems in general, including, for example, the dynamic behavior of meshing gears. [DOI: 10.1115/1.1574060]

1 Introduction

If a mechanical system contains no energy sources or dissipative mechanisms (such as friction or plasticity), the work done by the external forces must be equal to the increase in total potential energy of the system. This principle is one of the pillars of mechanics, but apparent counter examples can be produced if the system contains components whose stiffness changes with time. This inconsistency is a clear indication that the problem is in some sense ill-posed. In the present paper, we shall demonstrate that the energy conservation principle will be satisfied in such cases only if the local deformation of the components is taken into account in determining the direction of the contact force.

An important application in which the stiffness of a mechanical component varies with time concerns the meshing of two gears. In this case, the meshing stiffness changes as the contact point moves over the gear teeth and at the point where an additional pair of teeth comes into contact or leaves contact. Other examples include a vehicle moving over an elastic bridge or a loaded system in which the elastic modulus of the material changes as a function of temperature.

To introduce the subject, consider the simple case of a linear spring of stiffness k loaded by a force F. The extension of the spring, u, and the strain energy stored, U, are given by

$$u=\frac{F}{k};\quad U=\frac{F^2}{2k}, \tag{1}$$

respectively. If we now slowly change the stiffness of the spring by an amount δk (for example, by changing the temperature and hence the elastic modulus of the material), the force F will do work

$$\delta W=F\,\delta u=F\,\frac{\partial u}{\partial k}\,\delta k=-\frac{F^2\delta k}{k^2}, \tag{2}$$

but the strain energy will increase by

$$\delta U=\frac{\partial U}{\partial k}\,\delta k=-\frac{F^2\delta k}{2k^2}, \tag{3}$$

so the system appears to violate the principle of conservation of energy under a change of stiffness.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, July 16, 2001; final revision, Dec. 16, 2002. Associate Editor: L. T. Wheeler. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

In this example, the inconsistency will be resolved if the problem is reformulated in the context of thermodynamics and the apparent energy deficit will be associated with an exchange between thermal and mechanical energy. However, similar problems arise in purely mechanical problems, where the stiffness change is due to kinematic effects. These effects are generally not explicitly remarked in the literature. For example, the change in meshing stiffness of involute gears is sometimes approximated by representing the meshing stiffness by a sinusoidal function, [1]. The resulting equation of motion then takes the form of the Mathieu equation, which has domains of instability in which an initial perturbation from the steady periodic state will grow exponentially with time. Clearly this implies that the total energy increases with time. However, the mean power at input and output are equal and opposite, so the system as modeled violates the principle of conservation of energy.

2 A Cantilever Beam Problem

A simple example with a kinematically varying stiffness involves the cantilever beam of Fig. 1(*a*), loaded by a transverse force F at a distance x from the support. Elementary calculations show that the displacement under the force and the strain energy are

$$u=\frac{Fx^3}{3EI};\quad U=\frac{F^2x^3}{6EI}, \tag{4}$$

respectively, where EI is the flexural rigidity of the beam. If the point of application of the force now moves from x to $x+\delta x$, the displacement will change by

$$\delta u=\frac{\partial u}{\partial x}\,\delta x=\frac{Fx^2\delta x}{EI}, \tag{5}$$

allowing the forces to do work

$$\delta W=\frac{F^2x^2\delta x}{EI}. \tag{6}$$

However, the corresponding change in strain energy in the beam is only

$$\delta U=\frac{\partial U}{\partial x}\,\delta x=\frac{F^2x^2\delta x}{2EI}. \tag{7}$$

Alternatively, these results can be obtained from Eqs. (1)–(3) by substituting $k=3EI/x^3$.

This paradox is related to that remarked by Timoshenko and others, [2–6], in connection with the vibration of beams subject to moving transverse loads. If a vehicle drives across a bridge supported at both ends, the gravitational force does no net work, since the vehicle leaves at the same vertical level as it enters, but in

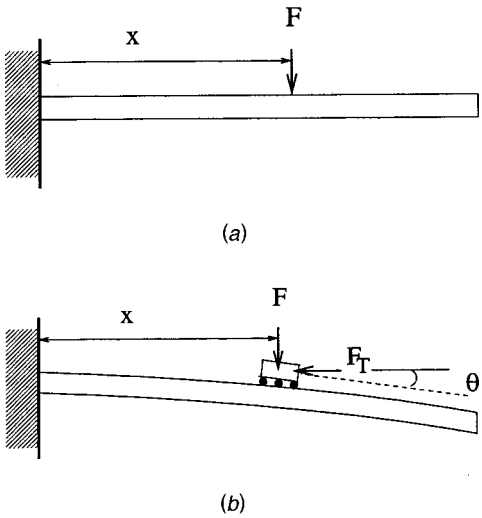


Fig. 1 (*a*) The cantilever beam loaded by a normal force, (*b*) the same beam loaded through a frictionless roller

general the bridge will be left in a state of vibration and hence will be in a higher energy state than it was before the transit. Where does the extra energy come from?

Timoshenko resolved the problem by noting that the instantaneous motion of the vehicle is not horizontal because of the deflected shape of the bridge. It follows that the brakes or the engine must be engaged to ensure a constant transit speed and this introduces additional energy terms. This argument can be applied to our cantilever beam problem by introducing the modified system of Fig. 1(*b*), in which the force is transmitted to the beam through a roller. If the roller is frictionless, it can be retained in equilibrium only by the application of a tangential force

$$F_T = F \tan\theta, \tag{8}$$

where

$$\tan\theta = \frac{Fx^2}{2EI} \tag{9}$$

is the slope of the beam at the point of application of the force. If the roller in Fig. 1(*b*) moves a distance δx to the right, an amount of work

$$\delta W_T = F_T \delta x = \frac{F^2 x^2 \delta x}{2EI} \tag{10}$$

will be done against the force F_T and the inclusion of this term completes the energy balance

$$\delta U = \delta W - \delta W_T. \tag{11}$$

Lee [3] showed that the same conclusion could be achieved without recourse to arguments from contact mechanics. We adapt the notion of Lee (who uses a convected time derivative) to the quasi-static case under consideration here. We decompose the motion of the force into two processes. In the first phase, the beam is ''frozen'' in its deformed state while the force moves from x to $x+\delta x$. During this phase, the force moves a distance $\delta u_1 = \delta x \tan\theta$ and hence does work,

$$\delta W_1 = F\,\delta u_1 = F\,\delta x \tan\theta = \frac{F^2 x^2 \delta x}{2EI}, \tag{12}$$

but none of this work is communicated to the beam. In the second phase, the beam is allowed to relax to its new equilibrium position. The additional displacement of the force is $\delta u_2 = \delta u - \delta u_1$ and the work done during this phase

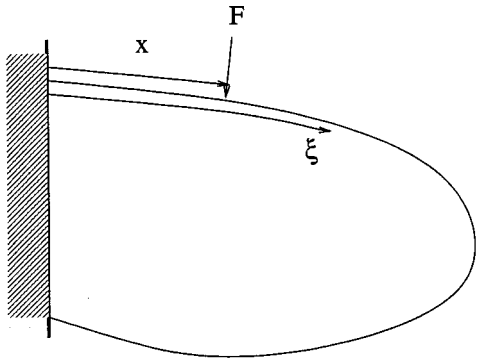


Fig. 2 General elastic structure loaded by a normal contact force

$$\delta W_2 = F\,\delta u_2 = \frac{F^2 x^2 \delta x}{2EI} \tag{13}$$

is communicated to the beam and results in the increase of strain energy δU.

Notice that if the direction of motion is reversed, the quantity δW_1 will be negative, showing that an external source of energy is required in phase 1 to move the force over the ''frozen'' beam.

Of course, in a real physical application, the work δW_1 cannot simply be lost to or generated from a fictitious energy source and any practical realization of the problem will bring us back into the realm of contact mechanics.

3 A More General Case

Figure 2 shows a more general elastic structure loaded by a normal force F at a point on the boundary characterized by a curvilinear coordinate x. We assume that strains and rotations are small, but the elastic behavior is not necessarily linear. We also assume that the force produces a bounded displacement at its point of application. This restriction will be removed for the linear case in Section 4.

We define the normal displacement at a general point on the boundary $x=\xi$ as

$$u = u(F,x,\xi), \tag{14}$$

in which case the local rotation of the deformed surface is

$$\theta(\xi) = \frac{\partial u}{\partial \xi}(F,x,\xi). \tag{15}$$

We also define the functions

$$f(F,x) \equiv u(F,x,x); \quad g(F,x) = \theta(F,x,x), \tag{16}$$

which are the normal displacement and rotation at the point of application of the force.

The strain energy in the structure can be found by applying F gradually, keeping its location fixed. It is therefore given by

$$U(F,x) = \int_0^F \frac{\partial u(F,x,x)}{\partial F} F\,dF = \int_0^F \left(\frac{\partial f}{\partial F}\right) F\,dF. \tag{17}$$

If, following Timoshenko's scenario, the force is applied through a frictionless roller, we shall require a restraining force

$$F_R = F\,\theta(F,x,x) = F g(F,x). \tag{18}$$

If the roller is now allowed to move a distance δx, the force F will do an increment of work

$$\delta W = F \frac{\partial f}{\partial x}\,\delta x, \tag{19}$$

but F_R will have work done against it equal to

$$\delta W_R = F_R \delta x = F g(F,x)\,\delta x. \tag{20}$$

Thus, the net work done on the structure will be

$$\delta W - \delta W_R = \left(F \frac{\partial f}{\partial x} - Fg \right) \delta x. \tag{21}$$

Equating this to the increase in strain energy in the structure, we obtain

$$\frac{\partial U}{\partial x} = F \frac{\partial f}{\partial x} - Fg = \int_0^F \left(\frac{\partial^2 f}{\partial F \partial x} \right) F dF, \tag{22}$$

from Eq. (4). Differentiating with respect to F, we then obtain

$$F \frac{\partial^2 f}{\partial F \partial x} = F \frac{\partial^2 f}{\partial F \partial x} + \frac{\partial f}{\partial x} - g - F \frac{\partial g}{\partial F} \tag{23}$$

and hence

$$\frac{\partial f}{\partial x} - g - F \frac{\partial g}{\partial F} = 0, \tag{24}$$

which defines a relationship between the displacement and the slope of the structure which must be satisfied if energy is to be conserved during the process.

Alternative Proof. An alternative proof of this result can be obtained by invoking the incremental form of Maxwell's reciprocal theorem, [7], for small perturbations about the reference state where the force F is applied at x, producing displacement u and rotation θ.

Moving F by a distance δx is equivalent to adding an infinitesimal moment $\delta M = F\,\delta x$ at x. The response of the structure can be linearized for small perturbations about the reference state, leading to the relation

$$\frac{\partial u}{\partial M} = \frac{\partial \theta}{\partial F} = \frac{\partial g}{\partial F}. \tag{25}$$

Following Lee's argument above, we consider the displacement due to the motion of the force from x to $x + \delta x$ as the sum of two parts. Freezing the beam in its deformed shape, we have $\delta u_1 = \theta(F,x,x)\,\delta x$ after which relaxation to the new equilibrium position gives an additional displacement δu_2 associated with the moment $F\,\delta x$. The total displacement of the force along its line of action is therefore

$$\theta(F,x,x)\,\delta x + \frac{\partial u}{\partial M} F\,\delta x = g(F,x)\,\delta x + \frac{\partial g}{\partial F} F\,\delta x, \tag{26}$$

but this displacement is also given by

$$\frac{\partial f}{\partial x}\,\delta x,$$

giving

$$\frac{\partial f}{\partial x} = g + F \frac{\partial g}{\partial F} \tag{27}$$

as before.

4 Distributed Forces

In problems of linear elasticity, the displacement field due to a concentrated force is singular at the point of application of the force and hence the functions f, g of the previous section are not well defined. However, the concentrated force solution can still be used as a Green's function to define the effect of a distributed force by superposition. Consider the case where Fig. 2 represents a two-dimensional linear elastic body and suppose that a concentrated normal force F at x produces a normal displacement u at ξ, where

$$u(\xi) = F \tilde{u}(x,\xi). \tag{28}$$

Now consider the case where the force is distributed in the vicinity of x, with intensity

$$p(x+r) = F f(r), \tag{29}$$

where the distribution function $f(r)$ is nonzero only in $-c < r < c$ and is normalized so that

$$\int_{-c}^{c} f(r)\,dr = 1. \tag{30}$$

The displacement due to this distribution can now be written by superposition as

$$u(\xi) = F \int_{-c}^{c} \tilde{u}(x+r,\xi) f(r)\,dr \tag{31}$$

and the local slope is

$$\theta(\xi) = \frac{\partial u}{\partial \xi} = F \int_{-c}^{c} \frac{\partial \tilde{u}}{\partial \xi}(x+r,\xi) f(r)\,dr. \tag{32}$$

The stored strain energy is

$$U = \frac{1}{2} \int_{x-c}^{x+c} u(\xi) p(\xi)\,d\xi = \frac{F^2}{2} \int_{-c}^{c} \int_{-c}^{c} \tilde{u}(x+r,x+s) f(r) f(s)\,dr\,ds, \tag{33}$$

writing $\xi = x + s$. If the distributed force is now displaced a distance δx, the increment in strain energy will be

$$\delta U = \frac{\partial U}{\partial x}\,\delta x = \frac{F^2 \delta x}{2} \int_{-c}^{c} \int_{-c}^{c} \left(\frac{\partial \tilde{u}}{\partial x} + \frac{\partial \tilde{u}}{\partial \xi} \right)(x+r,x+s) f(r) f(s)\,dr\,ds. \tag{34}$$

The additional work done by the normally directed distributed force during this motion is

$$\delta W = \delta x \int_{x-c}^{x+c} \left(\theta(\xi) + \frac{\partial u}{\partial x}(\xi) \right) p(\xi)\,d\xi = F^2 \delta x \int_{-c}^{c} \int_{-c}^{c} \left(\frac{\partial \tilde{u}}{\partial x} + \frac{\partial \tilde{u}}{\partial \xi} \right)(x+r,x+s) f(r) f(s)\,dr\,ds, \tag{35}$$

which is exactly twice the increment of strain energy. The remaining work is required to overcome the implied tangential restraining force F_T, which is the resultant of tractions equal and opposite to the component of p parallel to the *deformed* surface and is given by

$$F_T = \int_{x-c}^{x+c} \theta(\xi) p(\xi)\,d\xi = F^2 \int_{-c}^{c} \int_{-c}^{c} \frac{\partial \tilde{u}}{\partial \xi}(x+r,x+s) f(r) f(s)\,dr\,ds. \tag{36}$$

Notice that from Maxwell's reciprocal theorem we have

$$\tilde{u}(x,\xi) = \tilde{u}(\xi,x); \quad \frac{\partial \tilde{u}}{\partial \xi}(x+r,x+s) = \frac{\partial \tilde{u}}{\partial x}(x+s,x+r) \tag{37}$$

and hence

$$\int_{-c}^{c} \int_{-c}^{c} \frac{\partial \tilde{u}}{\partial \xi}(x+r,x+s) f(r) f(s)\,dr\,ds = \int_{-c}^{c} \int_{-c}^{c} \frac{\partial \tilde{u}}{\partial x}(x+r,x+s) f(r) f(s)\,dr\,ds, \tag{38}$$

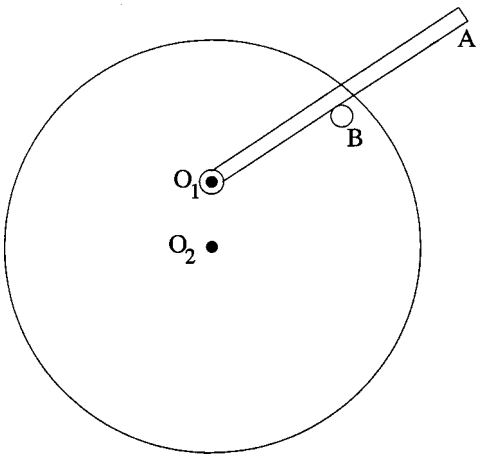


Fig. 3 Simple gear system using the cantilever beam of Fig. 1

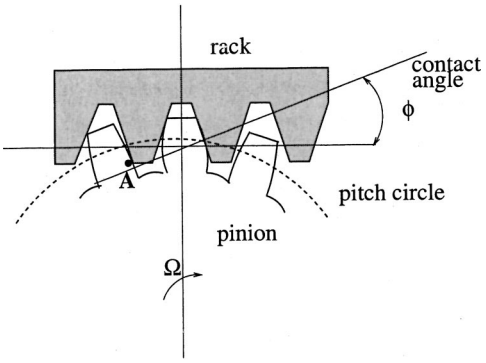


Fig. 4 A flexible involute gear meshing with a rigid rack

on interchanging the dummy variables s, r. Using this result in Eqs. (34), (35), we see that the system satisfies the principle of conservation of energy[1] if and only if we include the contribution of the tangential force F_T.

5 Implications in Contact Mechanics

These results show that in systems with kinematically varying stiffness, the direction of the contact forces or tractions must be chosen to be normal to the *deformed* contact surface in a Newtonian statement of the problem if energy conservation is to be preserved. This contradicts the conventional wisdom in contact mechanics, where the direction of the contact forces is referred to the undeformed configuration of the contacting bodies.

For dynamic systems involving moving contacts such as the meshing of two gears, failure to include this effect will generally lead to equations of motion that are incorrect because their solution violates the fundamental principle that the work done by the external forces equal the change in total potential energy of the system.

Figure 3 shows a simple illustrative example in which a ''gear'' comprising a flexible beam O_1A rotates clockwise at constant speed Ω about a center O_1 that is fixed in space. The rod drives a rigid pin B mounted on a rigid disk which rotates about center O_2. If O_1, O_2 are not coincident, the effective length O_1B of the beam will vary with angular position, leading to a kinematically varying stiffness.

The beam support can be brought to rest by superposing a counterclockwise rigid-body rotation Ω on the whole system. The disk will then be seen to execute a more complex motion whose effect however is merely to cause the pin B to slide quasi-sinusoidally along the beam. The results of Section 2 therefore show that the correct (energy conserving) equations of motion for this system will be obtained only if the local slope of the deflected beam is taken into account in determining the direction of the transmitted force and hence the torque transmitted by the disk O_2B.

5.1 Involute Gears. Similar considerations apply to the more complex system of the meshing of two involute gears. Once again, there will generally be a variation of effective stiffness as the contact point moves along the surfaces of the two meshing teeth. However, most gear systems will have a noninteger contact ratio, implying that the number of teeth in contact changes during the meshing cycle, resulting in a major change in contact stiffness. The present energy arguments show that there must be a sudden input of energy to the system at the point where the number of teeth in contact (and hence the total contact stiffness) *increases* and a sudden removal of energy when it *decreases*. The only external forces acting on the system that do work are the input and output torques, so we must conclude that there will be a significant change in the instantaneous torque ratio of the gears during these transitions, associated with the change in the line of action of the contact forces due to gear tooth deformation. The reader is invited to conduct a 'thought experiment' in which the output (reaction) torque is held constant and the input shaft is rotated at extremely slow speed. A significant spike will be needed in the input torque to force an extra tooth into contact. By contrast, when a tooth leaves contact, the input shaft will tend to spring ahead.

To explain this result, notice first that the tooth deflection causes a relative rigid-body motion of the gears, so at the point where an additional tooth would theoretically come into contact as a result of involute action, the new unloaded tooth would actually be in a position implying interpenetration. This is illustrated in Fig. 4 for a deformable gear meshing with a rigid rack. As a result, the contact of the new tooth will actually start before the theoretically correct point and it will involve contact of the noninvolute corner of the rack (Point A in Fig. 4) with the flank of the tooth. The line of action of the transmitted force will deviate considerably from the theoretical pressure line during this engagement period as a consequence of local tooth deformation. The analysis of this problem would be geometrically complex, but the results of Sections 3 and 4 show that the use of the true direction of the contact force would lead to the same result as the simpler energy-based analysis.

The tooth engagement period represents only a small proportion of the tooth period and an acceptable idealization in many cases is to assume it is instantaneous, leading to discontinuities (jumps) in stiffness. However, this implies the occurrence of discontinuities in strain energy and energy conservation requires corresponding discontinuities in kinetic energy and hence in rotational speed. The implications of these discontinuities for the dynamics of involute gear sets will be discussed in a separate paper, [8].

6 Conclusions

In this paper we have shown that a conventional statement of the equations of motion for a system with kinematically varying stiffness will generally lead to a solution that violates the fundamental mechanics principle that the work done by the external forces be equal to the change in total potential energy of the system. We cannot emphasize too strongly that such equations of motion are therefore incorrect.

To obtain a correct statement of the governing differential equations, it is necessary to allow for the local deformation of the components in determining the direction of the contact forces or tractions. This result, which is a generalization of the ''Timosh-

[1]Notice that for an elastic body, the tangential force F_T will also induce a local tangential deformation, but the associated work term is of second order relative to those considered above.

enko paradox,'' applies even in small strain problems where the problem statement is conventionally referred to the undeformed configuration.

References

[1] Benton, M., and Seireg, A., 1981, ''Factors Influencing Instability and Resonances in Geared Systems,'' Trans. ASME J. Mech. Des., **103**, pp. 372–378.

[2] Timoshenko, S. P., Young, D. H., and Weaver, W., 1974, *Vibration Problems in Engineering*, 4th Ed., John Wiley and Sons, New York, pp. 448–453.

[3] Lee, E. H., 1952, ''On a ''Paradox'' in Beam Vibration Theory,'' Q. J. Appl. Math., **10**, pp. 290–292.

[4] Maunder, L., 1960, ''On the Work of a Force Crossing a Beam,'' Q. J. Appl. Math., **17**, pp. 437–439.

[5] Adams, G. G., 1995, ''Critical Speeds and the Response of a Tensioned Beam on an Elastic Foundation to Repetitive Moving Loads,'' Int. J. Mech. Sci., **37**, pp. 773–781.

[6] Gavrilov, S., 2002, ''Nonlinear Investigation of the Possibility to Exceed the Critical Speed by a Load on a String,'' Acta Mech., **154**, pp. 47–60.

[7] Marsden, J. E., and Hughes, T. J. R., 1994, *Mathematical Foundations of Elasticity*, Dover, New York, pp. 290–292.

[8] Oh, Sejoong, Grosh, Karl, and Barber, J. R., 2003, ''Energy Conserving Equations for Gear Motion,'' ASME J. Vibr. Acoust., under review.

On the Strain Saturation Conditions for Polycrystalline Ferroelastic Materials

C. M. Landis
Mem. ASME
e-mail: landis@rice.edu
Department of Mechanical Engineering and
Materials Science, MS 321,
Rice University, P.O. Box 1892,
Houston, TX 77251

A phenomenological constitutive law is developed for the deformation of polycrystalline ferroelastic materials. The model is framed within a thermodynamic setting common to internal variable plasticity. The two significant inputs to this model are a switching (yield) surface, and a hardening potential. To maintain simplicity, the shape of the switching surface is assumed to be spherical in a modified deviatoric stress space. In order to ascertain the functional form of the hardening potential, micromechanical self-consistent simulations of multiple single crystals, with tetragonal crystal structure, embedded in an effective polycrystalline matrix, are performed for differing loading paths in remanent (plastic) strain space. As a result of the asymmetry in the tension versus compression behavior of these materials, it is shown that pure shear loading does not result in pure shear straining. This feature of the material behavior is demonstrated with the self-consistent simulations and predicted by the phenomenological constitutive law. Ultimately, the phenomenological theory is able to capture the complex constitutive behavior of ferroelastic materials predicted by the micromechanical model.
[DOI: 10.1115/1.1600472]

1 Introduction

Many smart materials, including ferroelectrics and shape memory alloys (below the martensite finish temperature, M_f), have a noncubic crystal structure. The simplest of these structures is tetragonal, but other structures such as rhombohedral and orthorhombic exist in technologically useful ferroelectrics with chemical compositions near a morphotropic phase boundary. A significant feature of these materials is that they exhibit irreversible deformation through a switching mechanism, e.g., a tetragonal variant oriented in the x direction can switch its orientation to the y direction. In shape memory alloys below M_f this switching is termed twinning and detwinning or twin reorientation. In all cases this type of switching, when it is induced by applied mechanical stress, is called ferroelasticity.

An interesting feature of ferroelastic deformation is that there exists an asymmetry in the uniaxial tension and uniaxial compression behavior [1–3]. In general, larger irreversible strains can be attained in tension than in compression (for the common case when the c-axis of the unit cell is longer and the a-axes). Consider a single crystal of tetragonal material. Assume that this crystal consists of equal quantities of three tetragonal variants as illustrated in Fig. 1(*a*), i.e. one third of the crystal has the c axis aligned in the x direction, one third in the y, and one third in the z. Within the crystal the variant types are divided into domains, with each domain separated by a domain wall or twin boundary (the terminology used depends on the material in consideration). If we assume that the strain state of a variant with its c axis aligned in the x, y, and z directions are given as $\varepsilon_{xx}=\varepsilon_0$, $-\varepsilon_0/2$,$-\varepsilon_0/2$, $\varepsilon_{yy}=-\varepsilon_0/2$,$\varepsilon_0$,$-\varepsilon_0/2$ and $\varepsilon_{zz}=-\varepsilon_0/2$,$-\varepsilon_0/2$,$\varepsilon_0$ respectively, then the initial volume averaged strain of the entire crystal vanishes. Now, apply a tensile stress in the x direction. This stress will do positive work if the variants aligned in the y and z directions switch their orientations towards the x direction. Generally, the mechanism for switching is the motion of domain walls/twin boundaries. Switching will proceed until all of the variants are aligned in the x direction, leaving the crystal with a strain of $\varepsilon_{xx}=\varepsilon_0$, $\varepsilon_{yy}=-\varepsilon_0/2$ and $\varepsilon_{zz}=-\varepsilon_0/2$. Now, if we apply a compressive stress in the x direction, the stress will be able to do positive work if the variants oriented in the x direction switch to either the y or z directions. Furthermore, the amount of work done by either of these switches is identical. In general, the exact switching sequence will depend on the geometry of the domain walls; however, if the crystal is large enough it is reasonable to expect that the final state of the crystal will have half of its variants aligned in the y direction and half in the z. Therefore, in compression the averaged strain state of the crystal is $\varepsilon_{xx}=-\varepsilon_0/2$, $\varepsilon_{yy}=\varepsilon_0/4$, and $\varepsilon_{zz}=\varepsilon_0/4$. Hence, for the single crystal loaded along any of its $\langle 100\rangle$ directions the maximum irreversible tensile strain is ε_0 and the maximum compressive strain is $-\varepsilon_0/2$, as illustrated in Fig. 1(*b*).

This observation has led some researchers to propose a saturation condition for ferroelastic *polycrystals* based on a minimum principal remanent strain criterion [4–6]; namely, that the minimum principal remanent strain in the material can never be less than $-\varepsilon_0/2$. This criterion leads to a tension-compression asymmetry ratio of 2:1 for the polycrystal loaded in any direction. More sophisticated models that treat the polycrystal as an aggregate of single crystals have found this asymmetry ratio to be about 1.37:1 [7–9]. Such a model will be used in this work to investigate the entire range of strain saturation conditions. For example, what is the maximum value of pure shear remanent strain that can be achieved? After the presentation of the micromechanical simulations a phenomenological model for ferroelasticity will be developed and compared to predictions from the micromechanical model.

2 Micromechanical Computations

The micromechanical model for the polycrystalline behavior implemented here is identical to the model presented in Ref. [10] without the effects of electric field. Furthermore, this model is analogous to the Hill-Hutchinson model [11–15] for polycrystal plasticity. The fundamental components to this model are a tangent constitutive law for the single crystals and a self-consistent averaging method to predict the properties of the polycrystal.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, July 29, 2002; final revision, Dec. 10, 2002. Associate Editor: H. Gao. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

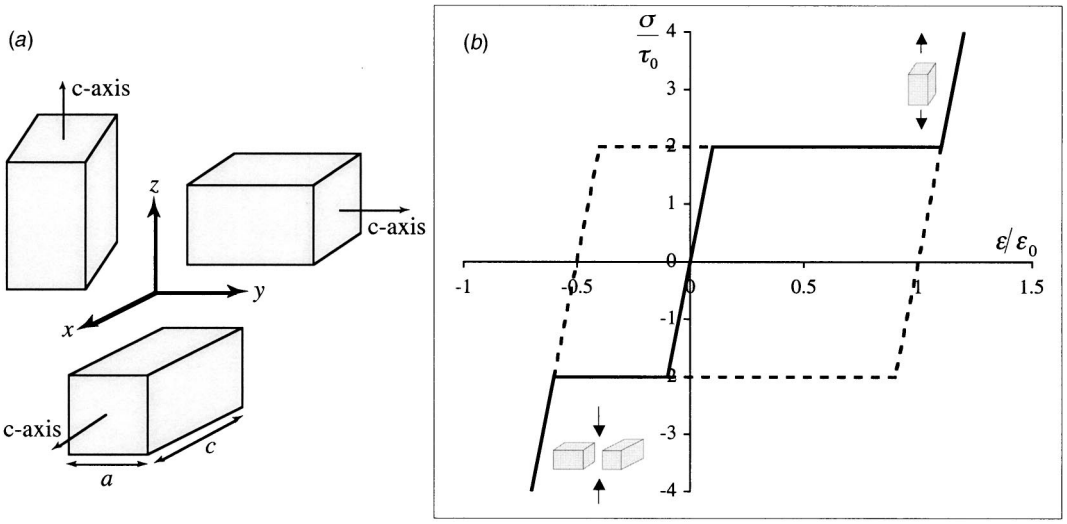


Fig. 1 (***a***) **The three possible orientations of tetragonal variants within a single crystal. Different variants within a single crystal will be separated by a domain wall or twin boundary.** (***b***) **The uniaxial stress-strain response of a model single crystal loaded along any of the ⟨100⟩ directions. Notice the asymmetry in tension versus compression.**

Each single crystal has a unique orientation and is treated as a spherical inclusion embedded in an infinite matrix. The tangent properties of the matrix are taken to be a self-consistent average of the incremental behavior of all of the single crystals. Hence, a stress or strain history can be applied to the polycrystal and the model can be used to determine the corresponding strain or stress history.

The single crystal constitutive law is analogous to continuum slip plasticity models with the added effect of strain saturation. In order to derive the single crystal constitutive law it is assumed that the stress within the single crystal is uniform and both the total strain and the remanent strain are the volume averages over the entire crystal. Note that by assuming the stress within the single crystal is uniform, we are neglecting elastic interactions between domains in the crystal. As illustrated in Fig. 1(*a*), three tetragonal variants can exist in each single crystal. Hence, the strain, remanent strain, and elastic compliance of the single crystal are

$$\varepsilon_{ij}^{sc}=\sum_{\mathrm{I}=1}^{3} c^{\mathrm{I}}\varepsilon_{ij}^{(\mathrm{I})}, \quad \varepsilon_{ij}^{r,sc}=\sum_{\mathrm{I}=1}^{3} c^{\mathrm{I}}\varepsilon_{ij}^{r(\mathrm{I})}, \quad \text{and} \quad s_{ijkl}^{sc}=\sum_{\mathrm{I}=1}^{3} c^{\mathrm{I}}s_{ijkl}^{(\mathrm{I})}, \tag{2.1}$$

where c^{I} represents the volume concentration of the Ith variant. Then, the remanent strains of each of the variants are given by

$$\varepsilon_{ij}^{r(1)}=\varepsilon_0(3\delta_{i1}\delta_{j1}-\delta_{ij})/2, \varepsilon_{ij}^{r(2)}=\varepsilon_0(3\delta_{i2}\delta_{j2}-\delta_{ij})/2,$$

and

$$\varepsilon_{ij}^{r(3)}=\varepsilon_0(3\delta_{i3}\delta_{j3}-\delta_{ij})/2, \tag{2.2}$$

where the 1, 2, and 3 coordinate directions are parallel to the ⟨100⟩ crystal directions, δ_{ij} is the Kronecker delta, and ε_0 has the same meaning as that for the discussion in the Introduction, and can be given in terms of the lattice parameters as $\varepsilon_0=2(c-a)/(c+2a)$. With these definitions of the remanent strains for the individual variants, the remanent strain of the entire crystal will be zero if there are equal volume concentrations of each of the variants.

In order to determine how the volume concentrations of the variants can change with applied loading, we will first consider the free energy of the crystal. Under isothermal conditions, the Helmholtz free energy of the single crystal, Ψ^{sc}, is equal to the stored elastic energy of the crystal and is given as

$$\Psi^{sc}=\frac{1}{2}c_{ijkl}^{sc}(\varepsilon_{ij}^{sc}-\varepsilon_{ij}^{r,sc})(\varepsilon_{kl}^{sc}-\varepsilon_{kl}^{r,sc}) \quad \text{with}$$

$$c_{ijkl}^{sc}=(s_{ijkl}^{sc})^{-1} \quad \text{and} \quad \sigma_{ij}^{sc}=\frac{\partial\Psi^{sc}}{\partial\varepsilon_{ij}^{sc}}=c_{ijkl}^{sc}(\varepsilon_{kl}^{sc}-\varepsilon_{kl}^{r,sc}). \tag{2.3}$$

Note that the elastic compliance of the crystal, and hence the elastic stiffness, are allowed to change as the volume concentrations of the variants changes. The dissipation rate in the single crystal is given as the work rate due to applied stresses minus the free energy rate. By applying the previously stated assumptions and an appropriate Legendre transformation, the dissipation rate can be shown to be

$$\dot{w}^D=\sigma_{ij}^{sc}\dot{\varepsilon}_{ij}^{sc}-\dot{\Psi}^{sc}=\sigma_{ij}^{sc}\dot{\varepsilon}_{ij}^{r,sc}+\frac{1}{2}\dot{s}_{ijkl}^{sc}\sigma_{ij}^{sc}\sigma_{kl}^{sc}$$

$$=\sum_{\mathrm{I}=1}^{3}\dot{c}^{\mathrm{I}}\left(\sigma_{ij}^{sc}\varepsilon_{ij}^{r(\mathrm{I})}+\frac{1}{2}s_{ijkl}^{\mathrm{I}}\sigma_{ij}^{sc}\sigma_{kl}^{sc}\right). \tag{2.4}$$

Note that six different switching transformations are possible, i.e., each of the three variants can switch to the other two. Furthermore, if a transformation is occurring, then the volume concentrations of the variants being switched to and from increase and decrease, respectively, at exactly the same rates. By applying these facts, it is then possible to rewrite the dissipation rate and define the transformation driving forces, G^{α}, as

$$\dot{w}^D=\sum_{\alpha=1}^{6}\dot{f}^{\alpha}\left(\sigma_{ij}^{sc}\Delta\varepsilon_{ij}^{r(\alpha)}+\frac{1}{2}\Delta s_{ijkl}^{\alpha}\sigma_{ij}^{sc}\sigma_{kl}^{sc}\right)$$

$$=\sum_{\alpha=1}^{6}\dot{f}^{\alpha}G^{\alpha}, \quad \text{where} \quad G^{\alpha}\equiv\sigma_{ij}^{sc}\Delta\varepsilon_{ij}^{r(\alpha)}+\frac{1}{2}\Delta s_{ijkl}^{\alpha}\sigma_{ij}^{sc}\sigma_{kl}^{sc}. \tag{2.5}$$

Here, α numbers the six possible transformation systems and the Δ represents the difference in the following quantity between the variant being transformed to and that being transformed from. For example, take $\alpha=1$ to represent the transformation from variant 1 to variant 2. Then, $\Delta\varepsilon_{ij}^{r(\alpha=1)}=3\varepsilon_0(\delta_{i2}\delta_{j2}-\delta_{i1}\delta_{j1})/2$ (note that this is a pure shear strain), $\Delta s_{ijkl}^{\alpha=1}=s_{ijkl}^{(\mathrm{I}=2)}-s_{ijkl}^{(\mathrm{I}=1)}$ and the volume

concentration rates are $\dot{c}^{(\mathrm{I}=1)}=-\dot{f}^{(\alpha=1)}$ and $\dot{c}^{(\mathrm{I}=2)}=\dot{f}^{(\alpha=1)}$. Note that in general $s_{ijkl}^{(\mathrm{I}=1)}$ and $s_{ijkl}^{(\mathrm{I}=2)}$ will be different due to differences in the orientations of the c axis of each variant.

It is assumed that if a quantity of a given variant exists, i.e., $c^{\mathrm{I}}>0$, then it is possible to incrementally switch that variant into one of the other two variants. Furthermore, this switching between variants is assumed to occur only if such a transformation results in a characteristic rate of nonnegative dissipation. Specifically, a transformation system is potentially active if

$$G^{\alpha}=3\tau_0\varepsilon_0\rightarrow\dot{f}^{\alpha}\geqslant 0 \tag{2.6}$$

and the transformation system is inactive if

$$G^{\alpha}<3\tau_0\varepsilon_0\rightarrow\dot{f}^{\alpha}=0. \tag{2.7}$$

If we neglect changes in the elastic properties, then the physical interpretation of Eq. (2.6) is as follows; when the resolved shear stress on a transformation system reaches the critical resolved shear stress, τ_0, then transformation is allowed to occur on that system. Note that if the volume fraction of a given variant vanishes, then the transformation systems that reduce the quantity of that variant cannot be activated. This feature enables the model to account for strain saturation at the single crystal level. In the absence of hardening of the transformation systems, i.e., τ_0 remains fixed during switching, the single crystal constitutive law requires the following inputs: the elastic properties of a single tetragonal variant, the critical resolved shear stress to induce transformation τ_0, and the lattice parameters of a tetragonal variant that determine ε_0.

From Eqs. (2.1)–(2.7) it is a mathematical exercise to determine the transformation rates, $\dot{f}^{\alpha}$, in terms of the applied strain rates, and then forms for the single crystal tangent moduli follow. This procedure will not be given here, but it is presented clearly in Ref. [10]. The resulting uniaxial stress-strain response of the single crystal loaded in any one of the $\langle 100\rangle$ directions is given in Fig. 1(*b*). Note however, that the single crystal response is anisotropic, and loading along other directions will yield different behaviors. For example, uniaxial stress applied in any of the $\langle 111\rangle$ directions does not create a driving force on any of the transformation systems, and hence $\langle 111\rangle$ loading will result in a perfectly linear elastic response of the crystal.

Using the single crystal constitutive law described above, a self-consistent model is applied to compute the overall stress-strain behavior for a polycrystal. For conceptual simplicity the polycrystal is viewed as an infinite collection of randomly oriented single crystals subjected to homogenous states of stress and strain. The Cartesian components of the macroscopic stress and strain increments of the polycrystal are taken to be the volume averages of the Cartesian components of the corresponding stress and strain increments in the single crystals. Each individual single crystal region is modeled as a spherical inclusion embedded in an infinite *effective medium* matrix. The tangent stiffness of the effective medium is taken to be that of the polycrystal. Since neighboring grains are not modeled explicitly, the constraint interactions between grains are not determined directly in this model. Instead, each grain is constrained by the effective medium matrix, and in this sense the model accounts for grain-to-grain constraints in an averaged sense. Ultimately, the stress and strain state in each single grain will depend on the applied loading history and the orientation of the crystal. For more details of both the single crystal constitutive law and the self-consistent averaging method the reader is referred to Refs. [10–15].

Note that in Eqs. (2.1)–(2.5) the quantities σ_{ij}^{sc}, ε_{ij}^{sc}, and $\varepsilon_{ij}^{r,sc}$ were used to represent stress, strain, and remanent strain in a *single crystal*. Throughout the remainder of this paper, σ_{ij}, ε_{ij}, and ε_{ij}^{r} will be used to represent stress, strain, and remanent strain in a polycrystal. To isolate the behavior of the polycrystal saturation, the elastic properties of a single tetragonal variant were taken to be isotropic in this study, with shear modulus μ and Poisson's ratio ν. Then, a dimensional analysis of the governing equations implies that the predictions of the model for the normalized stresses σ_{ij}/τ_0 versus the normalized strains $\varepsilon_{ij}/\varepsilon_0$ will only depend on the dimensionless parameters $\mu\varepsilon_0/\tau_0$ and ν. The goal of this investigation is to map out the saturation conditions for remanent strain states between axisymmetric extension and axisymmetric contraction. To do this the following remanent strain invariants are introduced:

$$J_2^e=\left(\frac{2}{3}e_{ij}^r e_{ij}^r\right)^{1/2} \quad \text{and} \quad J_3^e=\left(\frac{4}{3}e_{ij}^r e_{jk}^r e_{ki}^r\right)^{1/3}. \tag{2.8}$$

Here e_{ij}^r is the remanent strain deviator, $e_{ij}^r=\varepsilon_{ij}^r-\delta_{ij}\varepsilon_{kk}^r/3$. Of course, since the deformation processes of the single crystals are volume conserving, recall that the transformation strains are pure shears and therefore $\varepsilon_{kk}^r=0$; we have $e_{ij}^r=\varepsilon_{ij}^r$ and the introduction of the remanent strain deviator appears to be unnecessary. However, the theory presented in the next section will require derivatives of these invariants, and the derivatives will be affected by this distinction. With the definition of these two invariants, a full range of remanent strain saturation states can be probed by allowing the ratio of J_3^e/J_2^e to vary from -1 (axisymmetric contraction) to 0 (pure shear) to 1 (axisymmetric extension). If we consider any volume conserving remanent strain in the principal directions, then the remanent strain tensor and J_3^e/J_2^e can be written as

$$\varepsilon^r=\begin{pmatrix}\varepsilon^r & 0 & 0\\ 0 & b\varepsilon^r & 0\\ 0 & 0 & -(1+b)\varepsilon^r\end{pmatrix} \quad \text{and}$$

$$\frac{J_3^e}{J_2^e}=\frac{-\sqrt{3}(b+b^2)^{1/3}}{4^{1/6}\sqrt{1+b+b^2}}\,\mathrm{sgn}(\varepsilon^r), \tag{2.9}$$

where b can be any arbitrary constant. In other words, every multiaxial volume conserving remanent strain state can be described by the ratio of these two invariants, and this ratio will always lie in the range $-1\leqslant J_3^e/J_2^e\leqslant 1$.

Next, consider the problem of finding the saturation conditions for a pure shear remanent strain. At first, one might attempt to find this condition by applying a pure shear stress or pure shear total strain to the model polycrystalline material. However, as will be shown, such a procedure will not produce a pure shear *remanent* strain state. Due to the material's ability to deform more in tension than in compression, the ratio of J_3^e/J_2^e will approach 1 as an applied pure shear stress is increased. Therefore, it is necessary to devise a more sophisticated scheme for probing the range of J_3^e/J_2^e. Within the self-consistent model, the tangent moduli of the polycrystal, c_{ijkl}^t, are computed at each increment in the loading process. This knowledge of the instantaneous tangent moduli allows for the adjustment of the loading path in applied stress space, such that the ratio J_3^e/J_2^e remains constant during the entire load excursion. More specifically, by manipulating the relationships

$$\dot{\sigma}_{ij}=c_{ijkl}(\dot{\varepsilon}_{kl}-\dot{\varepsilon}_{kl}^r)=c_{ijkl}^t\dot{\varepsilon}_{kl} \tag{2.10}$$

the appropriate stress or strain increments that are required to maintain a constant ratio of J_3^e/J_2^e during loading can be determined.

Figure 2 plots a set of results for the normalized effective stress, σ_e/τ_0, versus the normalized effective remanent strain, J_2^e/ε_0, for the dimensionless material ratios of $\mu\varepsilon_0/\tau_0=2.36$ and $\nu=0.22$. Note that the effective stress is defined as

$$\sigma_e=\left(\frac{3}{2}s_{ij}s_{ij}\right)^{1/2}, \quad \text{where } s_{ij}=\sigma_{ij}-\frac{1}{3}\sigma_{kk}\delta_{ij}. \tag{2.11}$$

Each plot on Fig. 2 represent a different loading path that was prescribed in such a way that the ratio of the remanent strain invariants, J_3^e/J_2^e, remained constant throughout the loading. The

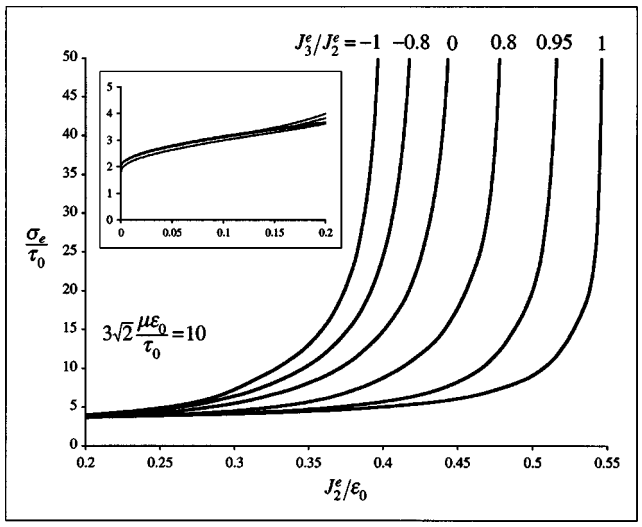


Fig. 2 Self-consistent computations of the deformation behavior of the ferroelastic material for different proportional remanent straining paths. The stresses are normalized by the critical resolved shear stress required to cause switching and the effective remanent strains are normalized by the tensile saturation strain of a single crystal. Note that $J_3^e/J_2^e=-1,0,1$ represents axisymmetric contraction, pure shear remanent strain, and axisymmetric extension, respectively. The inset is an expanded view of the region cut off from the larger plots. Note the lack of tension-compression asymmetry for small strains, but the significant asymmetry of the saturation strains.

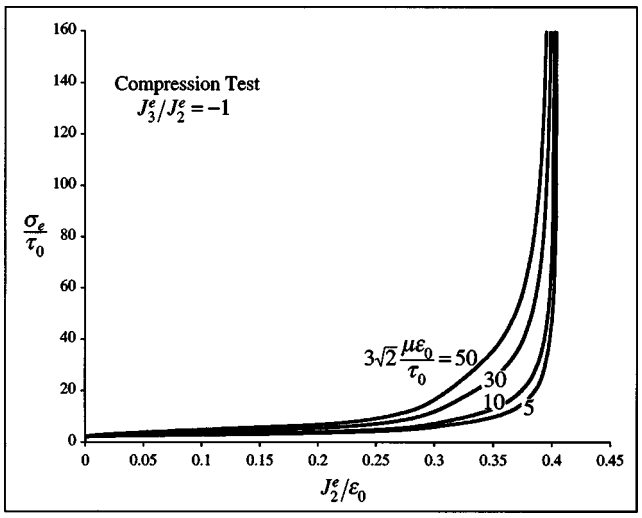


Fig. 3 Self-consistent computations of the uniaxial compression deformation behavior of the ferroelastic material for different levels of the dimensionless ratio $\mu\varepsilon_0/\tau_0$. Note that the saturation strain for each test is the same but the shape of the stress-strain curve differs.

inset contains an expanded view of the region excluded from the main graph. Note that a material with a switching (yield) surface described solely by the stress invariant σ_e and a hardening behavior dependent only on the remanent strain invariant J_2^e would have curves on this graph that are independent of the ratio J_3^e/J_2^e. At small remanent strains, $\varepsilon_{ij}^r<0.15\varepsilon_0$, the saturation behavior of the single crystals has not fully developed and has little effect on these stress-strain curves. In fact, the differences in the curves appearing on the inset figure arise primarily due to the fact that the switching surface for the model material follows a Tresca criterion [15]. As the remanent strain continues to increase, the saturation behavior of the single crystals begins to have an effect on the polycrystal, with the stress having to increase dramatically in order to cause further remanent straining. Ultimately, for each load path J_2^ε approaches some limiting value. The saturation values range from $0.40\varepsilon_0$ for axisymmetric contraction to $0.55\varepsilon_0$ for axisymmetric extension. These values are in agreement with other models with similar transformation system switching criteria, but less sophisticated polycrystalline averaging techniques [7–9].

In order to illustrate the effects of the dimensionless ratio $\mu\varepsilon_0/\tau_0$; Fig. 3 plots the normalized effective stress versus the normalized effective remanent strain in a uniaxial compression test for three values of $\mu\varepsilon_0/\tau_0$. Note that the curvature in the transition from J_2-like deformation to saturation is larger as $\mu\varepsilon_0/\tau_0$ decreases for these normalizations. Also note that the ultimate saturation level of remanent strain is the same for each level of $\mu\varepsilon_0/\tau_0$. The saturation strains are ultimately a function of the underlying crystal structure geometry; hence it is to be expected that the saturation levels of remanent strain are independent of the elastic properties of the material. In general, for any given untextured polycrystal, the saturation strain will depend only on the crystal structure of the variants, the switching criterion for the transformation systems, and the remanent strain invariant ratio J_3^e/J_2^e. Hence, the results for the saturation strain level to be presented in Fig. 4 are universally valid for tetragonal materials with the c axis longer than the a axis and transformation allowed at a critical resolved shear stress on the transformation system.

Lastly, Fig. 4 plots the saturation strain values for the model polycrystalline material versus the remanent strain ratio J_3^e/J_2^e. This result is the key component from these micromechanical simulations that will be applied to the phenomenological theory for ferroelastic switching presented in the next section. Figure 4 illustrates the entire range of multiaxial remanent strain states that are possible in the model material. Specifically, the region below the curve represents strain states that are achievable, and the region above the curve consists of unattainable remanent strains. Again, the observations that the low remanent strain region is relatively independent of J_3^e/J_2^e, and the saturation levels of

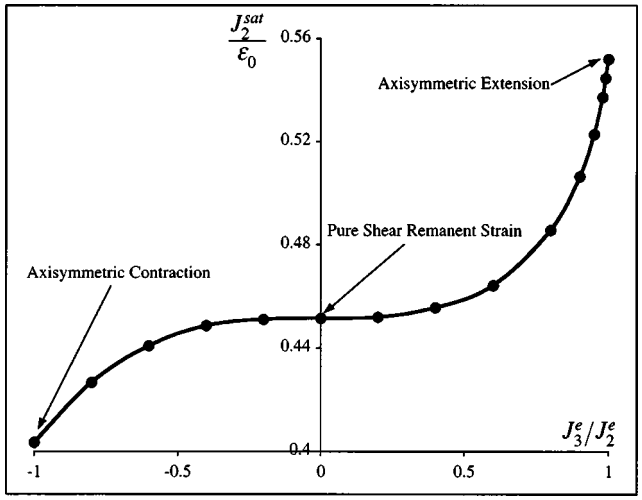


Fig. 4 The effective saturation strain level as a function of the remanent strain invariant ratio J_3^e/J_2^e. Note that $J_3^e/J_2^e=-1,0,1$ represents axisymmetric contraction, pure shear remanent strain, and axisymmetric extension respectively. This figure illustrates the anisotropic nature of the material in response to tension versus compression.

remanent strain are dependent on J_3^e/J_2^e will be used to devise a phenomenological constitutive law for these materials in the next section.

3 Phenomenological Model for a Polycrystalline Material

The phenomenological model follows the general formulation developed for ferroelectric materials developed in Ref. [6]. Here we will focus only on purely ferroelastic behavior. Hence this model applies to unpoled ferroelectric materials loaded only by applied stress and not by electric field. The model also applies to twin reorientation in shape memory alloys below the martensite finish temperature. The model will be cast within an isothermal, rate independent, small deformation framework. The primary assumption of the model is that the internal state of the material can be entirely characterized by the remanent, i.e., irreversible, strain state of the material ε_{ij}^r. Hence, we introduce the Helmholtz free energy per unit volume of the polycrystalline material, Ψ, as

$$\Psi = \Psi^s(\varepsilon_{ij}, \varepsilon_{ij}^r) + \Psi^r(\varepsilon_{ij}^r). \tag{3.1}$$

Here, Ψ^s represents the stored elastic energy per unit volume, ε_{ij} are the components of the total strain, and Ψ^r represents a contribution to the free energy associated only with the internal state of the material. The following development of the phenomenological theory will demonstrate that Ψ^r gives rise to ''back stresses.'' It is generally accepted [16–18] that the physical mechanism for back stresses is the existence of residual stresses in the material due to inhomogeneous remanent/plastic strain from grain to grain, i.e., ''locked-in'' energy [18]. Therefore, the physical interpretation of Ψ^r is that it accounts for the stored energy due to the internal residual stresses in the material.

Note that the elastic properties of the material can depend on the remanent strain state. For example, an initially unpoled ferroelectric will be elastically isotropic since the crystal variants with tetragonal elastic properties are randomly oriented. However, if this material is strained in tension in the x direction, most of the variants will align in the x direction and the elastic properties will now be transversely isotropic about the x direction. Then, assuming linear elastic behavior about a fixed remanent strain state, the stored free energy can be written as

$$\Psi^s = \frac{1}{2} c_{ijkl}(\varepsilon_{ij} - \varepsilon_{ij}^r)(\varepsilon_{kl} - \varepsilon_{kl}^r). \tag{3.2}$$

Here c_{ijkl} are the components of the elastic stiffness tensor that can depend on the remanent strain components ε_{ij}^r.

The second law of thermodynamics implies that the dissipation rate $\dot{\Delta}$ must be non-negative, i.e.,

$$\dot{\Delta} = \sigma_{ij}\dot{\varepsilon}_{ij} - \dot{\Psi} \geqslant 0. \tag{3.3}$$

A Legendre transformation along with the following definitions:

$$\overline{\sigma}_{ij} = \frac{1}{2} \frac{\partial s_{pqrs}}{\partial \varepsilon_{ij}^r} \sigma_{pq}\sigma_{rs}, \tag{3.4}$$

$$\sigma_{ij}^B = \frac{\partial \Psi^r}{\partial \varepsilon_{ij}^r}, \tag{3.5}$$

$$\hat{\sigma}_{ij} = \sigma_{ij} - \sigma_{ij}^B + \overline{\sigma}_{ij}, \tag{3.6}$$

are used to show that Eq. (3.3) can be written as

$$\dot{\Delta} = \hat{\sigma}_{ij}\dot{\varepsilon}_{ij}^r \geqslant 0. \tag{3.7}$$

Note that the s_{ijkl} are the components of the elastic compliance tensor, where the elastic compliance is the inverse of the elastic stiffness, i.e., $s_{ijkl} = (c_{ijkl})^{-1}$.

Now, assume that a switching surface $\Phi(\hat{\sigma}_{ij}, \varepsilon_{ij}^r) = 0$ exists in $\hat{\sigma}_{ij}$ space such that switching can occur if the state $\hat{\sigma}_{ij}$ is on the surface and the material response is elastic if the state $\hat{\sigma}_{ij}$ lies within the surface. Stress states outside of the switching surface are forbidden. If the material abides by the postulate of maximum plastic dissipation, such that $(\hat{\sigma}_{ij} - \hat{\sigma}_{ij}^*)\dot{\varepsilon}_{ij}^r \geqslant 0$ for any stress state $\hat{\sigma}_{ij}^*$ on or within the switching surface and the stress state $\hat{\sigma}_{ij}$ causes the remanent strain increment $\dot{\varepsilon}_{ij}^r$, then the switching surface must be convex and the flow law for the remanent strain rate is associative. In other words, convexity implies that the Hessian $\partial^2\Phi/\partial\hat{\sigma}_{ij}\partial\hat{\sigma}_{kl}$ is positive definite, and the associative flow rule implies that the remanent strain increment must be normal to the switching surface such that

$$\dot{\varepsilon}_{ij}^r = \lambda \frac{\partial \Phi}{\partial \hat{\sigma}_{ij}}. \tag{3.8}$$

Here, λ is a positive scalar multiplier that must be determined from the consistency condition. In addition to the convexity and normality constraints, the second law, Eq. (2.7), implies that the switching surface must enclose the origin in $\hat{\sigma}_{ij}$ space.

Along with the following definitions

$$U_{ijkl} = \frac{1}{2} \frac{\partial^2 s_{pqrs}}{\partial \varepsilon_{ij}^r \partial \varepsilon_{kl}^r} \sigma_{pq}\sigma_{rs}, \tag{3.9}$$

$$H_{ijkl} = \frac{\partial^2 \Psi^r}{\partial \varepsilon_{ij}^r \partial \varepsilon_{kl}^r}, \tag{3.10}$$

$$\tilde{\varepsilon}_{ij} = \frac{\partial \Phi}{\partial \hat{\sigma}_{ij}} + \frac{\partial \Phi}{\partial \hat{\sigma}_{kl}} \frac{\partial s_{mnij}}{\partial \varepsilon_{kl}^r} \sigma_{mn}, \tag{3.11}$$

$$\tilde{\sigma}_{ij} = c_{ijkl}\tilde{\varepsilon}_{kl}, \tag{3.12}$$

$$\mathcal{D} = \frac{\partial \Phi}{\partial \hat{\sigma}_{ij}}(H_{ijkl} - U_{ijkl}) \frac{\partial \Phi}{\partial \hat{\sigma}_{kl}} - \frac{\partial \Phi}{\partial \varepsilon_{ij}^r} \frac{\partial \Phi}{\partial \hat{\sigma}_{ij}}, \tag{3.13}$$

$$D = \mathcal{D} + \tilde{\sigma}_{ij}\tilde{\varepsilon}_{ij}, \tag{3.14}$$

and the consistency condition $\dot{\Phi} = 0$, we can solve for the multiplier λ as

$$\lambda = \frac{1}{\mathcal{D}} \tilde{\varepsilon}_{ij}\dot{\sigma}_{ij} = \frac{1}{D} \tilde{\sigma}_{ij}\dot{\varepsilon}_{ij} \tag{3.15}$$

and then write the incremental constitutive equations as

$$\dot{\varepsilon}_{ij} = \left(s_{ijkl} + \frac{1}{\mathcal{D}} \tilde{\varepsilon}_{ij}\tilde{\varepsilon}_{kl} \right) \dot{\sigma}_{kl} \tag{3.16}$$

or

$$\dot{\sigma}_{ij} = \left(c_{ijkl} - \frac{1}{D} \tilde{\sigma}_{ij}\tilde{\sigma}_{kl} \right) \dot{\varepsilon}_{kl}. \tag{3.17}$$

The definitions outlined in Eqs. (3.9)–(3.14) serve two purposes. First, these definitions allow for a compact notation, and second, the new variables allow for a simple realization of the symmetry of the tangent moduli.

4 Fitting the Model to the Material

The material specific variables that need to be specified for this model include the dependence of the elastic properties on the remanent strain, the switching surface Φ, and the remanent potential Ψ^r. For the sake of simplicity, we will assume that the elastic properties are isotropic and independent of the remanent strain from here on. As such, $\overline{\sigma}_{ij}$ and U_{ijkl} vanish. This assumption is in agreement with the self-consistent calculations since the individual crystallites were assumed to be elastically isotropic. However, real materials can exhibit significant anisotropy in their elastic properties and this assumption may need modification. Again, the emphasis of this work is to investigate the saturation behavior of these materials so this simplification is of secondary importance.

Due to the fact that the remanent strains arise due to the reorientation of the crystal variants it is reasonable to assume that the remanent strain is volume conserving, i.e., $\hat{\varepsilon}^r_{kk}=0$. Then the switching surface must be a function of invariants of the deviatoric stress, $\hat{s}_{ij}=\hat{\sigma}_{ij}-\delta_{ij}\hat{\sigma}_{kk}/3$. In general, the switching surface can change both size and shape with continued remanent straining. However, again for the sake of simplicity, we will assume the simple J_2 form for the switching surface, such that

$$\Phi=\frac{3}{2}\hat{s}_{ij}\hat{s}_{ij}-\sigma_0^2=0. \tag{4.1}$$

As noted in Sec. 2, the initial switching surface predicted by the self-consistent model follows the Tresca criterion. Furthermore, the switching surface will in general evolve into even more complex shapes than the Tresca hexagon [10,15]. However, it will be shown that this additional complexity need not be added to the phenomenological model in order to capture the salient features of the deformation behavior of ferroelastic materials, and that the switching surface form of Eq. (4.1) is adequate.

Finally, we must specify the form of the hardening potential Ψ^r. Of course, this potential must depend on the invariants of the remanent strain. Furthermore, to formally cause the back stress σ^B_{ij} to be purely deviatoric, we will write Ψ^r in terms of invariants of the remanent strain deviator, $e^r_{ij}=\varepsilon^r_{ij}-\delta_{ij}\varepsilon^r_{kk}/3$. Beyond this consideration, determining Ψ^r becomes a curve fitting exercise. However, this curve fitting must be done intelligently and should be informed by experimental observation and micromechanical models.

Recall that the remanent strain saturation behavior has a significant dependence on J^e_3. A first step to determining Ψ^r is to determine a reasonable fit to the saturation behavior. The saturation behavior depicted in Fig. 4 can be accurately determined in the following way. First, define a function f of the remanent strain invariant ratio J^e_3/J^e_2 as

$$f\left(\frac{J^e_3}{J^e_2}\right)=-0.0965\left(\frac{J^e_3}{J^e_2}\right)^3+0.01\left(\frac{J^e_3}{J^e_2}\right)^6+0.8935 \quad \text{for} \quad \left(\frac{J^e_3}{J^e_2}\right)<0, \tag{4.2}$$

$$f\left(\frac{J^e_3}{J^e_2}\right)=-0.1075\left(\frac{J^e_3}{J^e_2}\right)^3-0.027\left(\frac{J^e_3}{J^e_2}\right)^6-0.028\left(\frac{J^e_3}{J^e_2}\right)^{21}+0.8935 \quad \text{for} \quad \left(\frac{J^e_3}{J^e_2}\right)\geqslant 0. \tag{4.3}$$

Note that this fit for f has first, second, fourth and fifth derivatives equal to zero at $J^e_3/J^e_2=0$. The importance of these conditions on f at $J^e_3/J^e_2=0$ will be discussed shortly. Also note that this is a function for materials with tetragonal crystal structure only. Finally, one divided by this function is able to fit the normalized results of Fig. 4 (with J^e_2 normalized by ε_c instead of ε_0) to within 0.04% accuracy for $J^e_3/J^e_2<0$ and 0.15% for $J^e_3/J^e_2>0$.

Now, define a strainlike variable $\bar{\varepsilon}$ as

$$\bar{\varepsilon}=J^e_2 f(J^e_3/J^e_2). \tag{4.4}$$

When the remanent strain level characterized by $\bar{\varepsilon}$ reaches the compressive saturation level, ε_c, the remanent strain will be saturated, i.e., $\bar{\varepsilon}\leqslant\varepsilon_c$. Note that the micromechanical simulations described in Sec. 2 found that $\varepsilon_c=0.4035\varepsilon_0$. Then Eqs. (4.2) and (4.3) imply that the effective saturation level of remanent strain in tension is $\varepsilon_t=1.3679\varepsilon_c=0.5520\varepsilon_0$, and if the *remanent* strain state is pure shear, then the saturation level is $\varepsilon_s=1.1191\varepsilon_c=0.4515\varepsilon_0$. Hence, when devising a functional form for Ψ^r, the back stresses must become large as $\bar{\varepsilon}$ approaches ε_c.

The second significant observation from the self-consistent computations is that, discounting the Tresca nature of the initial switching surface, for small remanent strain levels the stress versus remanent strain behavior can be described solely by the strain invariant J^e_2, i.e., it is independent of J^e_3. This can be restated as: Ψ^r is only a function of J^e_2 as $\bar{\varepsilon}$ approaches zero. Furthermore, it is reasonable to conclude that Ψ^r is a function of $\bar{\varepsilon}$ as $\bar{\varepsilon}$ approaches ε_c. Hence, as part of this curve fitting procedure, a new strainlike variable is introduced,

$$\varepsilon^*=(1-w)J^e_2+w\bar{\varepsilon}, \tag{4.5}$$

where w is a weighting function that must be zero when $\bar{\varepsilon}=0$ and one when $\bar{\varepsilon}=\varepsilon_c$. A simple method for generating steplike weight functions is to take $w'(\bar{\varepsilon})=A\bar{\varepsilon}^p(1-\bar{\varepsilon})^q$. For this study $p=3$, $q=5$ and the functional form for w is

$$w(\bar{\varepsilon})=504\left[\frac{1}{4}\left(\frac{\bar{\varepsilon}}{\varepsilon_c}\right)^4-\left(\frac{\bar{\varepsilon}}{\varepsilon_c}\right)^5+\frac{5}{3}\left(\frac{\bar{\varepsilon}}{\varepsilon_c}\right)^6-\frac{10}{7}\left(\frac{\bar{\varepsilon}}{\varepsilon_c}\right)^7+\frac{5}{8}\left(\frac{\bar{\varepsilon}}{\varepsilon_c}\right)^8-\frac{1}{9}\left(\frac{\bar{\varepsilon}}{\varepsilon_c}\right)^9\right]. \tag{4.6}$$

Finally, we take Ψ^r to be a function of ε^*, i.e., $\Psi^r=\Psi^r(\varepsilon^*)$. The functional form for the derivative of Ψ^r with respect to ε^* can be fitted to either a uniaxial tension or compression test. For this study the functional form was taken as

$$\frac{d\Psi^r}{d\varepsilon^*}=H_0\left[\left(\frac{1}{1-\varepsilon^*/\varepsilon_c}\right)^m-1\right]. \tag{4.7}$$

Note that this function and hence the back stress components approach infinity as $\varepsilon^*\to\varepsilon_c$.

From Eq. (4.7) the required inputs to the phenomenological model can be obtained. Note that the ''primed'' variables denote first or second derivatives with respect to the argument of the function.

$$\sigma^B_{ij}=\frac{d\Psi^r}{d\varepsilon^*}\frac{\partial\varepsilon^*}{\partial\varepsilon^r_{ij}} \tag{4.8}$$

and

$$H_{ijkl}=\frac{d^2\Psi^r}{d\varepsilon^{*2}}\frac{\partial\varepsilon^*}{\partial\varepsilon^r_{ij}}\frac{\partial\varepsilon^*}{\partial\varepsilon^r_{kl}}+\frac{d\Psi^r}{d\varepsilon^*}\frac{\partial^2\varepsilon^*}{\partial\varepsilon^r_{ij}\partial\varepsilon^r_{kl}} \tag{4.9}$$

with

$$\frac{\partial\varepsilon^*}{\partial\varepsilon^r_{ij}}=(1-w)\frac{2e^r_{ij}}{3J^e_2}+(w+w'\bar{\varepsilon}-w'J^e_2)\frac{\partial\bar{\varepsilon}}{\partial\varepsilon^r_{ij}}, \tag{4.10}$$

$$\frac{\partial^2\varepsilon^*}{\partial\varepsilon^r_{ij}\partial\varepsilon^r_{kl}}=(1-w)\left[\frac{1}{3J^e_2}\left(\delta_{il}\delta_{jk}+\delta_{ik}\delta_{jl}-\frac{2}{3}\delta_{ij}\delta_{kl}\right)-\frac{4}{9(J^e_2)^3}e^r_{ij}e^r_{kl}\right]-\frac{2}{3}w'\left(\frac{e^r_{ij}}{J^e_2}\frac{\partial\bar{\varepsilon}}{\partial\varepsilon^r_{kl}}+\frac{e^r_{kl}}{J^e_2}\frac{\partial\bar{\varepsilon}}{\partial\varepsilon^r_{ij}}\right)+(w+w'\bar{\varepsilon}-w'J^e_2)\frac{\partial^2\bar{\varepsilon}}{\partial\varepsilon^r_{ij}\partial\varepsilon^r_{kl}}+(2w'+w''\bar{\varepsilon}-w''J^e_2)\frac{\partial\bar{\varepsilon}}{\partial\varepsilon^r_{ij}}\frac{\partial\bar{\varepsilon}}{\partial\varepsilon^r_{kl}}, \tag{4.11}$$

$$\frac{\partial\bar{\varepsilon}}{\partial\varepsilon^r_{ij}}=\frac{2}{3}\left(f-\frac{J^e_3}{J^e_2}f'\right)\frac{e^r_{ij}}{J^e_2}+\frac{4}{3}\frac{f'}{(J^e_3)^2}\left(e^r_{ik}e^r_{kj}-\frac{1}{2}J^{e2}_2\delta_{ij}\right), \tag{4.12}$$

and

$$\frac{\partial^2\overline{\varepsilon}}{\partial\varepsilon^r_{ij}\varepsilon^r_{kl}}=\delta_{ij}\delta_{kl}\left[\frac{4}{9}\frac{(J^e_2)^3}{(J^e_3)^4}f''-\frac{2}{9}\frac{1}{J^e_2}\left(f-\frac{J^e_3}{J^e_2}f'\right)-\frac{8}{9}\frac{(J^e_2)^4}{(J^e_3)^5}f'\right]+(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk})\left[\frac{1}{3}\frac{1}{J^e_2}\left(f-\frac{J^e_3}{J^e_2}f'\right)\right]+(\delta_{ij}e^r_{kl}+\delta_{kl}e^r_{ij})$$
$$\times\left[\frac{4}{9}\frac{1}{J^e_2J^e_3}f''-\frac{8}{9}\frac{1}{(J^e_3)^2}f'\right]+e^r_{ij}e^r_{kl}\left[\frac{4}{9}\frac{(J^e_3)^2}{(J^e_2)^5}f''-\frac{4}{9}\frac{1}{(J^e_2)^3}\left(f-\frac{J^e_3}{J^e_2}f'\right)\right]+(\delta_{ij}e^r_{km}e^r_{ml}+\delta_{kl}e^r_{im}e^r_{mj})\left[\frac{16}{9}\frac{(J^e_2)^2}{(J^e_3)^5}f'-\frac{8}{9}\frac{J^e_2}{(J^e_3)^4}f''\right]$$
$$+(\delta_{ik}e^r_{jl}+\delta_{jl}e^r_{ik}+\delta_{il}e^r_{jk}+\delta_{jk}e^r_{il})\left[\frac{2}{3}\frac{1}{(J^e_3)^2}f'\right]+(e^r_{ij}e^r_{km}e^r_{ml}+e^r_{im}e^r_{mj}e^r_{kl})\left[-\frac{8}{9}\frac{1}{(J^e_2)^3J^e_3}f''\right]$$
$$+e^r_{im}e^r_{mj}e^r_{kn}e^r_{nl}\left[\frac{16}{9}\frac{1}{J^e_2(J^e_3)^4}f''-\frac{32}{9}\frac{1}{(J^e_3)^5}f'\right]. \tag{4.13}$$

Note that Eq. (4.13) is the symmetric form of a similar expression appearing in Ref. [6]. The apparent complexity of these expressions is unfortunate, but is a result of the simple assumption that the internal state of the material can be described using only the remanent strain as an internal variable. The utility of these expressions will be demonstrated when the theory is compared to the more detailed micromechanical model used in Sec. 2. Note that the back stresses and hardening moduli depend directly on the expressions in Eqs. (4.12) and (4.13). In order for the back stresses and hardening moduli to be finite, certain restrictions must be placed on f and its derivatives at $J^e_3/J^e_2=0$. These conditions can be obtained by expanding f into a power series about zero and requiring the expressions in Eqs. (4.12) and (4.13) to be finite. Specifically, $f^{\mathrm{I}}_0=f^{\mathrm{II}}_0=f^{\mathrm{IV}}_0=f^{\mathrm{V}}_0=0$, where f^n_0 represents the nth derivative of f with respect to J^e_3/J^e_2 at $J^e_3/J^e_2=0$. These restrictions must be true for any f defining a strainlike variable that is used in the hardening potential. For example, Landis [6] assumed that saturation occurs when the minimum principal remanent strain reaches a critical value. The minimum principal remanent strain can be written as

$$\varepsilon^r_{\mathrm{III}}=J^e_2\left[-\frac{1}{2}\cos\left(\frac{\theta}{3}\right)-\frac{\sqrt{3}}{2}\sin\left(\frac{\theta}{3}\right)\right],\quad\text{with}\quad\theta=\arccos[(J^e_3/J^e_2)^3].$$

Note that the function in square brackets satisfies $f^{\mathrm{I}}_0=f^{\mathrm{II}}_0=f^{\mathrm{IV}}_0=f^{\mathrm{V}}_0=0$, and furthermore $f_0, f^{\mathrm{III}}_0, f^{\mathrm{VI}}_0\neq 0$.

The ultimate test of this phenomenological theory is to compare it to experimental measurements on real materials. However, especially for ferroelastic ceramics, it is difficult to investigate the strain saturation behavior due to the large stresses required. The ceramic material tends to fracture at large stresses. Furthermore, as noted in Sec. 2, the stress paths required to generate tests like those presented in Fig. 2 are not simple, and require more information than is possible to obtain in a single experiment. Hence, the next best test of the theory is to compare it to the predictions of a more detailed micromechanical model. Figure 5 is just such a comparison.

On Fig. 5 the bold lines represent the predictions of the self-consistent model and the thin lines are the predictions of the phenomenological theory for different proportional stress path loadings of the material. The parameters for the phenomenological theory were chosen to fit the uniaxial tension and compression data. On this graph are results for a uniaxial tension, uniaxial compression, and two other proportional stressing paths, including a pure shear stress test. The type of proportional loading path is characterized by the stress invariant ratio $\sigma_e/\sqrt[3]{9s_{ij}s_{jk}s_{ki}/2}$, with -1, 0, and 1 corresponding to axisymmetric compression, pure shear stressing, and axisymmetric tension respectively. Other computations for stress invariant ratios of -0.5 and 0.5 were also performed, but not included on Fig. 5 in order to avoid clutter around the pure shear results. The comparison of the phenomenological theory to these results was favorable as well. Notice that the pure shear stress test is not equal to the $J^e_3/J^e_2=0$ test of Fig. 2. For a material with a J_2 switching surface and J_2 hardening potential, the pure shear stress and $J^e_3/J^e_2=0$ tests would yield identical results. Note that the phenomenological theory predicts the saturation strain levels accurately. This is to be expected for the tension, compression, and in fact any of the proportional remanent strain path tests since the functional fit to the saturation behavior of Eqs. (4.2) and (4.3) has been used. However, the agreement of the model to the proportional stressing tests is a true prediction of the phenomenological theory. No parameters of the model have been adjusted in order to fit these micromechanical model tests. Furthermore, notice that even the computation with the constant stress invariant ratio of -0.8 (significant compressive stress) yields remanent strains that are more tensile than the remanent strain controlled test with $J^e_3/J^e_2=0.8$. Note, of course, that the tensile remanent strain is not aligned with the compressive stress direction, but rather the direction with the largest stress deviator component. The phenomenological model is able to capture this behavior to reasonable accuracy.

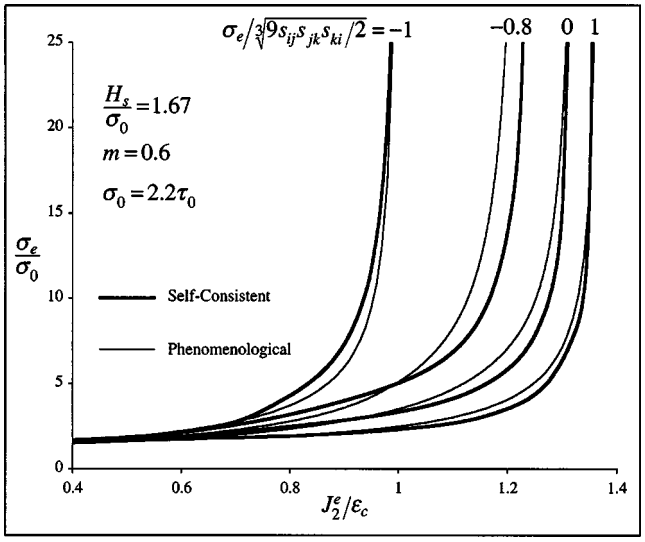


Fig. 5 A comparison of the phenomenological theory to the self-consistent calculations for different proportional stressing paths. The stresses and strains are now normalized by the parameters of the phenomenological model. Note that the ratio $\sigma_e/\sqrt[3]{9s_{ij}s_{jk}s_{ki}/2}=-1,0,1$ represents axisymmetric compression, pure shear stressing, and axisymmetric tension, respectively. Notice that even the stress path with significant compression results in an ultimate remanent strain state with significant extension. Of course, the tensile strain is not aligned with the compressive stress in this situation.

5 Discussion

In this work constitutive models for polycrystalline ferroelastic materials with an underlying tetragonal crystal structure have been investigated and developed. In Sec. 2 the self-consistent model of Huber et al. [10] was implemented to investigate the remanent strain saturation behavior of these materials. It was found that the saturation strain for axisymmetric extension was greater than the saturation strain for axisymmetric contraction by a factor of 1.37. Furthermore the entire range of remanent strain saturation states between axisymmetric extension and contraction were characterized using the remanent strain invariant ratio J_3^e/J_2^e. Equations (4.2) and (4.3) were proposed as an accurate fit for the derivation of the saturation curve presented in Fig. 4.

In Sec. 3 a phenomenological constitutive framework for ferroelastic materials was proposed. The theory assumes that the internal state of the material can be characterized completely by the remanent strain. In other words, the remanent strain components are the internal variables that determine the elastic properties, the switching surface, and the nonlinear hardening of the material. The theory allows for the elastic properties of the material to change with remanent deformation, and accounts for these changes in a thermodynamically consistent fashion within the definition of the switching surface space. Convexity of the switching surface in a modified stress space and an associated flow rule for the remanent strain increments follow from assuming that the material obeys the postulate of maximum plastic dissipation. Equations (3.16) and (3.17) represent the general form for the incremental constitutive relations from the phenomenological theory.

Section 4 of the paper is devoted to determining functional forms for the switching surface and hardening potential. For the micromechanical calculations of Sec. 2, the crystallites were assumed to be elastically isotropic. Hence, the functional form for the elastic coefficients in the phenomenological model were also taken to be isotropic and independent of remanent strain. To maintain some simplicity, the switching surface was taken to be a sphere in the modified deviatoric stress space. It was noted that the initial switching surface for the self-consistent material actually follows a Tresca criterion. Furthermore, the previous studies of Hutchinson [15] and Huber et al. [10] have shown that the shape of the switching surface evolves in a nontrivial way with remanent deformation. However, these features were not investigated or accounted for in this paper. Such modifications are left for possible future work, but are expected to yield only incremental improvements in the predictions of the model.

The remanent potential Ψ^r was the last component requiring a fit for the phenomenological model. The primary role of the remanent potential in the phenomenological theory is to account for the complicated dependence of the saturation strain on the remanent strain. Hence the first requirement for Ψ^r is to define a strainlike variable that predicts the strain saturation behavior displayed in Fig. 4. Such a variable is defined in Eq. (4.4) with the function f given by Eqs. (4.2) and (4.3). This variable, $\overline{\varepsilon}$, must be less than some critical value ε_c. Hence, as $\overline{\varepsilon}$ approaches ε_c the remanent potential must grow without bound. Then, as a result of the observation from the micromechanical calculations that the initial deformation is essentially independent of J_3^e/J_2^e, a new strainlike variable, ε^*, was introduced in Eq. (4.5). Due to the weighting function w ε^* is close to J_2^e for small values of $\overline{\varepsilon}$, and approaches $\overline{\varepsilon}$ as $\overline{\varepsilon}$ approaches ε_c. Finally, the ultimate specification of Ψ^r requires a fit to a stress versus remanent strain curve. The likely candidates for this fit are either a uniaxial tension or compression curve. To summarize, three observations were used to specify Ψ^r. First, the function f was introduced to fit the saturation strain as a function of J_3^e/J_2^e, then w was implemented to fit the observation that small deformations are independent of J_3^e/J_2^e, and finally $d\Psi^r/d\varepsilon^*$ was fit to a uniaxial stress versus remanent strain curve.

Note that each of the functions introduced to fit specific behaviors is in no way connected to specified proportional stressing paths. For example, the loading path traversed to determine the strain saturation condition when $J_3^e/J_2^e=0$ is *not* a pure shear loading path. In fact, the load path required to keep the remanent strain ratio equal to zero is not even constant. Initially, the loading path is pure shear stressing. Recall that pure shear is an extension in one direction and an equal contraction in an orthogonal direction. Since the material is able to extend more easily than it can contract, in order to maintain pure shear remanent strain the compressive load in the contraction direction must be greater than the than the tensile load in the extension direction as the strain approaches saturation. In fact, as saturation is approached with the remanent strain ratio remaining constant at $J_3^e/J_2^e=0$, the applied stress approaches a uniaxial compression. Therefore, comparing the phenomenological model to the micromechanical model for different proportional stressing paths is a valid test of the phenomenological theory. Figure 5 is a comparison of the phenomenological theory to the self-consistent calculations, and the agreement is quite good.

Again, it is emphasized that the self-consistent computations performed in Sec. 2 and the results of that model are valid for materials with tetragonal crystal structure only. In order to investigate other crystal structures, like orthorhombic or rhombohedral, and the possibility of combinations of these with tetragonal domains, the single crystal constitutive law used in the self-consistent model would need to be modified. However, the primary result that would be obtained from such calculations is a strain saturation curve like the one presented in Fig. 4 for tetragonal materials. Furthermore, simpler micromechanical models that do not account for grain to grain interactions can be used to obtain this strain saturation curve. Then, the only significant change to the phenomenological theory would be to the functional form of f defined in Eqs. (4.2) and (4.3).

The goal of this work was to understand the strain saturation behavior of ferroelastic materials and incorporate this information into a phenomenological constitutive theory. The phenomenological model is useful for structural stress analyses on systems containing ferroelastic materials. Such calculations will most likely be performed within a finite element framework. Hence, it is important that the constitutive response can be integrated rapidly. This consideration of computational speed rules out the more detailed self-consistent model, since any one curve (aside from the axisymmetric cases) on Figs. 2, 3, or 5 requires a few days to compute, whereas the corresponding phenomenological theory requires a few seconds. Finally, while the phenomenological theory has been shown to be in good agreement with the more detailed micromechanical model, the ultimate test of this new constitutive law will be against careful multiaxial loading experiments.

Acknowledgment

This work was supported by Contract No. DAAD10-02-1-0241 from the Army Research Office.

References

[1] Fett, T., Müller, S., Munz, D., and Thun, G., 1998, "Nonsymmetry in the Deformation Behavior of PZT," J. Mater. Sci. Lett., **17**, pp. 261–265.
[2] Fett, T., and Thun, G., 1998, "Nonsymmetric Deformation Behavior of Lead Zirconate Titanate Determined in Bending Tests," J. Am. Ceram. Soc., **81**, pp. 269–272.
[3] Karastamatis, T., Lupascu, D., Lucato, S., and Lynch, C. S., 2001, "The Effect of Grain Size on the R-Curve Behavior of Lead Zirconate Titanate," Proc. SPIE, **4333**, pp. 38–40.
[4] Landis, C. M., and McMeeking, R. M., 1999, "A Phenomenological Constitutive Law for Ferroelastic Switching and a Resulting Asymptotic Crack Tip Solution," J. Intell. Mater. Syst. Struct., **10**, pp. 155–163.
[5] Huber, J. E., and Fleck, N. A., 2001, "Multi-Axial Electrical Switching of a Ferroelectric: Theory versus Experiment," J. Mech. Phys. Solids, **49**, pp. 785–811.
[6] Landis, C. M., 2002, "Fully Coupled, Multi-Axial, Symmetric Constitutive Laws for Polycrystalline Ferroelectric Ceramics," J. Mech. Phys. Solids, **50**, pp. 127–152.

[7] Lu, W., Fang, D. N., Li, C. Q., and Hwang, K. C., 1999, ''Nonlinear Electric-Mechanical Behavior and Micromechanics Modelling of Ferroelectric Domain Evolution,'' Acta Mater., **47**, pp. 2913–2926.
[8] Frölich, A., 2001, ''Mikromechanisches Modell zur Ermittlung Effektiver Materialeigenschaften von Piezoelektrischen Polykristallen,'' Dissertation thesis, Universität Karlsruhe, Germany.
[9] Kamlah, M., 2001, ''Ferroelectric and Ferroelastic Piezoceramics—Modeling and Electromechanical Hysteresis Phenomena,'' Continuum Mech. Thermodyn., **13**, pp. 219–268.
[10] Huber, J. E., Fleck, N. A., Landis, C. M., and McMeeking, R. M., 1999, ''A Constitutive Model for Ferroelectric Polycrystals,'' J. Mech. Phys. Solids, **47**, pp. 1663–1697.
[11] Hill, R., 1965, ''Continuum Micro-Mechanics of Elastoplastic Polycrystals,'' J. Mech. Phys. Solids, **13**, pp. 89–101.
[12] Hill, R., 1965, ''A Self-Consistent Mechanics of Composite Materials,'' J. Mech. Phys. Solids, **13**, pp. 213–222.
[13] Hill, R., 1966, ''Generalized Constitutive Relations for Incremental Deformation of Metal Crystals by Multislip,'' J. Mech. Phys. Solids, **14**, pp. 95–102.
[14] Hill, R., 1967, ''The Essential Structure of Constitutive Laws for Metal Composites and Polycrystals,'' J. Mech. Phys. Solids, **15**, pp. 79–95.
[15] Hutchinson, J. W., 1970, ''Elastic-Plastic Behavior of Polycrystalline Metals and Composites,'' Proc. R. Soc. London, Ser. A, **319**, pp. 247–272.
[16] Rice, J. R., 1970, ''On the Structure of Stress-Strain Relations for Time-Dependent Plastic Deformation in Metals,'' J. Appl. Mech., **37**, pp. 728–737.
[17] Rice, J. R., 1971, ''Inelastic Constitutive Relations for Solids: An Internal-Variable Theory and Its Applications to Metal Plasticity,'' J. Mech. Phys. Solids, **19**, pp. 433–455.
[18] Rice, J. R., 1975, ''Continuum Mechanics and Thermodynamics of Plasticity in Relation to Microscale Deformation Mechanisms,'' in Constitutive Relations in Plasticity edited by A. S. Argon, MIT Press, Cambridge, Massachusetts, pp. 23–79.

Relationship Among Coefficient Matrices in Symmetric Galerkin Boundary Element Method for Two-Dimensional Scalar Problems

G. Y. Yu
School of Civil & Environmental Engineering,
Nanyang Technological University,
Nanyang Avenue, Singapore 639798
Tel: (65) 6790 5284, Fax: (65) 6792 1650
e-mail: cgyyu@ntu.edu.sg

Based on the assumption that solutions from different methods should be the same, the relationship among weakly singular, strongly singular and hypersingular matrices associated with symmetric Galerkin boundary element method (SGBEM) is derived in this paper. Hypersingularity is avoided through matrix manipulations so that only weakly and strongly singularities need to be solved. Compared with the advantages brought about by symmetry, the additional computation caused by matrix manipulations is not so important in many cases, especially for time-domain problems or when one wants to couple BEM with other symmetric schemes. Simplicity is the advantage of the current method over the traditional SGBEM. Both steady-state and time-domain potential problems have been studied, and two numerical examples are included to show the effectiveness and accuracy of the present formulation. [DOI: 10.1115/1.1598478]

1 Introduction

The traditional collocation boundary element method (TCBEM) and the asymmetric Galerkin BEM (AGBEM) have some certain unpleasant features. The most pertinent is the lack of symmetry for some coefficient matrices, which will make it less efficient to couple BEM with other symmetric schemes. The symmetric Galerkin BEM (SGBEM) formulation was first proposed by Sirtori [1] for linear elastic analysis. The main problem encountered is that one has to calculate the three kinds of coefficient matrices, which may respectively contain weakly singular, strongly singular, and hypersingular integrals [2]. Although numerous papers have been published to deal with the hypersingular integrals, there are still many spaces that need more research works. Carini et al. [3] classified the existing strategies, proposed by different researchers for solving the hypersingular integration, into three principle groups: regularization, finite part, and direct approach. The regularization procedure analytically converts the strongly singular and hypersingular integrals into, at most, weakly singular integrals which can then be calculated directly by numerical methods [4–6]. The finite-part approach is based on the numerical method for the finite part of the principal value of the singular integral [7]. The difference of this approach from the direct integration approach is that the final integration is performed numerically [8,9]. The direct treatment of singular integrals in BEM is to part the source and field points temporarily to allow an analytical evaluation of the singular integral, and then derive the limit as the source and field points come together [10–12]. All of these methods have been proved to be efficient but need a complex procedure; therefore a simple method is necessary as an alternative, which can be easily used to convert an existing AGBEM code to a SGBEM one.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Apr. 23, 2002; final revision, Dec. 17, 2002. Associate Editor: T. E. Tezduyar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

In fact, formulations for SGBEM and AGBEM are not independent. If the numerical error is negligible, their solutions should be same, while the rigid body movement is considered as a special solution here. Based on this assumption, the relationship among the three kinds of coefficient matrices appearing in SGBEM is given in this paper, from which the hypersingular integral can be avoided through matrix manipulations. As only weak and strong singularities need to be solved, such an indirect method makes SGBEM much easier than it was before. The indirect SGBEM formulation is derived directly from the AGBEM one; there is no doubt that they yield exactly the same result. However, the SGBEM formulation is symmetric which makes it more efficient to couple BEM with other symmetric numerical methods.

The relationship is given for both steady-state and time domain problems [13,14] in this paper. Compared with the traditional SGBEM, the use of the indirect method may need more computer time, especially for steady-state problems. But for time-domain problems, such a cost increment is not significant. For both steady-state and time-domain problems, the indirect method provides an alternative for those who want an easier method to convert directly an existing AGBEM code to a SGBEM one. Two numerical examples are included to illustrate the effectiveness and accuracy of the present method.

2 Relationship among the Three Kinds of Coefficient Matrices

2.1 Steady-State Potential Problems. The two-dimensional (2D) steady-state problem can be described as a typical Laplace problem. Consider the potential u at any point ξ in the domain Ω bounded by Γ having unit outward normal $\mathbf{n}^0$ at the boundary point X (see Fig. 1). The problem can be described in boundary integration form as

$$2\pi c(\xi)u(\xi)=\int_\Gamma u^*(X,\xi)p(X)d\Gamma(X)-\int_\Gamma p^*(X,\xi)u(X)d\Gamma(X), \tag{1}$$

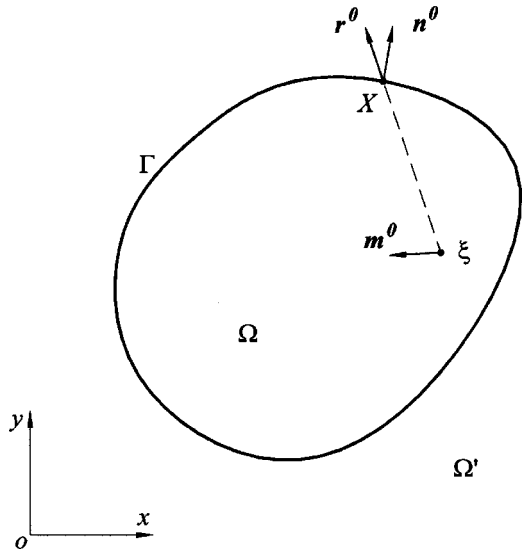


Fig. 1 Definition of some symbols

$$\text{where } c(\xi)=\begin{cases} 1 & \xi \text{ in the internal domain } \Omega, \\ 1/2 & \xi \text{ on the smooth boundary}, \\ 0 & \xi \text{ in the outer domain } \Omega'. \end{cases}$$

$u(X)$ and $p(X)=\partial u/\partial \mathbf{n}$ are the potential and flux at the boundary point X respectively, $u^*(X,\xi)=-\ln r$ is the fundamental solution,

$$p^*(X,\xi)=\frac{\partial u^*(X,\xi)}{\partial \mathbf{n}}=-\frac{(\mathbf{r}^0\cdot\mathbf{n}^0)}{r}$$

is the normal derivative of $u^*(X,\xi)$, and r is the distance between points X and ξ.

The space derivative of Eq. (1) in the direction of $\mathbf{m}$ becomes

$$2\pi c(\xi)\frac{\partial u(\xi)}{\partial \mathbf{m}}=\int_\Gamma \frac{\partial u^*(X,\xi)}{\partial \mathbf{m}}p(X)d\Gamma(X) - \int_\Gamma \frac{\partial p^*(X,\xi)}{\partial \mathbf{m}}u(X)d\Gamma(X). \tag{2}$$

If ξ is on the boundary and $\mathbf{m}$ is the outward normal, the above equation becomes

$$2\pi c(\xi)p(\xi)=\int_\Gamma \frac{\partial u^*(X,\xi)}{\partial \mathbf{m}}p(X)d\Gamma(X) - \int_\Gamma \frac{\partial p^*(X,\xi)}{\partial \mathbf{m}}u(X)d\Gamma(X). \tag{3}$$

Consider a set of discrete points $X_j, j=1,2,\ldots J$, on the boundary Γ. $u(X)$ and $p(X)$ are approximately expressed by functions $\eta_j(X)$ as follows:

$$\begin{cases} u(X)=\sum_{j=1}^{J}\eta_j(X)u_j \\ p(X)=\sum_{j=1}^{J}\eta_j(X)p_j \end{cases}, \tag{4}$$

where the same spatial interpolation function is used for u and p, $u_j=u(X_j)$, and $p_j=p(X_j)$.

Substituting Eq. (4) into Eqs. (1) and (3) leads to

$$2\pi c(\xi)u(\xi)=\sum_{j=1}^{J}G_j^{uu}(\xi)p_j-\sum_{j=1}^{J}G_j^{pu}(\xi)u_j \tag{5}$$

and

$$2\pi c(\xi)p(\xi)=\sum_{j=1}^{J}G_j^{up}(\xi)p_j-\sum_{j=1}^{J}G_j^{pp}(\xi)u_j, \tag{6}$$

where

$$G_j^{uu}(\xi)=\int_\Gamma u^*(X,\xi)\,\eta_j(X)d\Gamma(X), \tag{7}$$

$$G_j^{pu}(\xi)=\int_\Gamma p^*(X,\xi)\,\eta_j(X)d\Gamma(X), \tag{8}$$

$$G_j^{up}(\xi)=\int_\Gamma \frac{\partial u^*(X,\xi)}{\partial \mathbf{m}}\,\eta_j(X)d\Gamma(X), \tag{9}$$

$$G_j^{pp}(\xi)=\int_\Gamma \frac{\partial p^*(X,\xi)}{\partial \mathbf{m}}\,\eta_j(X)d\Gamma(X). \tag{10}$$

As r approaches to zero, hypersingularity appears in $G_j^{pp}(\xi)$, which requires the spatial interpolation function $\eta_j(X)$ to be $C^{1,\alpha}$ around the point ξ. In order to remove this requirement, the Galerkin method is applied to Eqs. (5) and (6). The AGBEM formulations, corresponding to the boundary node ξ_i, can be written as

$$\sum_{j=1}^{J}C_{Gij}u_j=\sum_{j=1}^{J}G_{Gij}^{uu}p_j-\sum_{j=1}^{J}G_{Gij}^{pu}u_j, \tag{11}$$

$$\sum_{j=1}^{J}C_{Gij}p_j=-\sum_{j=1}^{J}G_{Gji}^{pu}p_j+\sum_{j=1}^{J}G_{Gij}^{pp}u_j, \tag{12}$$

where

$$G_{Gij}^{uu}=\int_\Gamma\int_\Gamma u^*(X,\xi)\,\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi), \tag{13}$$

$$G_{Gij}^{pu}=\int_\Gamma\int_\Gamma p^*(X,\xi)\,\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi), \tag{14}$$

$$G_{Gij}^{pp}=-\int_\Gamma\int_\Gamma \frac{\partial p^*(X,\xi)}{\partial \mathbf{m}}\,\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi) =-\int_\Gamma\int_\Gamma \frac{\partial}{\partial \mathbf{m}}\left[\frac{\partial u^*(X,\xi)}{\partial \mathbf{n}}\right]\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi), \tag{15}$$

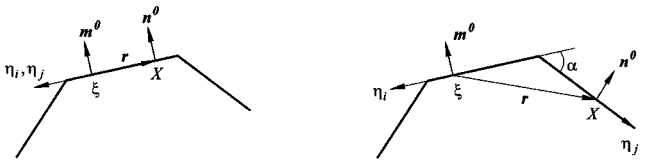
$$C_{Gij}=\pi\int_\Gamma \eta_i(\xi)\,\eta_j(\xi)d\Gamma(\xi), \tag{16}$$

in which $c(\xi)$ is set equal to 0.5 for ξ lying on the boundary. The subscript "G" denotes the variables associated with the Galerkin method.

It should be noted that Eq. (13) contains only weakly singular integrals which can be evaluated either through standard numerical procedure or through analytical method [15]. Figure 2 shows two special cases using linear elements. In case (a), when integration in Eq. (14) runs along the same element (see Fig. 2(a)) with respect to the two spatial variables, X and ξ,

$$p^*(X,\xi)=\frac{\partial u^*(X,\xi)}{\partial \mathbf{n}}=\frac{\partial u^*(X,\xi)}{\partial r}(\mathbf{r}^0\cdot\mathbf{n}^0)=0.$$

In case (b), when integration in Eq. (14) runs along two adjacent elements (see Fig. 2(b)), it leads to strongly singular integral as the distance between X and ξ tends to zero. Such a strongly sin-



(*a*) ξ and X are in the same element (*b*) ξ and X are in two adjacent elements

Fig. 2 Two special cases for linear elements

gular integral can be evaluated through analytical procedure given in the Appendix for G^{pu}_{Gij} or by the regularization method [6]. However, when the higher-order interpolation function is used, the analytical procedure should be used for both of these two cases so as to increase the accuracy.

The hypersingular integral in Eq. (15), which needs a complex procedure to deal with, is evaluated here by an indirect method: writing Eqs. (11) and (12) in matrix form as

$$\{[C_G]+[G^{pu}_G]\}\{u\}=[G^{uu}_G]\{p\} \tag{17}$$

and

$$\{[C_G]+[G^{pu}_G]^{\mathrm{T}}\}\{p\}=[G^{pp}_G]\{u'\}=[G^{pp}_G](\{u\}+\{c\})=[G^{pp}_G]\{u\}, \tag{18}$$

where $\{u\}$ and $\{u'\}$ are the displacements corresponding to $\{p\}$, calculated from Eqs. (17) and (18), respectively. It has been well recognized that Eqs. (17) and (18) will yield the correct result which should be unique except the rigid body movement. Therefore the only difference between $\{u\}$ and $\{u'\}$ is the constant displacement $\{c\}$ corresponds to $\{p\}=0$ and $[G^{pp}_G]\{c\}=0$.

From Eqs. (17) and (18) one can get

$$[G^{pp}_G]\{u\}=([C_G]+[G^{pu}_G]^{\mathrm{T}})[G^{uu}_G]^{-1}([C_G]+[G^{pu}_G])\{u\} \tag{19}$$

which is true for any $\{u\}$ and corresponding $\{p\}$. Therefore the relationship among $[G^{pp}_G]$, $[G^{pu}_G]$, and $[G^{uu}_G]$ for steady-state problems can be written as

$$[G^{pp}_G]=([C_G]+[G^{pu}_G]^{\mathrm{T}})[G^{uu}_G]^{-1}([C_G]+[G^{pu}_G]), \tag{20}$$

where as $[C_G]$ and $[G^{uu}_G]$ are symmetric matrices, $[G^{pp}_G]$ is also symmetric.

From Eq. (17), one can see that if no element coincides with another one, only $\{p\}=0$ can satisfy the solution for zero or constant $\{u\}$ which makes the left part of Eq. (17) equal zero. Otherwise, if $\{p\}\neq 0$, $\{u\}$ cannot be zero or constant, as $p=\partial u/\partial n$ (no internal source or initial boundary condition is considered here). Therefore $[G^{uu}_G]$ is invertible if no element coincides with another one. Equation (20) is the formulation to calculate $[G^{pp}_G]$ through matrix manipulations. As matrices $[C_G]$, $[G^{uu}_G]$, and $[G^{pu}_G]$ have all appeared in the AGBEM formulation, no hypersingularity needs to be considered, the indirect method is much easier compared with the traditional SGBEM. With Eq. (20), an existing AGBEM computer code can be easily modified to a SGBEM one.

For the case when only BEM is used, symmetric formulation can be established through Eqs. (17) and (18): Equation (17) is applied to all those sections of the boundary where displacement components are prescribed (Γ_1), while Eq. (18) is applied to the sections where traction components are prescribed (Γ_2). For the case of BEM/FEM coupling procedure, both Eqs. (17) and (18) should be used on the interface.

2.2 Time-Domain Potential Problems. Consider a set of discrete quantities at time t_n, $n=1,2,\ldots,N$. $u(X,t)$ and $p(X,t)$ can be approximated through interpolations as [15,16]

$$\begin{cases} u(X,t)=\sum_{j=1}^{J}\sum_{m=1}^{N}\phi^m(t)\,\eta_j(X)u_j^m \\ p(X,t)=\sum_{j=1}^{J}\sum_{m=1}^{N}\gamma^m(t)\,\eta_j(X)p_j^m \end{cases}, \tag{21}$$

where $u^m_j=u(X_j,t_m)$ and $p^m_j=p(X_j,t_m)$, and ϕ^m, γ^m, η^m are interpolation functions.

Similar to Eq. (1), the boundary integral equation can be written as [15,16]

$$4\pi c(\xi)u(\xi,t)=\int_\Gamma\int_0^{t^+}u^*(X,t;\xi,\tau)p(X,\tau)\,d\tau\,d\Gamma(X) - \int_\Gamma\int_0^{t^+}p^*(X,t;\xi,\tau)u(X,\tau)\,d\tau\,d\Gamma(X). \tag{22}$$

The space derivative of Eq. (22) in the direction of outward normal $\mathbf{m}$ can be written as

$$4\pi c(\xi)p(\xi,t)=\int_\Gamma\int_0^{t^+}\frac{\partial u^*(X,t;\xi,\tau)}{\partial\mathbf{m}}p(X,\tau)\,d\tau\,d\Gamma(X) - \int_\Gamma\int_0^{t^+}\frac{\partial u^*(X,t;\xi,\tau)}{\partial\mathbf{m}}u(X,\tau)\,d\tau\,d\Gamma(X), \tag{23}$$

where

$$u^*(X,\xi)=\frac{2c}{\sqrt{c^2(t-\tau)^2-r^2}}H[c(t-\tau)-r]$$

is the fundamental solution,

$$p^*(X,t;\xi,\tau)=\frac{\partial u^*(X,t;\xi,\tau)}{\partial\mathbf{n}}=\frac{\partial u^*(X,t;\xi,\tau)}{\partial r}(\mathbf{r}^0\cdot\mathbf{n}^0),$$

$H[c(t-\tau)-r]$ is the Heaviside function, and c is the wave velocity.

Applying the Galerkin method to Eqs. (22) and (23) with respect to the boundary node ξ_i, it results in

$$\int_\Gamma 4\pi c(\xi)u(\xi,t)\,\eta_i(\xi)\,d\Gamma(\xi) = \int_\Gamma\int_\Gamma\int_0^{t^+}u^*(X,t;\xi,\tau)p(X,\tau)\,d\tau\,d\Gamma(X)\,\eta_i(\xi)\,d\Gamma(\xi) - \int_\Gamma\int_\Gamma\int_0^{t^+}p^*(X,t;\xi,\tau)u(X,\tau)\,d\tau\,d\Gamma(X)\,\eta_i(\xi)\,d\Gamma(\xi) \tag{24}$$

and

$$\int_\Gamma 4\pi c(\xi)p(\xi,t)\,\eta_i(\xi)\,d\Gamma(\xi) = \int_\Gamma\int_\Gamma\int_0^{t^+}\frac{\partial u^*(X,t;\xi,\tau)}{\partial\mathbf{m}}p(X,\tau)\,d\tau\,d\Gamma(X)\,\eta_i(\xi)\,d\Gamma(\xi) - \int_\Gamma\int_\Gamma\int_0^{t^+}\frac{\partial p^*(X,t;\xi,\tau)}{\partial\mathbf{m}}u(X,\tau)\,d\tau\,d\Gamma(X)\,\eta_i(\xi)\,d\Gamma(\xi). \tag{25}$$

The discrete form,

Table 1 Time integrated kernels

Case	$F_F^{(m-1)n}(X,\xi)$	$F_I^{mn}(X,\xi)$	$D_F^{(m-1)n}(X,\xi)$	$D_I^{mn}(X,\xi)$
1	A_1A_2	$-A_1A_2-A_5A_7$	$\frac{A_1}{A_2}$	$-\frac{A_1A_2}{r}$
2	$A_1A_2+A_6A_7$	$-A_1A_2-A_5A_7$	$\frac{1}{r}\frac{A_1(A_4)^2}{A_2}$	$-\frac{A_1A_2}{r}$
3	$-A_3A_4+A_1A_2+A_6A_8$	$A_3A_4-A_1A_2-A_5A_7$	$\frac{A_4}{A_2}\left(\frac{2c\Delta t}{A_1A_4+A_2A_3}\right)$	$-\frac{A_2}{A_4}\left(\frac{2c\Delta t}{A_1A_4+A_2A_3}\right)$
4	0	0	0	0

$$\sum_{j=1}^{J} C_{GDij}u_j^n+\sum_{m=1}^{N}\sum_{j=1}^{J} G_{GDij}^{pumn}u_j^m=\sum_{m=1}^{N}\sum_{j=1}^{J} G_{GDij}^{uumn}p_j^m, \quad (26)$$

$$\sum_{j=1}^{J} C_{GDij}p_j^n+\sum_{m=1}^{N}\sum_{j=1}^{J} G_{GDji}^{pumn}p_j^m=\sum_{m=1}^{N}\sum_{j=1}^{J} G_{GDij}^{ppmn}u_j^m, \quad (27)$$

where

$$G_{GDij}^{uumn}=\int_\Gamma\int_\Gamma\int_0^{t^+} u^*(X,t_n;\xi,\tau)\gamma^m(\tau)d\tau\eta_j(X)d\Gamma(X) \times\eta_i(\xi)d\Gamma(\xi), \quad (28)$$

$$G_{GDij}^{pumn}=\int_\Gamma\int_\Gamma\int_0^{t^+} p^*(X,t_n;\xi,\tau)\gamma^m(\tau)d\tau\eta_j(X)d\Gamma(X) \times\eta_i(\xi)d\Gamma(\xi), \quad (29)$$

$$G_{GDij}^{ppmn}=-\int_\Gamma\int_\Gamma\int_0^{t^+}\frac{\partial p^*(X,t_n;\xi,\tau)}{\partial\mathbf{m}} \times\phi^m(\tau)d\tau\eta_j(X)d\Gamma(X)\eta_i(\xi)d\Gamma(\xi) =-\int_\Gamma\int_\Gamma\int_0^{t^+}\frac{\partial}{\partial\mathbf{m}}\frac{\partial u^*(X,t_n;\xi,\tau)}{\partial\mathbf{n}} \times\phi^m(\tau)d\tau\eta_j(X)d\Gamma(X)\eta_i(\xi)d\Gamma(\xi), \quad (30)$$

$$C_{GDij}=2\pi\int_\Gamma\eta_i(\xi)\eta_j(\xi)d\Gamma(\xi)=2C_{Gij}, \quad (31)$$

in which the second subscript "*D*" denotes time-domain problems, to distinguish it from those quantities in previous steady-state problems.

Linear time interpolation function is used for both u and p:

$$\gamma^m(\tau)=\phi^m(\tau)=\begin{cases}\dfrac{\tau-t_{m-1}}{\Delta t} & \text{if } t_{m-1}\leqslant\tau\leqslant t_m\\ \dfrac{t_{m+1}-\tau}{\Delta t} & \text{if } t_m\leqslant\tau\leqslant t_{m+1}\\ 0 & \text{otherwise}\end{cases}. \quad (32)$$

Substitution of Eq. (32) into Eqs. (28) and (29) leads to [16]

$$G_{GDij}^{uumn}=\int_\Gamma\int_\Gamma\int_{t_{m-1}}^{t_m} u^*(X,t_n;\xi,\tau)\frac{\tau-t_{m-1}}{\Delta t}d\tau\eta_j(X)d\Gamma(X) \times\eta_i(\xi)d\Gamma(\xi)+\int_\Gamma\int_\Gamma\int_{t_m}^{t_{m+1}} u^*(X,t_n;\xi,\tau) \times\frac{t_{m+1}-\tau}{\Delta t}d\tau\eta_j(X)d\Gamma(X)\eta_i(\xi)d\Gamma(\xi) =\frac{2}{c\Delta t}\int_\Gamma\int_\Gamma[F_I^{mn}(X,\xi) +F_F^{mn}(X,\xi)]d\tau\eta_j(X)d\Gamma(X)\eta_i(\xi)d\Gamma(\xi), \quad (33)$$

$$G_{GDij}^{pumn}=\int_\Gamma\int_\Gamma\int_{t_{m-1}}^{t_m} p^*(X,t_n;\xi,\tau)\frac{\tau-t_{m-1}}{\Delta t}d\tau\eta_j(X)d\Gamma(X) \times\eta_i(\xi)d\Gamma(\xi)+\int_\Gamma\int_\Gamma\int_{t_m}^{t_{m+1}} p^*(X,t_n;\xi,\tau) \times\frac{t_{m+1}-\tau}{\Delta t}d\tau\eta_j(X)d\Gamma(X)\eta_i(\xi)d\Gamma(\xi) =\frac{2}{c\Delta t}\int_\Gamma\int_\Gamma[D_I^{mn}(X,\xi)+D_F^{mn}(X,\xi)]\frac{\partial r}{\partial\mathbf{n}}\eta_j(X) \times d\Gamma(X)\eta_i(\xi)d\Gamma(\xi), \quad (34)$$

where expressions for the kernels $D_I^{mn}(X,\xi)$, $D_F^{mn}(X,\xi)$, $F_I^{mn}(X,\xi)$, and $F_F^{mn}(X,\xi)$ depend on the position of the wave front described in the fundamental solution. They are characterized by the retarded time $t_r=t_n-r/c$, with respect to the time interval considered. Four possible situations are considered as follows:

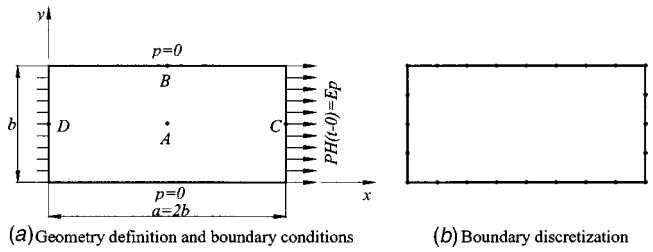
case 1: $t_m=t_n, t_{m-1}<t_r\leqslant t_m$,

case 2: $t_m<t_n, t_{m-1}<t_r\leqslant t_m$,

case 3: $t_m\leqslant t_r$,

case 4: $t_r\leqslant t_{m-1}$.

The expressions corresponding to cases 1–4, obtained by the analytical method, are shown in Table 1. In order to shorten the final expressions, the following notations are used:



(a) Geometry definition and boundary conditions (b) Boundary discretization

Fig. 3 One-dimensional rod under Heaviside-type forcing function

$$A_1=\sqrt{c(t_n-t_{m-1})-r}, \qquad A_2=\sqrt{c(t_n-t_{m-1})+r},$$

$$A_3=\sqrt{c(t_n-t_m)-r}, \qquad A_4=\sqrt{c(t_n-t_m)+r},$$

$$A_5=c(t_n-t_{m-1}), \qquad A_6=c(t_n-t_m),$$

$$A_7=\ln\left(\frac{r}{A_1A_2+A_5}\right), \qquad A_8=\ln\left(\frac{A_3A_4+A_5}{A_1A_2+A_6}\right).$$

Similar to the steady-state problems, Eq. (33) contains only weakly singular integrals which can be evaluated either through standard numerical procedure or through analytical method [15]. Figure 2 shows two special cases using linear elements. In case (a), when integration in Eq. (34) runs along the same element (see Fig. 2(a)) with respect to the two spatial variables, X and ξ,

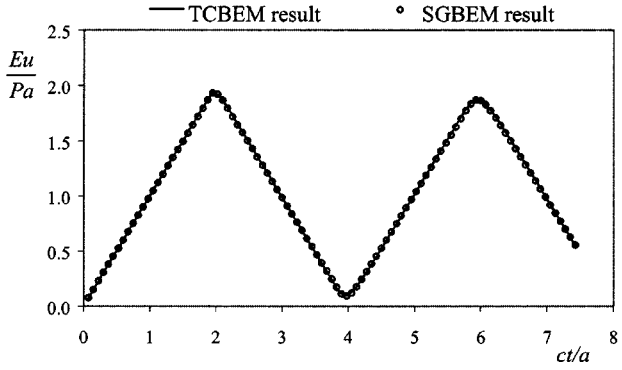
$$p^*(X,t_n;\xi,\tau)=\frac{\partial u^*(X,t_n;\xi,\tau)}{\partial \mathbf{n}}=\frac{\partial u^*(X,t_n;\xi,\tau)}{\partial r}(\mathbf{r}^0\cdot\mathbf{n}^0)=0.$$

In case (b), when integration in Eq. (34) runs along two adjacent elements (see Fig. 2(b)), it leads to strongly singular integral as the distance between X and ξ tends to zero and $m=n$. Such a strongly singular integral can be evaluated through the analytical procedure given in the Appendix for G^{pu}_{GDij} or by the regularization method [6]. However, when higher-order spatial interpolation function is used, the analytical procedure should be used for both of these two cases in order to increase the accuracy.

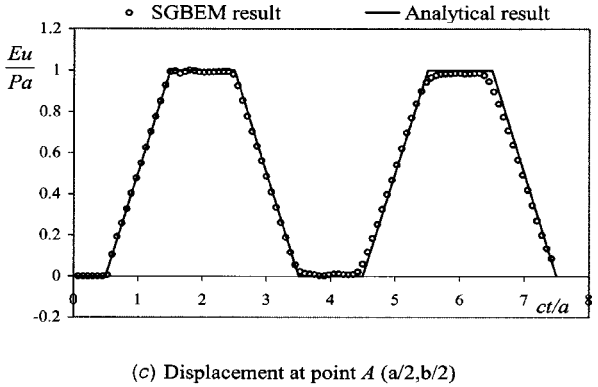
From here onward, unless causing ambiguity, G^{uunn}_{GDij}, G^{punn}_{GDij}, and G^{ppnn}_{GDij} will be written as G^{uu}_{GDij}, G^{pu}_{GDij}, and G^{pp}_{GDij}, respectively.

The hypersingular integral appearing in expression (30), which is difficult to deal with for direct methods, can also be evaluated by an indirect method. Writing in matrix form, Eqs. (26) and (27) become

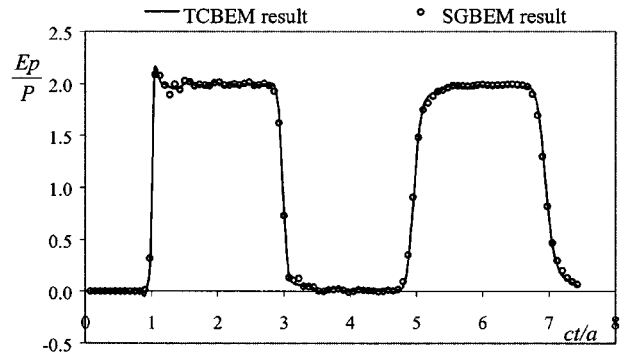
$$\{[C_{GD}]+[G^{pu}_{GD}]\}\{u\}^n=[G^{uu}_{GD}]\{p\}^n+\sum_{m=1}^{n-1}[G^{uu}_{GD}]^{mn}\{p\}^m-\sum_{m=1}^{n-1}[G^{pu}_{GD}]^{mn}\{u\}^m, \tag{35}$$



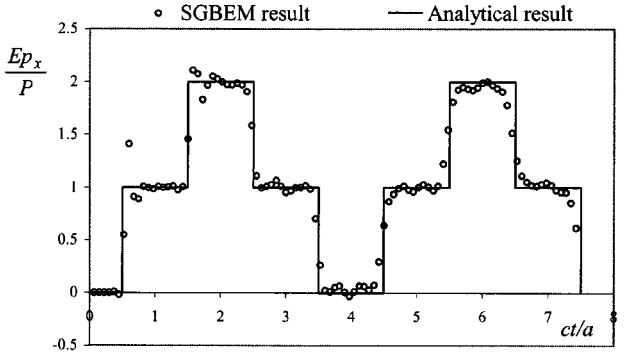
(a) Displacement at point $C(a,0)$



(c) Displacement at point A (a/2,b/2)



(b) Flux at point $D(0,0)$



(d) Flux at point A (a/2,b/2)

Fig. 4 Comparison between the results from TCBEM and SGBEM, $\theta=1.4$ and $\beta=0.6$

$$\{[C_{GD}]+[G_{GD}^{pu}]^{\mathrm{T}}\}\{p\}^n=[G_{GD}^{pp}]\{u\}^n+\sum_{m=1}^{n-1}[G_{GD}^{pp}]^{mn}\{u\}^m-\sum_{m=1}^{n-1}[G_{GD}^{pu}]^{mn}\{p\}^m, \quad (36)$$

where the rigid body movement need not be considered as it has been stated in the steady-state problem before.

Multiplying Eq. (35) by $([C_{GD}(\xi)]+[G_{GD}^{pu}]^{\mathrm{T}})[G_{GD}^{uu}]^{-1}$, one gets

$$\begin{aligned}&\{[C_{GD}(\xi)]+[G_{GD}^{pu}]^{\mathrm{T}}\}\{p\}^n\\&=\{[C_{GD}(\xi)]+[G_{GD}^{pu}]^{\mathrm{T}}\}[G_{GD}^{uu}]^{-1}\{[C_{GD}(\xi)]+[G_{GD}^{pu}]\}\{u\}^n\\&\quad+([C_{GD}(\xi)]+[G_{GD}^{pu}])^{\mathrm{T}}[G_{GD}^{uu}]^{-1}\\&\quad\times\left(\sum_{m=1}^{n-1}[G_{GD}^{pu}]^{mn}\{u\}^m-\sum_{m=1}^{n-1}([G_{GD}^{uu}]^{mn})^{\mathrm{T}}\{p\}^m\right). \quad (37)\end{aligned}$$

Comparing Eqs. (36) and (37), and bearing in mind that no matter what kind of $\{u^m\}$ and $\{p^m\}$ are used, the results from these two equations for $\{u^n\}$ and $\{p^n\}$ are the same, the relationship among $[G_{GD}^{pp}]^{mn}$, $[G_{GD}^{pu}]^{mn}$, and $[G_{GD}^{uu}]^{mn}$ for time-domain problems can be written as

$$[G_{GD}^{pp}]=([C_{GD}(\xi)]+[G_{GD}^{pu}])^{\mathrm{T}}[G_{GD}^{uu}]^{-1}([C_{GD}(\xi)]+[G_{GD}^{pu}]), \quad (38)$$

$$[G_{GD}^{pp}]^{mn}=([C_{GD}]+[G_{GD}^{pu}]^{\mathrm{T}})[G_{GD}^{uu}]^{-1}[G_{GD}^{pu}]^{mn}, \quad (39)$$

where $m=1,2,\ldots,n-1$, $[G_{GD}^{pp}]$ is a symmetric matrix.

As there is no singularity in $[G_{GD}^{pp}]^{mn}$ $(m<n)$, it can be calculated directly from Eq. (30) by the numerical method. Therefore the only cost increment for the indirect method is from Eq. (38) which appears only at the first time step. If one prefers to use Eq. (39) for $[G_{GD}^{pp}]^{mn}$ and write $([C_{GD}]+[G_{GD}^{pu}]^{\mathrm{T}})[G_{GD}^{uu}]^{-1}$ as a new matrix $[A]$, the additional cost increment for Eq. (39) at each time step is the multiplication of matrix $[A]$ with a vector (refer to Eq. (37)). Compared with the huge amount of matrix manipulations for time-domain problems, such an increment is usually not significant. With Eqs. (38) and (39), an existing time domain AGBEM computer code can be easily modified to a SGBEM one.

The same with steady-state problems, time domain SGBEM formulation can be established through Eqs. (35) and (36): Equation (35) is applied to all those sections of the boundary where displacement components are prescribed (Γ_1), while Eq. (36) isapplied to the sections where traction components are prescribed (Γ_2). For the case of BEM/FEM coupling procedure, both Eqs. (35) and (36) should be used on the interface.

3 Numerical Examples

Two numerical examples for time-domain potential problems are given in this paper. Since SGBEM cannot guarantee the stability of the dynamic scheme, the linear θ method [17] is used here. Linear time and space interpolation functions are used for both u and p.

3.1 One-Dimensional Rod. The first example is for a finite domain problem as depicted in Fig. 3, which was presented previously by Mansur [16]. It consists of a one-dimensional rod fixed at one extremity and subjected to a Heaviside-type force function applied at the other extremity. The displacements **u** are assumed to be zero at $x=0$ and the tractions **p** are also taken as null at $y=0$ and $y=b$ for any time t. At $x=a$ and $t=0$, a load $P=Ep$ is suddenly applied and kept constant until the end of the analysis (E is the Young's modulus). Twenty-four boundary elements with the same length L_j are used for the numerical model.

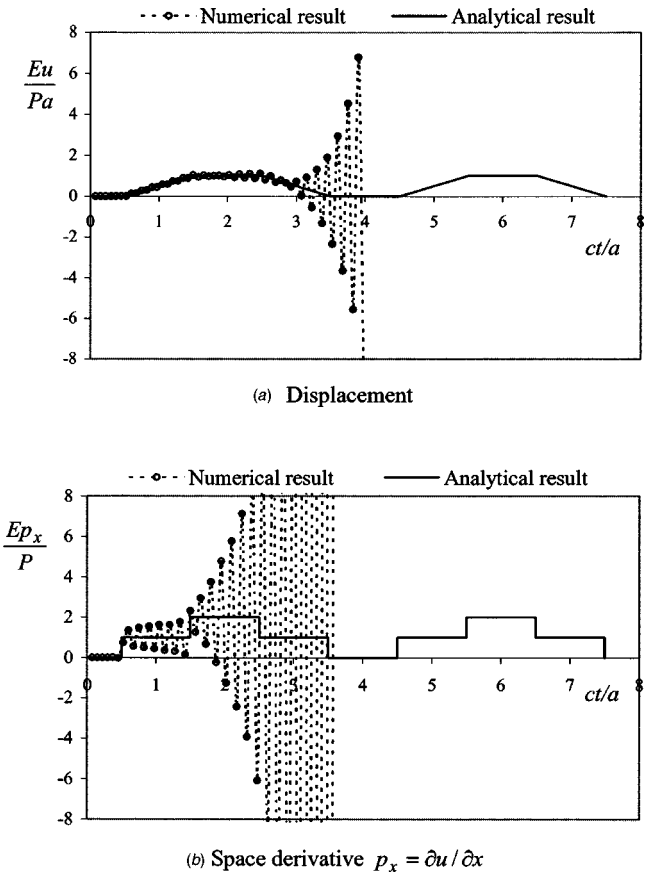
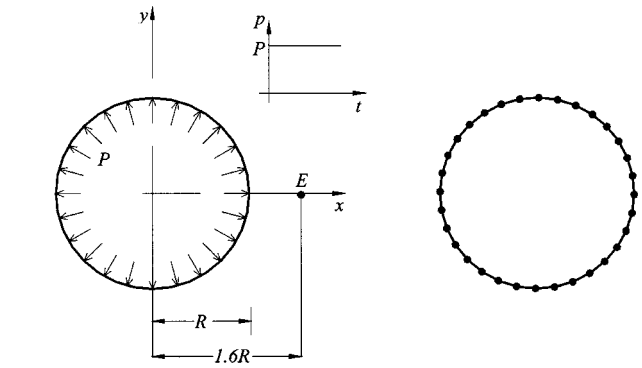


Fig. 5 Results at point $A(a/2,b/2)$ from SGBEM, $\theta=1.0$ and $\beta=0.6$

Comparisons between TCBEM and SGBEM results for different points, with $\theta=1.4$ and $\beta=c\Delta t/L_j=0.6$ (c is the wave velocity, Δt is the time step), are shown in Fig. 4, where one can see only slight differences. This provides a numerical proof to the relationship among different kinds of coefficient matrices given inthis paper. When the linear θ method is not used (i.e., $\theta=1.0$), both displacement and its space derivative at point $A(a/2,b/2)$ exhibit instability as shown in Fig. 5, which means that SGBEM cannot guarantee the stability for the dynamic scheme. Against this background, the linear θ method should be used.

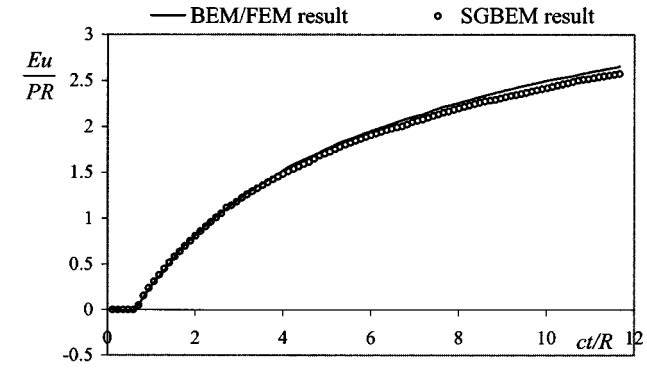
3.2 Two-Dimensional Cavity. In order to show the effectiveness of the proposed method for infinite domain problems, Fig. 6 gives a 2D cylindrical cavity of radius R in an infinite space. At $t=0$, a boundary flux p is suddenly applied and kept constant until the end of the analysis. Thirty-two boundary elements with the same length l_j are used in the modeling.

Time histories of displacement and its space derivative at the internal point E, obtained from the BEM/FEM coupling formulation [18] and from the indirect SGBEM formulation ($\theta=1.4$ and $\beta=0.6$), are show in Figs. 7a and 7b, respectively. From Fig. 7, one can observe the good accord.

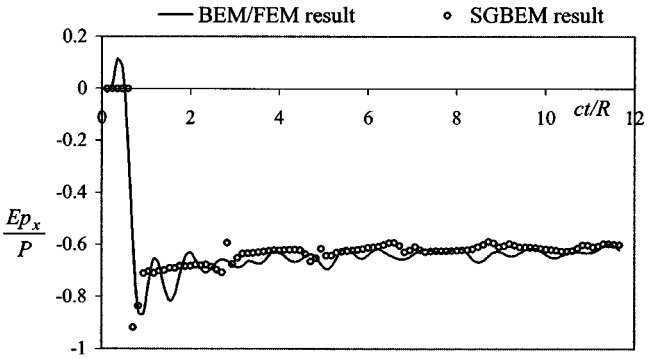


(*a*) Geometry definition and boundary conditions (*b*) Discretization

Fig. 6 2D cavity problem



(*a*) Displacement



(*b*) Space derivative $p_x = \partial u / \partial x$

Fig. 7 Results at point $E(1.6R,0)$ from SGBEM and BEM/FEM coupling method, $\theta=1.4$ and $\beta=0.6$

4 Conclusions

The relationship among the three kinds of coefficient matrices is given in this paper, through which hypersingularity is avoided for SGBEM. With the indirect method, it is very easy to modify an existing AGBEM computer code to a SGBEM one. There is no limitation for the application of the indirect method. It can be used to get symmetric BEM/FEM coupling procedure, as normally FE domain has more nodes than BE domain. It can also be used to all kinds of SGBEM problems, such as elastodynamic problem, heat transfer problem, fluid problem, etc. A similar method can be applied to evaluate the hypersingular integrals related to collocation BEM, or related to boundary stress calculation.

Appendix

As shown in Fig. 2, when the linear spatial interpolation function is used and the integration runs along two adjacent elements (with subscript "1"), strong singularity appears in Eq. (14) as r approaches zero. Although the numerical method can be used for the Galerkin procedure, the analytical method is given here for the inner space integration so as to increase its accuracy, as

$$
\begin{aligned}
(G_{Gij}^{pu})_1 &= \int_0^{l_i}\int_0^{l_j} p^*(X,\xi)\,\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi)\\
&= \int_0^{l_i}\int_0^{l_j}\frac{1}{r}(\mathbf{r}^0\cdot\mathbf{n}^0)\left(1-\frac{\eta_j}{l_j}\right)d\eta_j\left(1-\frac{\eta_j}{l_j}\right)d\eta_i\\
&= \int_0^{l_i}\int_0^{l_j}\frac{\eta_i\sin\alpha}{(\eta_j+\eta_i\cos\alpha)^2+(\eta_i\sin\alpha)^2}\\
&\quad\times\left(1-\frac{\eta_j}{l_j}\right)d\eta_j\left(1-\frac{\eta_i}{l_i}\right)d\eta_i\\
&= \int_0^{l_i}\left\{-\frac{\eta_i\sin\alpha}{2l_j}[\ln(l_j^2+\eta_i^2+2l_j\eta_i\cos\alpha)-2\ln\eta_i]\right.\\
&\quad\left.+\frac{l_j+\eta_i\cos\alpha}{l_j}\left(tg^{-1}\frac{l_j+\eta_i\cos\alpha}{\eta_i\sin\alpha}-tg^{-1}\frac{\cos\alpha}{\sin\alpha}\right)\right\}\\
&\quad\times\left(1-\frac{\eta_i}{l_i}\right)d\eta_i, \qquad (A1)
\end{aligned}
$$

where one can also see that if $\alpha=0$, $(G_{Gij}^{pu})_1=0$. As there is no singularity in the outer space integration in expression (*A*1), the numerical method can be used.

From Eq. (34) and Table 1, one can see that when $t_m=t_n$ and $r=0$, strong singularity appears in $D_I^{mn}(X,\xi)$. Therefore $G_{GDij}^{pumn}=G_{GDij}^{pu}$ can be calculated analytically, as

$$
\begin{aligned}
(G_{GDij}^{pu})_1 &= \int_0^{l_i}\int_0^{l_j} D_I^{mn}(X,\xi)(\mathbf{r}^0\cdot\mathbf{n}^0)\,\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi)\\
&= -\int_0^{l_i}\int_0^{l_j}\frac{A_1A_2}{r}H[c\Delta t-r](\mathbf{r}^0\cdot\mathbf{n}^0)\\
&\quad\times\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi)\\
&= -\int_0^{l_i}\int_0^{l_j}\left(\frac{r}{A_1A_2+A_5}-\frac{c\Delta t}{r}\right)H[c\Delta t-r](\mathbf{r}^0\cdot\mathbf{n}^0)\\
&\quad\times\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi)\\
&= -\int_0^{l_i}\int_0^{l_j}\frac{r}{A_1A_2+A_5}H[c\Delta t-r](\mathbf{r}^0\cdot\mathbf{n}^0)\,\eta_j(X)d\Gamma(X)\\
&\quad\times\eta_i(\xi)d\Gamma(\xi)+\int_0^{l_i}\int_0^{l_j}\frac{c\Delta t}{r}H[c\Delta t-r](\mathbf{r}^0\cdot\mathbf{n}^0)\\
&\quad\times\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi)=D_1+D_2, \qquad (A2)
\end{aligned}
$$

where $\Delta t=t_n-t_{n-1}$, $A_1=\sqrt{c\Delta t-r}$, $A_2=\sqrt{c\Delta t+r}$, and $A_5=c\Delta t$. While there is no singularity in D_1, the strongly singular integration in D_2 can be calculated analytically by a similar way with Eq. (*A*1), but special attention should be given to the Heaviside function,

$$D_2 = \int_0^{l_i}\int_0^{l_j} \frac{c\Delta t}{r} H[c\Delta t - r](\mathbf{r}^0 \cdot \mathbf{n}^0)\,\eta_j(X)d\Gamma(X)\,\eta_i(\xi)d\Gamma(\xi)$$

$$= \int_0^{l_i}\int_0^{l_{0j}} \frac{c\Delta t\,\eta_i \sin\alpha}{(\eta_j + \eta_i\cos\alpha)^2 + (\eta_i \sin\alpha)^2} \times \left(1 - \frac{\eta_j}{l_j}\right) d\eta_j \left(1 - \frac{\eta_i}{l_i}\right) d\eta_i$$

$$= c\Delta t \int_0^{l_i} \left\{ -\frac{\eta_i \sin\alpha}{2l_j}[\ln(l_{0j}^2 + \eta_i^2 + 2l_{0j}\eta_i\cos\alpha) - 2\ln\eta_i] + \frac{l_j + \eta_i\cos\alpha}{l_j}\left(tg^{-1}\frac{l_{0j} + \eta_i\cos\alpha}{\eta_i\sin\alpha} - tg^{-1}\frac{\cos\alpha}{\sin\alpha}\right)\right\} \times \left(1 - \frac{\eta_i}{l_i}\right) d\eta_i, \quad (A3)$$

where $l_{0j} = \min\{l_j, \sqrt{(c\Delta t)^2 - (\eta_i \sin\alpha)^2} - \eta_i\cos\alpha\}$.

As there is no singularity in the outer space integration in Eq. (*A*3), the numerical method can be used.

References

[1] Sirtori, S., 1979, ''General Stress Analysis Method by Means of Integral Equations and Boundary Elements,'' Meccanica, **14**, pp. 210–218.

[2] Bonnet, M., Maier, G., and Polizzotto, C., 1998, ''Symmetric Galerkin Boundary Element Methods,'' Appl. Mech. Rev., **51**(11), pp. 669–704.

[3] Carini, A., Diligent, M., Maranesi, P., and Zanella, M., 1999, ''Analytical Integrations for Two-Dimensional Elastic Analysis by the Symmetric Galerkin Boundary Element Method,'' Comput. Mech., **23**, pp. 308–323.

[4] Krishnasamy, G., Rizzo, F. J., and Rudolphi, T. J., 1991, ''Hypersingular Boundary Integral Equations: Their Occurrence, Interpretation, Regularization and Computation,'' *Developments in Boundary Element Methods*, Vol. 7, P. K. Banerjee, and S. Kobayashi, eds., Elsevier Applied Science Publishers, New York.

[5] Dominguez, J., Ariza, M. P., and Gallego, R., 2000, ''Flux and Traction Boundary Elements without Hypersingular or Strongly Singular Integrals,'' Int. J. Numer. Methods Eng., **48**, pp. 111–135.

[6] Tanaka, M., Sladek, V., and Sladek, J., 1994, ''Regularization Techniques Applied to Boundary Element Methods,'' Appl. Mech. Rev., **47**(10), pp. 457–499.

[7] Hadamard, J., 1923, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, CT.

[8] Monegato, G., 1994, ''The Numerical Evaluation of Hypersingular Integrals,'' J. Comput. Appl. Math., **50**, pp. 9–31.

[9] Aimi, A., Carini, A., Diligenti, M., and Monegato, G., 1998, ''Numerical Integration Schemes for Evaluation of (Hyper) singular Integrals in 2D BEM,'' Comput. Mech., **22**, pp. 1–11.

[10] Gray, L. J., Martha, L. F., and Ingraffea, A. R., 1990, ''Hypersingular Integrals in Boundary Element Fracture Analysis,'' Int. J. Numer. Methods Eng., **29**, pp. 1135–1158.

[11] Guiggiani, M., 1994, ''Hypersingular Formulation for Boundary Stress Evaluation,'' Eng. Anal. Boundary Elem., **13**, pp. 169–179.

[12] Guiggiani, M., 1995, ''Hypersingular Boundary Integral Equations Have an Additional Free Term,'' Comput. Mech., **16**, pp. 245–248.

[13] Gallego, R., and Dominguez, J., 1996, ''Hypersingular BEM for Transient Elastodynamics,'' Int. J. Numer. Methods Eng., **39**, pp. 1681–1705.

[14] Carini, A., Diligenti, M., and Salvadori, A., 1999, ''Implementation of a Symmetric Boundary Element Method in Transient Heat Conduction with Semi-Analytical Integrations,'' Int. J. Numer. Methods Eng., **46**, pp. 1819–1843.

[15] Yu, G. Y., Mansur, W. J., Carrer, J. A. M., and Gong, L., 2000, ''Stability of Galerkin and Collocation Time Domain Boundary Element Methods as Applied to Scalar Wave Equation,'' Comput. Struct., **74**, pp. 495–506.

[16] Mansur, W. J., 1983, ''A Time-Stepping Technique to Solve Wave Propagation Problems Using the Boundary Element Method,'' Ph.D. thesis, University of Southampton.

[17] Yu, G. Y., Mansur, W. J., Carrer, J. A. M., and Gong, L., 1998, ''A Linear θ Method Applied to 2D Time-Domain BEM,'' Commun. Numer. Methods Eng., **14**(12), pp. 1171–1180.

[18] Mansur, W. J., Yu, G. Y., Carrer, J. A. M., Lie, S. T., and Siquiera, E. F. N., 2000, ''The θ Scheme for Time-Domain BEM/FEM Coupling Applied to the 2-D Scalar Wave Equation,'' Commun. Numer. Methods Eng., **16**, pp. 439–448.

Intensity of Singular Stress Fields at the End of a Cylindrical Inclusion

N.-A. Noda

T. Genkai

Q. Wang

Mechanical Engineering Department,
Kyushu Institute of Technology,
Kitakyushu 804-8550, Japan

In short fiber reinforced composite it is known that the singular stress at the end of fibers causes crack initiation, propagation, and final failure. The singular stress field is controlled by the generalized stress intensity factors defined at the end of the inclusion. In this study the stress intensity factors are discussed for an elastic cylindrical inclusion in an infinite body under (A) asymmetric uniaxial tension in the x direction, and (B) symmetric uniaxial tension in the z direction. These problems are formulated as a system of integral equations with Cauchy-type or logarithmic-type singularities, where densities of body force distributed in infinite bodies having the same elastic constants as those of the matrix and inclusion are unknown. In the numerical analysis, the unknown body force densities are expressed as fundamental density functions and weight functions. Here, fundamental density functions are chosen to express the symmetric and skew-symmetric stress singularities. Then, the singular stress fields at the end of a cylindrical inclusion are discussed with varying the fiber length and elastic ratio. The results are compared with the ones of a rectangular inclusion under longitudinal and transverse tension.
[DOI: 10.1115/1.1598479]

1 Introduction

In short fiber reinforced composite it is known that the singular stress at the end of fibers causes crack initiation, propagation, and final failure. Recently, Chen and Nisitani [1,2] indicated that the singular stress field is controlled by the generalized stress intensity factors (SIF's) defined for inclusion corners, and Chen [3,4] discussed the singular stress of a rectangular inclusion as a two-dimensional (2D) model. Since actual fibers always have 3D shapes and dimensions, 3D analysis is necessary to evaluate the strength of composites. From this viewpoint, a cylindrical inclusion is important as a 3D model of a fiber in matrix. In previous studies, Kasano et al. [7] treated a rigid cylindrical inclusion, and Hasegawa and Yoshiya [8] solved an elastic cylindrical inclusion with rounded ends. Takao et al. [9], Hasegawa et al. [10], and Wu and Du [11,12] discussed stress fields induced by uniform eigenstrain given within a cylindrical domain. Usually to obtain generalized SIF's for 3D problems is more difficult than to obtain normal SIF's defined for ordinal cracks. Therefore, when Chen and Nisitani [3–5] applied the body force method [6] to 2D problems, they examined two types of numerical procedures, one of which is obtained from the values of unknown body force densities, and the other of which is from the values of stress around the inclusion corner. In both procedures the final results were extrapolated from the results of finite numbers of collocation points [4–6].

In this study the stress intensity factors are discussed for an elastic cylindrical inclusion in an infinite body under (A) asymmetric uniaxial tension in the x direction, and (B) symmetric uniaxial tension in the z direction. The asymmetric problem (A) is solved on the superposition of two auxiliary loads; (i) biaxial tension and (ii) plane state of pure shear. Those problems are formulated as a system of integral equations with Cauchy-type or logarithmic-type singularities, where densities of body force distributed in infinite bodies having the same elastic constants as those of the matrix and inclusion are unknown. In the numerical analysis, the unknown functions of the body force densities are expressed as fundamental density functions and weight functions. Here, the weight functions are approximated as power series instead of step or linear functions used usually in the body force method [3–6]. Then, the singular stress fields at the end of a cylindrical inclusion are discussed for the wide range of the fiber length and elastic ratio.

2 Theory and Solution

Consider a cylindrical bar in an infinite body under asymmetric uniaxial tension in the x direction shown in Fig. 1*a*. This problem is composed of the superposition of two auxiliary loads; biaxial tension in the xy plane, shown in Fig. 1*c*, and pure shear in the xy plane, shown in Fig. 1*d*. The method of analysis will be explained for the problem of pure shear in the xy plane. Here, L and D are dimensions of the inclusion, and σ^{∞} is a stress at infinity. The notations (G_M, ν_M) denote the shear modulus and Poisson's ratio of the matrix, and (G_I, ν_I) denotes the ones of the inclusion. Rectangular coordinate (x,y,z) and cylindrical coordinate (r,θ,z) are defined in Fig. 1. Here, (ξ,η,ζ), (ρ,ϕ,ζ) are rectangular and cylindrical coordinates where body forces are applied. The body force method [6] is used to formulate the problem as a system of singular integral equations. Here, the fundamental solutions are stress and displacement fields when two ring forces acting symmetrically to the plane $z=0$. In this case the boundary conditions only on $z \geqslant 0$ can be considered due to symmetry. The two ring forces have three types [13,14], that is, (1) ring forces in the r direction with the magnitude of $\cos 2\phi$, (2) ring forces in the θ direction with the magnitude of $\sin 2\phi$, (3) ring forces in the z direction with the magnitude of $\cos 2\phi$. In the following discussion, how to satisfy the boundary conditions around corner A will be explained.

The problem can be expressed by the following equations in terms of the unknown body force densities $(F_{nM}, F_{tM}, F_{\theta M})$ and $(F_{nI}, F_{tI}, F_{\theta I})$ distributed at infinitesimal area $\rho d\theta dr_A$ in infinite bodies M and I. Here, bodies M and I have the same elastic constants as those of the matrix and inclusion, respectively, and r_A is a distance from the corner A as shown in Fig. 1*e*. In the following equation, the notation $\int_L$ means integrating body forces on both the side and ends of the cylinder. Here, for example, the notation

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, July 30, 1999; final revision, Jan. 16, 2003. Associate Editor: J. W. Ju. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

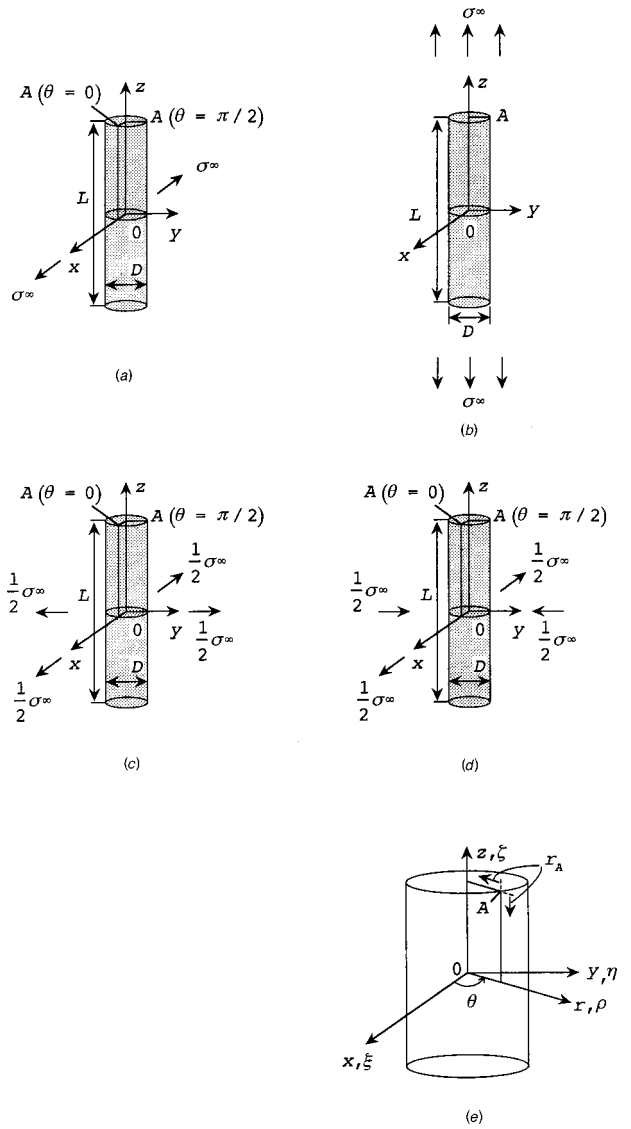


Fig. 1 Problem and coordinate system: (a) Uniaxial tension perpendicular to the axis of the inclusion (*x* direction); (b) Uniaxial tension in the axis of the inclusion (*z* direction); (c) Hydrostatic tension in a plane perpendicular to the axis of the inclusion (*xy* plane); (d) Pure shear in a plane perpendicular to the axis of the inclusion (*xy* plane); (e) Coordinate system

$h_{nn}^{F_{nM}}(r_A,s)$ denotes the normal stress induced at the collocation point s induced by the body force F_{nM} acting at the point r_A. Since the integral includes the Cauchy-type or logarithmic-type singularities, the integration should be interpreted in the Cauchy principal value sense,

$$-\frac{1}{2}F_{nM}(s)-\frac{1}{2}F_{nI}(s)+\int_L h_{nn}^{F_{nM}}(r_A,s)F_{nM}(r_A)dr_A + \int_L h_{nn}^{F_{tM}}(r_A,s)F_{tM}(r_A)dr_A+\int_L h_{nn}^{F_{\theta M}}(r_A,s)F_{\theta M}(r_A)dr_A - \int_L h_{nn}^{F_{nI}}(r_A,s)F_{nI}(r_A)dr_A-\int_L h_{nn}^{F_{tI}}(r_A,s)F_{tI}(r_A)dr_A - \int_L h_{nn}^{F_{\theta I}}(r_A,s)F_{\theta I}(r_A)dr_A = -\{\sigma_{nM}^{\infty}(s)-\sigma_{nI}^{\infty}(s)\}\cos 2\theta, \quad (1a)$$

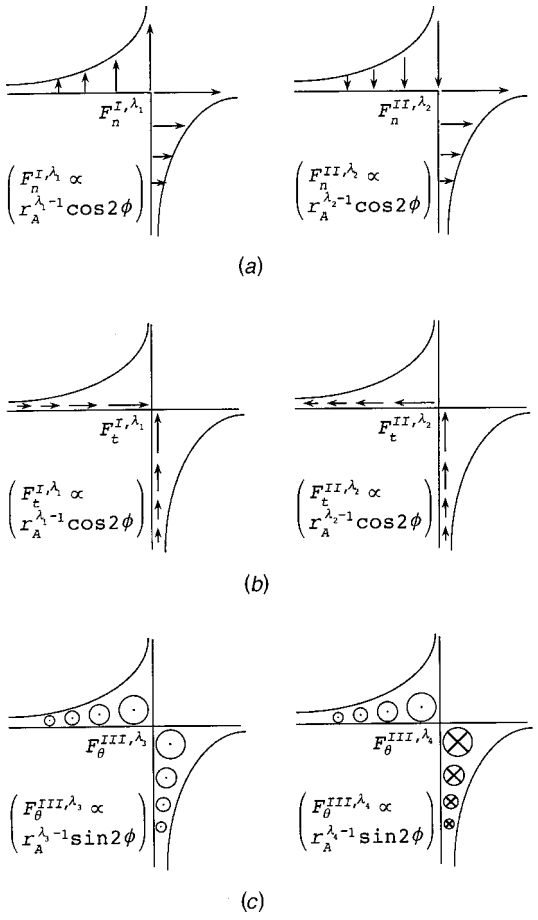
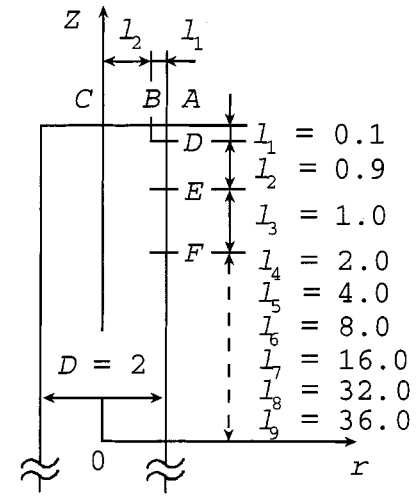


Fig. 2 Two types of body force distributed around the corner in the (*a*) normal, (*b*) tangential, and (*c*) circumferential directions

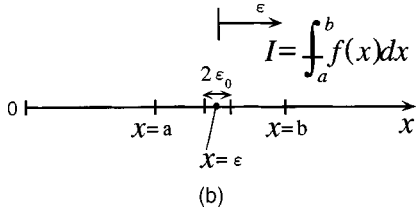
$$\int_L h_v^{F_{nM}}(r_A,s)F_{nM}(r_A)dr_A+\int_L h_v^{F_{tM}}(r_A,s)F_{tM}(r_A)dr_A + \int_L h_v^{F_{\theta M}}(r_A,s)F_{\theta M}(r_A)dr_A-\int_L h_v^{F_{nI}}(r_A,s)F_{nI}(r_A)dr_A - \int_L h_v^{F_{tI}}(r_A,s)F_{tI}(r_A)dr_A-\int_L h_v^{F_{\theta I}}(r_A,s)F_{\theta I}(r_A)dr_A = -(u_{\theta M}^{\infty}-u_{\theta I}^{\infty})\sin 2\theta. \quad (1b)$$

Equations (1*a*) and (1*b*) enforce the boundary conditions $\sigma_{nM}-\sigma_{nI}=0$, and $u_{\theta M}-u_{\theta I}=0$, respectively; other boundary conditions, that is, $\tau_{ntM}-\tau_{ntI}=0$, $\tau_{n\theta M}-\tau_{n\theta I}=0$, $u_{rM}-u_{rI}=0$, $u_{zM}-u_{zI}=0$, can be expressed in a similar way. In Eqs. (1*a*) and (1*b*) σ_{nM}^{∞} and $u_{\theta M}^{\infty}$ are stress and displacement components induced by the stress at infinity σ^{∞}; here we assume body *I* is also under the stresses $(\sigma_{rI}^{\infty},\sigma_{\theta I}^{\infty},\tau_{r\theta I}^{\infty})$, which induce σ_{nI}^{∞}, and $u_{\theta I}^{\infty}$ in Eqs. (1*a*) and (1*b*).

The singular stress fields near corner A can be expressed by two types of body force distributions; symmetric and skew-symmetric types to the bisector of the corners. Figure 2 indicates the two types of body forces distributed in the normal, tangential, and circumferential directions to the boundary. In the vicinity of corner A, plain strain condition can be assumed; then, the eigenvalues λ_1, λ_2, λ_3, λ_4 controlling the singular stress fields is determined from the eigenequations for 2D problems [1,2,15]. In this study



(a)



(b)

Fig. 3 (*a*) Typical boundary division for Eqs. (3) and (4). (*b*) Boundary division for singular integrals.

therefore unknown body force densities $F_{n,M}(r_A)\sim F_{\theta,M}(r_A)$ in body M and $F_{n,I}(r_A)\sim F_{\theta,I}(r_A)$ in body I are approximated by using fundamental densities $r_A^{\lambda_1-1}\sim r_A^{\lambda_4-1}$ and weight functions $W_{n,M}\sim W_{\theta,I}^{III,\lambda_4}$, where unknown constants are $a_{n,M}\sim f_{n,I}$,

$$F_{n,j}(r_A)=F_{n,j}^{I,\lambda_1}(r_A)+F_{n,j}^{II,\lambda_2}(r_A)=W_{n,j}^{I}(r_A)r_A^{\lambda_1-1}+W_{n,j}^{II}(r_A)r_A^{\lambda_2-1}$$
$$F_{t,j}(r_A)=F_{t,j}^{I,\lambda_1}(r_A)+F_{t,j}^{II,\lambda_2}(r_A)=W_{t,j}^{I}(r_A)r_A^{\lambda_1-1}+W_{t,j}^{II}(r_A)r_A^{\lambda_2-1} \quad (2)$$

$$F_{\theta,j}(r_A)=F_{\theta,j}^{III,\lambda_3}(r_A)+F_{\theta,j}^{III,\lambda_4}(r_A)=W_{\theta,j}^{III,\lambda_3}(r_A)r_A^{\lambda_3-1}+W_{\theta,j}^{III,\lambda_4}(r_A)r_A^{\lambda_4-1} \quad (j=M,I),$$

$$W_{n,j}^{I,\lambda_1}(r_A)=\sum_{n=1}^{M} a_{n,j}r_A^{n-1},$$

$$W_{t,j}^{I,\lambda_1}(r_A)=\sum_{n=1}^{M} b_{n,j}r_A^{n-1}, \quad W_{\theta,j}^{III,\lambda_3}(r_A)=\sum_{n=1}^{M} c_{n,j}r_A^{n-1}, \quad (3)$$

$$W_{n,j}^{II,\lambda_2}(r_A)=\sum_{n=1}^{M} d_{n,j}r_A^{n-1},$$

$$W_{t,j}^{II,\lambda_2}(r_A)=\sum_{n=1}^{M} e_{n,j}r_A^{n-1},$$

$$W_{\theta,j}^{III,\lambda_4}(r_A)=\sum_{n=1}^{M} f_{n,j}r_A^{n-1} \quad (j=M,I).$$

Here, fundamental density functions are chosen to express the symmetric stress singularity of the forms $1/r^{1-\lambda_1}$, $1/r^{1-\lambda_3}$ and the skew-symmetric stress singularity of the forms $1/r^{1-\lambda_2}$, $1/r^{1-\lambda_4}$. Equations (2) and (3) do not include the terms expressing local uniform streching and shear distortion at corner A. Therefore the stresses $(\sigma_{rI}^{\infty},\sigma_{\theta I}^{\infty},\tau_{r\theta I}^{\infty})$ applied in body I. In the numerical calculation, we may set the values of $(\sigma_{rI}^{\infty},\sigma_{\theta I}^{\infty},\tau_{r\theta I}^{\infty})$ in body I so as to produce the same strains appearing in body M under the stress σ^{∞}. The eigenvalues λ_1, λ_2 are given as the roots of the eigenequations for in-plane deformation [15,2], and the eigenvalues λ_3, λ_4 are given as the roots of the eigenequations for antiplane deformation [1]. For example, the eigenequations for antiplane deformation are shown in Eqs. (4) and (5).

Table 1 Convergence of F_{I,λ_1}, F_{II,λ_2}, F_{III,λ_3} ($L/D=10$, $G_I/G_M=10^2$, $\nu_I=\nu_M=0.3$)

M	F_{I,λ_1}			F_{II,λ_2}			F_{III,λ_3}
	from $W_{nM}^{I}(0)$	from $W_{tM}^{I}(0)$	Average	from $W_{nM}^{II}(0)$	from $W_{tM}^{II}(0)$	Average	from $W_{\theta M}^{III}(0)$
2	0.290	0.417	0.354	-0.723	-0.763	-0.743	0.070
3	0.321	0.396	0.359	-0.792	-0.795	-0.794	0.027
4	0.333	0.388	0.361	-0.790	-0.789	-0.789	0.019
5	0.338	0.383	0.360	-0.788	-0.788	-0.788	0.014
6	0.340	0.379	0.360	-0.789	-0.788	-0.789	0.011

For a symmetric state of stress singularity,

$$\frac{\sin\lambda(\gamma-\pi)}{\sin\lambda\pi}=+\frac{\Gamma+1}{\Gamma-1}. \quad (4)$$

For a skew-symmetric state of stress singularity,

$$\frac{\sin\lambda(\gamma-\pi)}{\sin\lambda\pi}=-\frac{\Gamma+1}{\Gamma-1}, \quad (5)$$

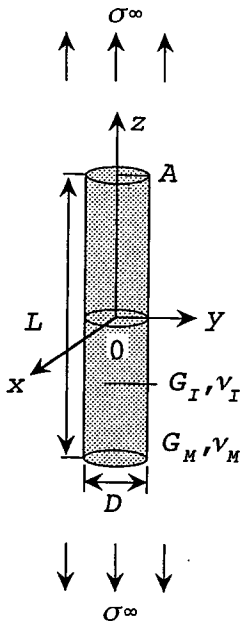
where the corner angle for matrix $\gamma=3\pi/2$ and the elastic ratio $\Gamma=G_M/G_I$. When $G_M>G_I$, there is a real root λ_3 of Eqs. (4), but no roots λ_4 of Eqs. (5). On the contrary, when $G_M<G_I$, there is a real root λ_4 of Eqs. (4), but no roots λ_3 of Eqs. (5).

Figure 3*a* illustrates an example of boundary divisions for $L/D=10^2$. In the numerical solutions for elliptical inclusions, we do not have to divide the boundaries because the "fundamental densities" to express an elliptical inclusion exactly are available [16]. On the other hand, the boundary division is introduced here because in this problem the fundamental densities are only useful near the corner. Then, the fundamental densities with singularities of symmetric and skew-symmetric types are employed on boundaries C-B-A-D-E in Fig. 3. It is confirmed that the results are not affected until almost to the third digit by changing a region over which the fundamental densities are used. Except along the boundaries C-B-A-D-E in Fig. 3, body forces are simply distributed in the normal, circumferential, and tangential directions without using symmetric and skew-symmetric distributions. On the numerical solution as shown in Eqs. (2) and (3), the singular integral Eqs. (1*a*) and (1*b*) are reduced to algebraic equations for the determination of the unknown coefficients, for example, $a_{n,M}\sim f_{n,I}$ in Eq. (3). These coefficients are determined from the boundary conditions at suitably chosen collocation points. The

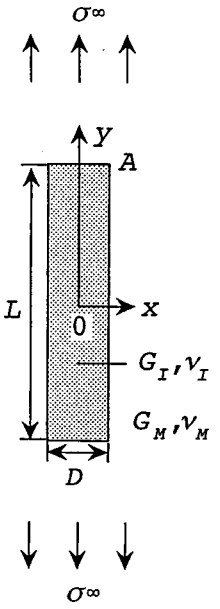
Table 2 Convergence of F_{I,λ_1}, F_{II,λ_2}, F_{III,λ_4} ($L/D=10$, $G_I/G_M=10^{-5}$, $\nu_I=\nu_M=0.3$)

M	F_{I,λ_1}			F_{II,λ_2}			F_{III,λ_4}
	from $W_{nM}^{I}(0)$	from $W_{tM}^{I}(0)$	Average	from $W_{nM}^{II}(0)$	from $W_{tM}^{II}(0)$	Average	from $W_{\theta M}^{III}(0)$
2	0.191	0.221	0.206	1.260	0.989	1.125	0.044
3	0.212	0.227	0.220	1.123	1.024	1.074	0.021
4	0.214	0.223	0.219	1.058	1.051	1.055	0.014
5	0.215	0.222	0.219	1.060	1.048	1.054	0.010
6	0.216	0.221	0.219	1.056	1.047	1.052	0.008

L/D \ G_I/G_M	F_{I,λ_1}						F_{II,λ_2}					
	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}
$10^{0.0}=1.0$	0.393	0.374	0.281	0.219	0.292	0.322	1.652	1.687	1.935	3.000	0.624	0.549
$10^{0.5}=3.162$	0.382	0.366	0.278	0.249	0.437	0.659	1.616	1.641	1.877	2.776	0.891	0.959
$10^{1.0}=10.0$	0.379	0.363	0.277	0.255	0.599	1.298	1.604	1.627	1.862	2.780	1.087	1.816
$10^{1.5}=31.62$	0.379	0.363	0.276	0.255	0.618	1.982	1.602	1.623	1.860	2.780	1.120	2.692
$10^{2.0}=100.0$	0.379	0.363	0.276	0.255	0.618	2.113	1.600	1.624	1.860	2.780	1.121	2.891
$10^{2.5}=316.2$	0.379	0.363	0.276	0.255	0.618	2.099	1.600	1.624	1.859	2.780	1.119	2.891
$10^{3.0}=1000$	0.379	0.362	0.276	0.255	0.618	2.118	1.600	1.631	1.860	2.780	1.119	2.892



L/D \ G_I/G_M	F_{I,λ_1}						F_{II,λ_2}					
	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}
$10^{0.0}=1.0$	0.591	0.555	0.368	0.198	0.230	0.244	2.208	2.226	2.415	2.386	0.531	0.439
$10^{0.5}=3.162$	0.540	0.511	0.359	0.240	0.359	0.413	2.044	2.060	2.233	2.581	0.707	0.646
$10^{1.0}=10.0$	0.513	0.489	0.351	0.258	0.495	0.673	1.968	1.979	2.135	2.741	0.944	1.018
$10^{1.5}=31.62$	0.502	0.484	0.348	0.264	0.593	1.064	1.938	1.944	2.099	2.804	1.137	1.619
$10^{2.0}=100.0$	0.498	0.484	0.347	0.265	0.635	1.544	1.940	1.930	2.083	2.825	1.224	2.377
$10^{2.5}=316.2$	0.497	0.483	0.346	0.266	0.649	1.893	1.924	1.931	2.084	2.831	1.248	2.933
$10^{3.0}=1000$	0.493	0.483	0.346	0.266	0.651	1.950	1.924	1.926	2.081	2.832	1.255	2.977



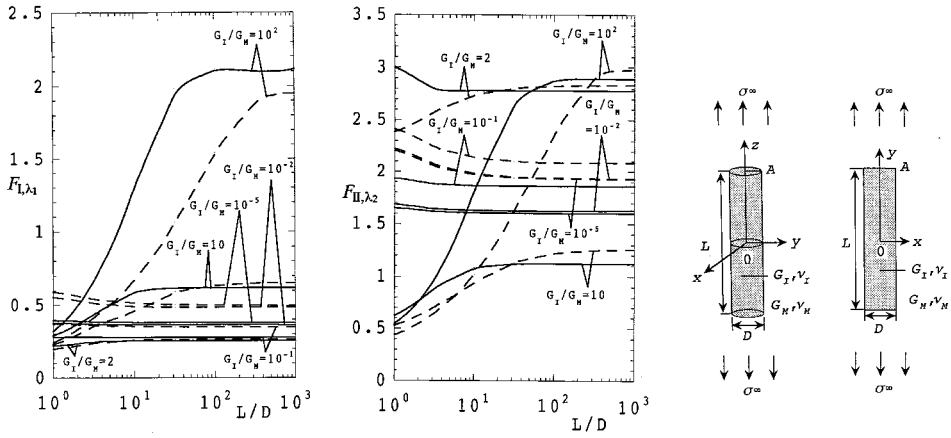


Fig. 4 F_{I,λ_1} **and** F_{II,λ_2} **for a cylindrical inclusion (solid line) and a rectangular inclusion (broken line) under longitudinal tension (**$\nu_M=\nu_I=0.3$**)**

stress intensity factors K_{I,λ_1}, K_{II,λ_2}, K_{III,λ_3}, K_{III,λ_4} for corner A can be obtained from the values of weight functions at the corner tip $W_n^I(0)$, $W_l^I(0)$, $W_n^{II}(0)$, $W_l^{II}(0)$, $W_\theta^{III,\lambda_3}(0)$, $W_\theta^{III,\lambda_4}(0)$. The expressions may be found in Nisitani et al. ([5]), and Noda et al. ([17]).

3 Numerical Results and Discussion

The generalized stress intensity factors K_{I,λ_1}, K_{II,λ_2}, K_{III,λ_3}, K_{III,λ_4} defined in Refs. [1], [2] are analyzed with varying the aspect ratio L/D and elastic ratio G_I/G_M. In the following discussion, dimensionless stress intensity factors $F_{I,\lambda_1}\sim F_{III,\lambda_4}$ are shown assuming Poisson's ratio $\nu_I=\nu_M=0.3$,

$$\begin{aligned} F_{I,\lambda_1}&=K_{I,\lambda_1}/\sigma^\infty\sqrt{\pi}(D/2)^{1-\lambda_1},\\ F_{II,\lambda_2}&=K_{II,\lambda_2}/\sigma^\infty\sqrt{\pi}(D/2)^{1-\lambda_2},\\ F_{III,\lambda_3}&=K_{III,\lambda_3}/\sigma^\infty\sqrt{\pi}(D/2)^{1-\lambda_3},\\ F_{III,\lambda_4}&=K_{III,\lambda_4}/\sigma^\infty\sqrt{\pi}(D/2)^{1-\lambda_4}. \end{aligned} \tag{6}$$

Convergence of the results are shown for the problem of pure shear in the xy plane in Fig. 1*d*. Table 1 shows the values of F_{I,λ_1}, F_{II,λ_2} at the corner $A(\theta=0)$ and the values of F_{III,λ_3} at the corner $A(\theta=\pi/4)$ for $L/D=10$, $G_I/G_M=10^2$. Also Table 2 shows the values of F_{I,λ_1}, F_{II,λ_2} at the corner $A(\theta=0)$ and the values of F_{III,λ_4} at the corner $A(\theta=\pi/4)$ for $L/D=10$, $G_I/G_M=10^{-5}$. In Tables 1 and 2, M is the number of collocation points at each boundary division, and the total number of collocation points is $7M$. As shown in Tables 1 and 2, The F_{I,λ_1} values can be determined from the values $W_{nM}^I(0)$, $W_{tM}^I(0)$, and the F_{II,λ_2} values can be determined from the values of $W_{nM}^{II}(0)$, $W_{tM}^{II}(0)$. From the examination as shown in Tables 1 and 2, we can see the following.

(1) When $G_I/G_M>1$, the difference of the F_{I,λ_1}, F_{II,λ_2} values obtained from different components of unknown functions $W_{nM}^I(0)$, $W_{tM}^I(0)$ is larger, about 10%. When $G_I/G_M<1$, the difference is smaller, about a few percent. Similar tendency was seen in the analysis of a 2D rectangular inclusion [4].

(2) The average values of F_{I,λ_1}, F_{II,λ_2}, which is obtained from different components, always have good convergence for $M=5$, 6 and look reliable.

(3) The values of F_{III,λ_3}, F_{III,λ_4} are only a few percent compared with the values of F_{I,λ_1}, F_{II,λ_2}.

(4) The final results are obtained without using extrapolation because the weight functions are approximated as power series instead of step or linear functions [3–5] used usually in the body force method. The convergence of the present solution is better than the cases of Nisitani [5] and Chen [3,4].

The following values of F_{I,λ_1}, F_{II,λ_2} are obtained confirming the convergence of the average values for various aspect ratio L/D and elastic ratio G_I/G_M.

In Table 3, the results for a cylindrical bar under uniaxial tension in the z direction are shown. For comparison, Table 4 shows the results for a rectangular inclusion under longitudinal tension obtained in the similar way of the present analysis. Chen's results [3] are in good agreement with Table 4. Results of Tables 3 and 4 are plotted in Fig. 4 as a comparison between the 3D and 2D models. With increasing the value L/D, the stress intensity factors increase and finally become saturated. For the same elastic ratio G_I/G_M, 3D and 2D results have a similar tendency with the difference under $\pm 30\%$ in most cases.

Table 5 shows the values of F_{I,λ_1}, F_{II,λ_2} at the corner $A(\theta=0)$, and Table 6 shows the ones at the corner $A(\theta=\pi/2)$ for uniaxial tension of a cylindrical inclusion in the x direction. Figures 5 and 6 are the plots of Tables 5 and 6. As shown in these tables and figures, the stress intensity factors have the largest values at $A(\theta=0)$ in most cases. From the comparison between the results of Figs. 4 and 5, it is found the F_{I,λ_1} values for z-directional tension are one to four times larger than the ones for x-directional tension in most cases, although the F_{II,λ_2} values are in the same order. From Figs. 4 and 5, it may be concluded that the stress intensity factors take saturated values at nearly the same value of L/D; for example, when $G_I/G_M=10^2$, the F_{I,λ_1} values become saturated when $L/D=100$ in Figs. 4 and 5. In Fig. 5, for high G_I/G_M, the stress intensity F_{I,λ_1} decreases with L/D, becomes almost zero at $L/D=10$, and then increases. Usually, F_{I,λ_1} increases with increasing L/D; however, as $L/D\to 1$, the interaction between both ends of the cylinder seems to make F_{I,λ_1} larger. In Fig. 6, when $G_I/G_M=10^2$ the stress intensity F_{I,λ_1} changes in sign as L/D increases, because in two auxiliary problems in Figs. 1*c* and *d*, F_{I,λ_1} depends on L/D differently. In Table 7 and Fig. 7 the results for a rectangular inclusion under transverse tension are shown. The difference between the results for Figs. 5 and 7 is very large, in other words, it seems difficult to use 2D solution to evaluate the 3D results if the load is applied in the transverse direction.

Table 5 F_{I,λ_1} and F_{II,λ_2} for a cylindrical inclusion at the corner A when θ=0 under uniaxial tension in the x direction (ν_I=ν_M=0.3)

G_I / G_M L / D	F_{I,λ_1}						F_{II,λ_2}					
	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}
$10^{0.0}$=1.0	0.300	0.288	0.226	0.205	0.263	0.294	1.402	1.426	1.673	-3.278	-0.724	-0.624
$10^{0.5}$=3.162	0.302	0.290	0.226	0.193	0.206	0.186	1.317	1.342	1.587	-3.391	-0.833	-0.805
$10^{1.0}$=10.0	0.312	0.299	0.230	0.189	0.166	-0.019	1.280	1.305	1.550	-3.431	-0.898	-1.058
$10^{1.5}$=31.62	0.313	0.300	0.231	0.188	0.165	0.572	1.274	1.299	1.544	-3.442	-0.917	-1.314
$10^{2.0}$=100.0	0.313	0.300	0.231	0.188	0.166	0.614	1.274	1.299	1.544	-3.442	-0.918	-1.371
$10^{2.5}$=316.2	0.314	0.300	0.231	0.188	0.164	0.615	1.274	1.298	1.544	-3.443	-0.918	-1.373
$10^{3.0}$=1000	0.314	0.300	0.231	0.188	0.164	0.616	1.274	1.298	1.544	-3.443	-0.918	-1.373

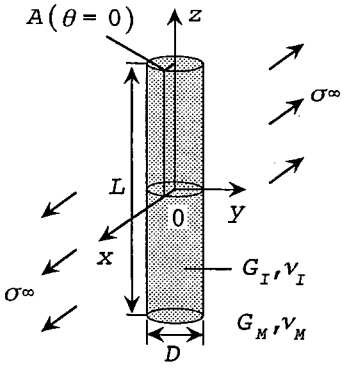
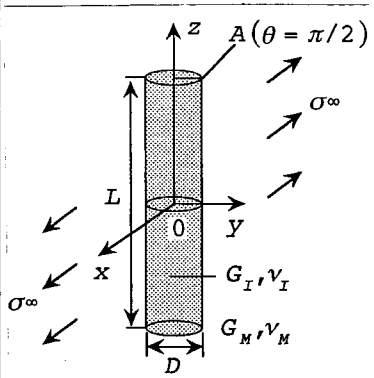


Table 6 F_{I,λ_1} and F_{II,λ_2} for a cylindrical inclusion at corner A when θ=π/2 under uniaxial tension in the x direction

G_I / G_M L / D	F_{I,λ_1}						F_{II,λ_2}					
	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}
$10^{0.0}$=1.0	0.067	0.062	0.033	-0.019	-0.068	-0.091	0.160	0.160	0.164	0.168	0.080	0.069
$10^{0.5}$=3.162	0.083	0.077	0.041	-0.022	-0.103	-0.171	0.241	0.242	0.246	0.217	0.060	-0.018
$10^{1.0}$=10.0	0.093	0.086	0.045	-0.027	-0.145	-0.379	0.228	0.229	0.233	0.191	0.000	-0.269
$10^{1.5}$=31.62	0.095	0.087	0.046	-0.027	-0.146	0.212	0.222	0.223	0.227	0.180	-0.019	-0.525
$10^{2.0}$=100.0	0.095	0.087	0.046	-0.027	-0.145	0.254	0.222	0.223	0.227	0.180	-0.020	-0.582
$10^{2.5}$=316.2	0.095	0.087	0.046	-0.027	-0.146	0.255	0.222	0.222	0.227	0.179	-0.020	-0.584
$10^{3.0}$=1000	0.095	0.087	0.046	-0.027	-0.146	0.256	0.222	0.222	0.227	0.179	-0.020	-0.584



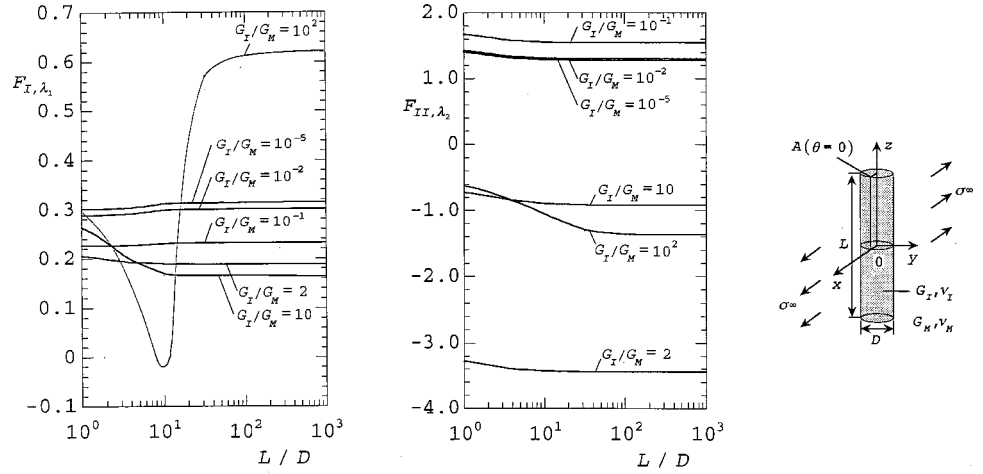


Fig. 5 F_{I,λ_1} and F_{II,λ_2} for a cylindrical inclusion under uniaxial tension in the *x* direction (at corner A with $\theta=0$, $\nu_M=\nu_I=0.3$)

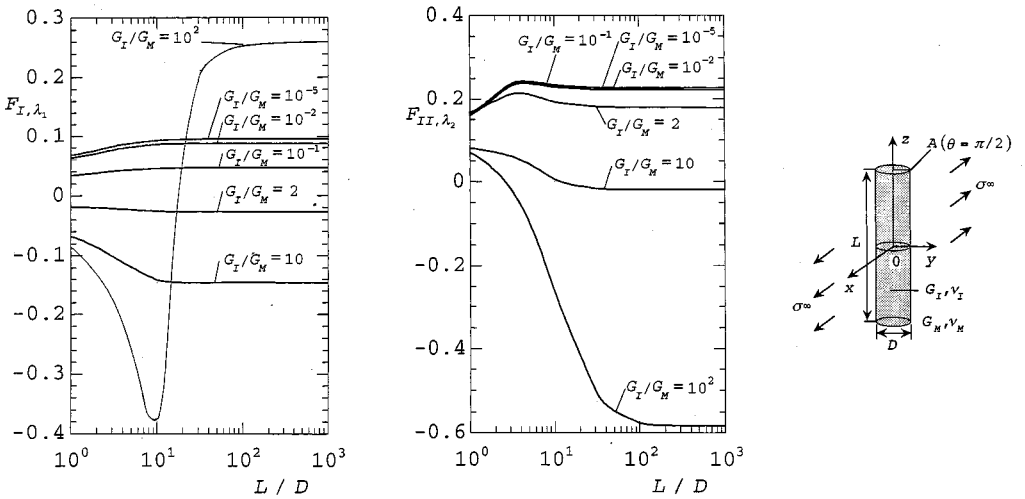


Fig. 6 F_{I,λ_1} and F_{II,λ_2} for a cylindrical inclusion under uniaxial tension in the *x* direction (at corner A with $\theta=\pi/2$, $\nu_M=\nu_I=0.3$)

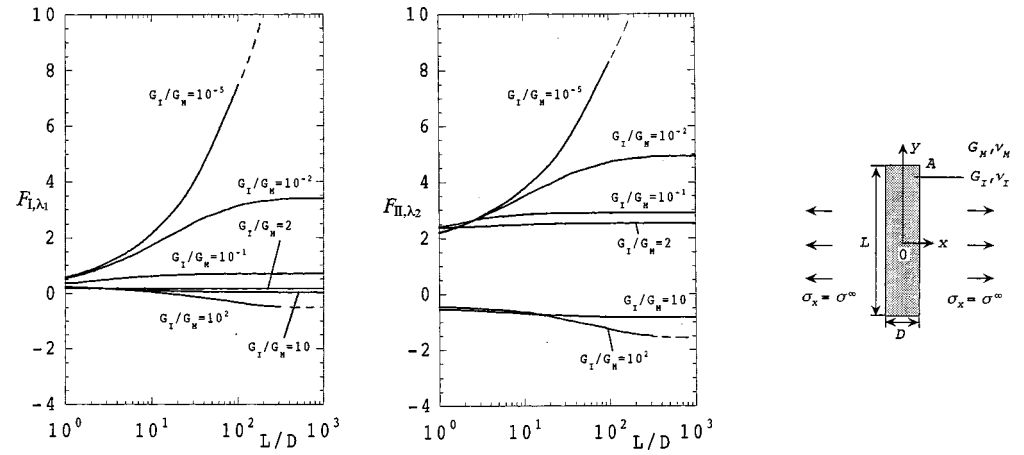


Fig. 7 F_{I,λ_1} and F_{II,λ_2} for a rectangular inclusion under transverse tension, the case of plane strain $\nu_M=\nu_I=0.3$

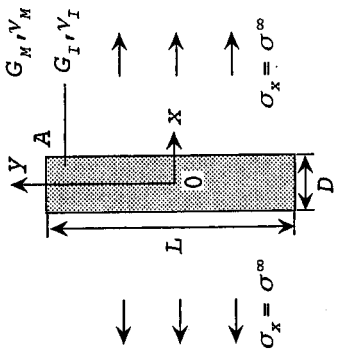


Table 7 F_{I,λ_1} and F_{II,λ_2} for a rectangular inclusion under transverse tension, the case of plane strain $\nu_M = \nu_I = 0.3$

L/D \ G_I/G_M	F_{I,λ_1}						F_{II,λ_2}					
	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}	10^{-5}	10^{-2}	10^{-1}	2	10^{1}	10^{2}
$10^{0.0}=1.0$	0.591	0.555	0.368	0.198	0.230	0.244	2.208	2.226	2.415	2.386	-0.532	-0.439
$10^{0.5}=3.162$	1.124	1.007	0.517	0.176	0.167	0.165	2.798	2.777	2.719	2.427	-0.595	-0.518
$10^{1.0}=10.0$	2.148	1.741	0.628	0.165	0.103	0.051	3.774	3.505	2.856	2.494	-0.696	-0.634
$10^{1.5}=31.62$	4.052	2.595	0.678	0.159	0.057	-0.122	5.423	4.263	2.892	2.524	-0.776	-0.929
$10^{2.0}=100.0$	7.484	3.169	0.692	0.158	0.038	-0.330	8.279	4.738	2.902	2.533	-0.815	-1.254
$10^{2.5}=316.2$	-	3.390	0.697	0.158	0.032	-0.484	-	4.920	2.903	2.532	-0.826	-1.500
$10^{3.0}=1000$	-	3.390	0.697	0.158	0.032	-	-	4.920	2.903	2.536	-0.826	-

4 Conclusion

In this paper, a cylindrical inclusion as a 3D model of a short fiber in composites was analyzed. Using the body force method the problem is formulated as singular integral equations. The generalized stress intensity factors are calculated with varying the aspect ratio L/D and elastic ratio G_I/G_M. The conclusions can be made as follows.

(1) In the numerical solution of the singular integral equations of the body force method, the unknown functions were approximated by the products of the fundamental density functions and the power series along the short segments into which whole boundary is discretized. The convergence of the present solution is better than the cases of Nisitani [5] and Chen [3,4], where the final results are obtained by using extrapolation. The average values of F_{I,λ_1}, F_{II,λ_2}, which is obtained from different components of unknown functions, always have good convergence to the third digit, and look reliable, even for the collocation number of each division $M=5$, 6. The results are shown in the tables and figures.

(2) When the cylindrical inclusion under z-directional tension the F_{I,λ_1} values are one to four times larger than the ones when the cylindrical inclusion under x-directional tension in most cases although the F_{II,λ_2} values are almost in the same order.

(3) From the comparison between the results of a 3D cylindrical inclusion and a 2D rectangular inclusion, it appears though 3D and 2D results have a similar tendency with the difference under $\pm 30\%$ in most cases when the load is applied in the longitudinal direction. However, the difference is very large if the load is applied in the transverse direction. Care should be taken if the 2D solution is applied to evaluate the 3D results.

(4) The values of F_{III,λ_3}, F_{III,λ_4} are only a few percent compared with the values of F_{I,λ_1}, F_{II,λ_2}.

Acknowledgments

This research was supported by the Kyushu Institute of Technology Fellowship, and the Japan Society for the Promotion of Science Postdoctral Fellowship. The authors wish to express their thanks to the members of their group, especially Mr. T. Morodomi and Mr. Y. Kumagae, who carried out much of the constructional work.

Appendix: How to Evaluate Singular Integrals

In this analysis it is important to evaluate integrals in Eqs. (1*a*) and (1*b*) accurately because they have singularities when the integral interval includes boundary collocation points. In the previous studies these integrals were evaluated as shown in the following way [6,18]. The integral interval is divided into three parts as shown in Fig. 3*b* and Eq. (6),

$$I=\int_a^b f(x)dx=\int_a^{\varepsilon-\varepsilon_0} f(x)dx+\int_{\varepsilon-\varepsilon_0}^{\varepsilon+\varepsilon_0} f(x)dx+\int_{\varepsilon+\varepsilon_0}^{b} f(x)dx$$
$$=I_1+I_2+I_3. \tag{A1}$$

If we take suitable small values of $2\varepsilon_0$ the integral I can be evaluated as Eq. (*A*2),

$$I=\int_{-\varepsilon_0}^{\varepsilon_0} f(x)dx$$
$$=\int_{-\varepsilon_0}^{\varepsilon_0}\left(\frac{C_{-1}}{\varepsilon}+C_0+D_0\ln|\varepsilon|+C_1\varepsilon+D_1\varepsilon\ln|\varepsilon|\right.$$
$$\left.+D_1\varepsilon^2\ln|\varepsilon|+\cdots\right)$$
$$=2C_0\varepsilon_0+2D_0(\varepsilon_0\ln\varepsilon_0-\varepsilon_0). \tag{A2}$$

Here, C_{-1}, C_0, D_0 are constants, which may be obtained from expansion forms around $x=\varepsilon$ with painstaking tasks. In this study therefore the following method is applied. First, we set

$$I_{2\varepsilon}=I-\int_{-\varepsilon_0}^{\varepsilon_0} f(\varepsilon)d\varepsilon, \quad I_{4\varepsilon}=I-\int_{-2\varepsilon_0}^{2\varepsilon_0} f(\varepsilon)d\varepsilon,$$

$$I_{8\varepsilon}=I-\int_{-4\varepsilon_0}^{4\varepsilon_0} f(\varepsilon)d\varepsilon. \tag{A3}$$

These integrals can be expressed by

$$I_{2\varepsilon}=I-2C_0\varepsilon_0-2D_0(\varepsilon_0\ln\varepsilon_0-\varepsilon_0)=I-C'-D'\varepsilon_0\ln(2\varepsilon_0),$$

$$I_{4\varepsilon}=I-2C'\varepsilon_0-D'(2\varepsilon_0)\ln(2\varepsilon_0), \tag{A4}$$

$$I_{8\varepsilon}=I-4C'\varepsilon_0-D'(4\varepsilon_0)\ln(4\varepsilon_0),$$

where $C'=2(C_0-D_0)$, $D'=2D_0$. Since the integrals $I_{2\varepsilon}$, $I_{4\varepsilon}$, $I_{8\varepsilon}$ exclude singular points, they can be evaluated accurately through normal numerical procedure. Finally, we can evaluate I from

$$I=4I_{2\varepsilon}-4I_{4\varepsilon}-I_{8\varepsilon}. \tag{A5}$$

References

[1] Chen, D. H., and Nisitani, H., 1991, ''Singular Stress Fields at a Corner of Jointed Dissimilar Material Under Antiplane Loads,'' Trans. Jpn. Soc. Mech. Eng., Ser. A, **57**-542, pp. 2499–2503 (in Japanese) [1992, JSME Int. J., Ser. I, **35**-4, pp. 399–403].

[2] Chen, D. H., and Nisitani, H., 1993, ''Singular Stress Near the Corner of Jointed Dissimilar Materials,'' Trans. ASME, J. Appl. Mech., **60**, pp. 607–613.

[3] Chen, D. H., 1992, ''Analysis for Corner Singularity in Composite Materials Based on the Body Force Method,'' *Localized Damage II*, Vol. 1, pp. 397–421.

[4] Chen, D. H., and Nisitani, H., 1992, ''Analysis of Intensity of Singular Stress Fields at Fiber End,'' Trans. Jpn. Soc. Mech. Eng., Ser. A, **58**-554, pp. 1834–1838 (in Japanese).

[5] Nisitani, H., Chen, D. H., and Shibako, A., 1993, ''Singular Stress at a Corner of Lozenge Inclusion Under Antiplane Shear,'' Trans. Jpn. Soc. Mech. Eng., Ser. A, **59**-561, pp. 1191–1195 (in Japanese).

[6] Nisitani, H., 1967, ''The Two-Dimensional Stress Problem Solved Using an Electric Digital Computer,'' J. Jpn. Soc. Mech. Eng., **70**, pp. 627–632 [1968, Bull. JSME, **11**, pp. 14–23].

[7] Kasano, H., Matsumoto, H., and Nakahara, I., 1981, ''Tension of a Rigid Circular Cylindrical Inclusion in an Infinite Body,'' Trans. Jpn. Soc. Mech. Eng., Ser. A, **47**-413, pp. 18–26 (in Japanese).

[8] Hasegawa, H., and Yoshiya, K., 1994, ''Tension of Elastic Solid With Elastic Circular-Cylindrical Inclusion,'' Trans. Jpn. Soc. Mech. Eng., Ser. A, **60**-575, pp. 1585–1590 (in Japanese).

[9] Takao, Y., Taya, M., and Chou, T. W., 1981, ''Stress Field Due to a Cylindrical Inclusion With Constant Axial Eigenstrain in an Infinite Elastic Body,'' ASME J. Appl. Mech., **48**, pp. 853–858.

[10] Hasegawa, H., Lee, V.-G., and Mura, T., 1992, ''The Stress Fields Caused by a Circular Cylindrical Inclusion,'' ASME J. Appl. Mech., **59**, pp. 107–114.

[11] Wu, L., and Du, S., 1995, ''The Elastic Field Caused by a Circular Cylindrical Inclusion-Part I,'' ASME J. Appl. Mech., **62**, pp. 579–584.

[12] Wu, L., and Du, S., 1995, ''The Elastic Field Caused by a Circular Cylindrical Inclusion-Part II,'' ASME J. Appl. Mech., **62**, pp. 585–589.

[13] Noguchi, H., Nisitani, H., Goto, H., and Mori, K., 1987, ''Semi-Infinite Body With a Semi-Ellipsoidal Pit Under Tension,'' Trans. Jpn. Soc. Mech. Eng., Ser. A, (in Japanese), **53**-488, pp. 820–826 [1989, JSME Int. J., Ser. I, **32**-1, pp. 14–22].

[14] Noda, N., and Tomari, K., 1997, ''Fundamental Solution and its Application for Stress Analysis of Axisymmetric Body Under Asymmetric Uniaxial Tension,'' Bulletin of the Kyushu Institute of Technology, **67**, pp. 7–12.

[15] Bogy, D. B., and Wang, K. C., 1971, ''Stress Singularities at Interface Corners in Bonded Dissimilar Isotropic Elastic Materials,'' Int. J. Solids Struct., **7**, pp. 993–1005.

[16] Noda, N.-A., and Matsuo, T., 1998, ''Singular Integral Equation Method for Interaction Between Elliptical Inclusions,'' ASME J. Appl. Mech., **65**, pp. 310–319.

[17] Noda, N.-A., Kawashima, Y., Moriyama, S., and Oda, K., 1996, ''Interaction of Newly Defined Stress Intensity Factors for Angular Corners in a Row of Diamond-Shaped Inclusions,'' Int. J. Fract., **82**, pp. 267–295.

[18] Noda, N.-A., 1984, ''Stress Concentration Analysis for Notched Round Bar Under Torsion, Tension, and Bending,'' Ph.D. Dissertation, Kyushu University, pp. 50–64 (in Japanese).

General Solutions of Anisotropic Laminated Plates

W.-L. Yin
School of Civil and Environmental Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332-0355
e-mail: wanlee.yin@ce.gatech.edu

Anisotropic laminates with bending-stretching coupling possess eigensolutions that are analytic functions of the complex variables $x+\mu_k y$, where the eigenvalues μ_k and the corresponding eigenvectors are determined in the present analysis, along with the higher-order eigenvectors associated with repeated eigenvalues of degenerate laminates. The analysis and the resulting expressions are greatly simplified by using a mixed formulation involving a new set of elasticity matrices $\mathbf{A}^*$, $\mathbf{B}^*$, *and* $\mathbf{D}^*$. *There are 11 distinct types of laminates, each with a different expression of the general solution. For an infinite plate with an elliptical hole subjected to uniform in-plane forces and moments at infinity, closed-form solutions are obtained for all types of anisotropic laminates in terms of the eigenvalues and eigenvectors.* [DOI: 10.1115/1.1576804]

1 Introduction

Symmetric laminates belong to the class of laminates in which the coupling between bending and stretching is absent. For this class, the general problem of equilibrium under various types of loads and boundary conditions can be separated into an out-of-plane bending problem and an in-plane stretching/shearing problem. The latter is identical to the plane-stress anisotropic elasticity problem. The similarity between this problem and the bending problem of symmetric laminates was observed by Lekhnitskii [1]. Specifically, the deflection function of bending satisfies a fourth-order governing differential equation that is similar to the differential equation for the Airy's stress function. Lekhnitskii obtained the general solution of the bending of (nondegenerate) symmetric laminates by using the complex variable theory, in a way similar to his investigation of two-dimensional anisotropic elasticity, [2].

In this work, we consider anisotropic laminates that generally manifest bending-stretching coupling. Although the literature on the numerical analysis of laminated plates is vast, there appear to be few systematic attempts to investigate the analytical representation and mathematical structure of the solutions, to the extent that such researches have been pursued both in the classical theory of *isotropic* plates and in the two-dimensional anisotropic elasticity theory. Half a century after its publication, Lekhnitskii's work [1] still remains the standard reference on anisotropic plates.

The mature development of linear *isotropic* elasticity was due substantially to the systematic investigation of Muskhelishvili [3] and his associates, in which the use of complex function representations of general solutions, Green's functions and integral equations, are essential. In anisotropic elasticity, the representation is broadened by using multiple complex variables involving material eigenvalues. Such representations are especially important for finding singularity solutions of multimaterial wedges in composite structures near free edges and crack tips, [4–6], and for the boundary element formulation of elastostatic problems, [7,8]. However, similar representations of the general solution are presently available only for anisotropic laminates with simple (distinct) eigenvalues and for isotropic laminates, but not for the various other classes of degenerate laminates.

In this paper, we obtain explicit analytical expressions of the general solutions of all types of anisotropic laminates. By expressing the displacements, the stress potentials and the moment potentials in terms of complex analytic functions, the governing equilibrium equations and the constitutive relations are reduced to an algebraic eigenvalue problem, which possesses four complex conjugate pairs of eigenvalues and eigenvectors. Each simple eigenvalue μ_k and its eigenvector determine an eigensolution containing an arbitrary analytical function of $x+\mu_k y$. If repeated eigenvalues occur, then high-order eigenvectors may be needed to supplement the zeroth-order eigenvectors. In such degenerate cases, the eigensolutions have more complicated forms of expression.

In the conventional formulation of laminated plate theory, the constitutive relation is given by three symmetric stiffness matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{D}$, which reduce the equilibrium equations of the forces and moments to a system of differential equations governing the three displacement components u, v, and w, [9]. This formulation is ill-suited for the analytical task of determining the general solution. In this work, a *mixed* formulation is adopted by taking the curvatures and the in-plane forces as the primary unknown functions. The conventional stiffness matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ are replaced by new elasticity matrices $\mathbf{A}^*$, $\mathbf{B}^*$, and $\mathbf{D}^*$ (where $\mathbf{A}^*=\mathbf{A}^{-1}$, the other two matrices are defined by Eqs. (13*c*,*d*), and $\mathbf{B}^*$ is not symmetric). In this formulation, the determination of eigenvalues and eigenvectors of coupled laminates is reduced to an eigenvalue problem associated with a 2×2 matrix function, which can be solved effortlessly. It is shown, moreover, that the higher-order eigenvectors and eigensolutions associated with a repeated eigenvalue μ of degenerate laminates may be obtained in a formalistic way by differentiating appropriate analytical expressions of the lower-order eigenvectors and eigensolutions with respect to μ, which is regarded provisionally as a variable prior to evaluation at a specific multiple eigenvalue. Therefore, all eigenvectors and eigensolutions of the various types of anisotropic laminates may be obtained in explicit forms.

The present analysis bears a formal resemblance to the author's recent work on plane anisotropic elasticity, [10,11]. However, the eigenvectors in the present problem have the dimension eight, whereas those in the generalized two-dimensional elasticity (including coupling between the in-plane and the antiplane modes) have the dimension six. The present problem yields eleven different classes of anisotropic plates, while there are only five different classes of materials in plane anisotropic elasticity. Each one of the eleven types possesses distinct representations of the eigenvectors and of the general solution, depending on the multiplicity of the eigenvalues, and on whether they are normal, abnormal, or superabnormal. A classification of eight distinct types of eigenvalues is shown in Section 5, accompanied by a simple proof that no eigenvalue can be real if the laminate has a positive-definite strain

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, June 5, 2002; final revision, Nov. 22, 2002. Associate Editor: J. R. Barber. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Chair, Department of Mechanics 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.

energy function. In Section 7, we apply the general solution to an infinite laminate with a load-free elliptical hole subjected to uniform force and moment loads at infinity. An analytical solution is obtained for *all* types of laminates in terms of the respective eigenvectors, thereby extending the previous solution of Lu and Mahrenholtz [12] for nondegenerate laminates, in which the eigenvectors are shown implicitly but not explicitly due to the inherent complexity of the Stroh formalism used in their analysis.

2 Eigenvalues, Eigenvectors, and Eigensolutions

Let $u(x,y)$ and $v(x,y)$ denote the midplane tangential displacements of a laminated plate and let $w(x,y)$ denote the normal displacement. The midplane strains and curvatures are usually grouped in the order $\boldsymbol{\epsilon}=\{\epsilon_x,\epsilon_y,2\epsilon_{xy}\}^T=\{u_{,x},v_{,y},u_{,y}+v_{,x}\}^T$ and $\boldsymbol{\kappa}=\{w_{,xx},w_{,yy},2w_{,xy}\}^T$. They are related to the stress and moment resultants, $\mathbf{n}=\{N_x,N_y,N_{xy}\}^T$ and $\mathbf{m}=\{M_x,M_y,M_{xy}\}^T$, by the constitutive equations of an anisotropic laminated plate

$$\mathbf{n}=\mathbf{A}\boldsymbol{\epsilon}+\mathbf{B}\boldsymbol{\kappa},\quad \mathbf{m}=\mathbf{B}\boldsymbol{\epsilon}+\mathbf{D}\boldsymbol{\kappa}, \tag{1}$$

where the 3×3 symmetric matrices **A**, **B**, and **D** characterize the extensional stiffness, extension-bending coupling, and bending stiffness, respectively, [9].

An equilibrium solution of the plate problem with kinetic or kinematical boundary conditions and subjected to distributed tangential and normal surface loads may be decomposed into a particular solution under the same distributed loads and a complementary problem involving boundary loads and boundary displacements only. A particular solution is any solution of the inhomogeneous differential equations (containing the specified distributed loads) with no regard to boundary conditions. It may be obtained by integrating Green's functions (fundamental solutions) for the infinite domain weighted by the distributed load intensities. The present paper is concerned only with the complementary problem, which is governed by homogeneous equilibrium equations

$$N_{x,x}+N_{xy,y}=0,\quad N_{xy,x}+N_{y,y}=0, \tag{2}$$

$$M_{x,x}+M_{xy,y}+Q_x=0,\quad M_{xy,x}+M_{y,y}+Q_y=0,\quad Q_{x,x}+Q_{y,y}=0. \tag{3}$$

These equations have the following general solution in terms of three arbitrary functions $F(x,y)$, $\Psi_1(x,y)$ and $\Psi_2(x,y)$:

$$N_x=F_{,yy},\quad N_y=F_{,xx},\quad N_{xy}=-F_{,xy}, \tag{4}$$

$$M_y=\Psi_{1,x},\quad M_x=\Psi_{2,y},\quad -2M_{xy}=\Psi_{1,y}+\Psi_{2,x}, \tag{5}$$

$$Q_x=1/2(\Psi_{1,y}-\Psi_{2,x})_{,y},\quad Q_y=-1/2(\Psi_{1,y}-\Psi_{2,x})_{,x}. \tag{6}$$

Notice that M_y, M_x, and $-2M_{xy}$ are expressed as the derivatives of Ψ_1 and Ψ_2 in the same way that the midplane strains ϵ_x, ϵ_y, and $2\epsilon_{xy}$ are expressed in terms of the corresponding derivatives of the displacement functions u and v.

We seek solutions of the six functions u, v, w, F, Ψ_1, and Ψ_2 in the form

$$\boldsymbol{\chi}\equiv\{w_{,y},-w_{,x},F_{,y},-F_{,x},\Psi_1,\Psi_2,-u,-v\}^T=f(x+\mu y)\boldsymbol{\xi}, \tag{7}$$

where f is an arbitrary analytic function and the complex constant μ and the eight-dimensional complex vector $\boldsymbol{\xi}=\{\xi_1,\xi_2,\ldots,\xi_8\}^T$ will be determined subsequently. The superscript T indicates the transpose operation on a matrix or a row vector. The relations

$$w_{,xy}=-f'(x+\mu y)\mu\xi_2=w_{,yx}=f'(x+\mu y)\xi_1,$$

$$F_{,xy}=-f'(x+\mu y)\mu\xi_3=F_{,yx}=f'(x+\mu y)\mu\xi_4$$

require that

$$\xi_1=-\mu\xi_2,\quad \xi_3=-\mu\xi_4. \tag{8}$$

If we define

$$\boldsymbol{\eta}\equiv\{\xi_2,\xi_4\}^T, \tag{9a}$$

$$\boldsymbol{\phi}\equiv\{w_{,yy},w_{,xx},-w_{,xy},F_{,yy},F_{,xx},-F_{,xy}\}^T, \tag{9b}$$

$$\boldsymbol{\theta}\equiv\{M_y,M_x,-2M_{xy},-\epsilon_x,-\epsilon_y,-2\epsilon_{xy}\}^T. \tag{9c}$$

Then, from Eq. (7) follows that

$$\boldsymbol{\phi}=f'(x+\mu y)\boldsymbol{\Phi}(\mu)\boldsymbol{\xi}=f'(x+\mu y)\mathbf{P}(\mu)\boldsymbol{\eta}, \tag{10a}$$

$$\boldsymbol{\theta}=f'(x+\mu y)\mathbf{E}(\mu)\boldsymbol{\xi}, \tag{10b}$$

where

$$\boldsymbol{\Phi}(\mu)\equiv\begin{bmatrix}\mu&0&0&0&0&0&0&0\\0&-1&0&0&0&0&0&0\\-1&0&0&0&0&0&0&0\\0&0&\mu&0&0&0&0&0\\0&0&0&-1&0&0&0&0\\0&0&-1&0&0&0&0&0\end{bmatrix}, \tag{11a}$$

$$\mathbf{P}(\mu)\equiv\begin{bmatrix}-\mu^2&0\\-1&0\\\mu&0\\0&-\mu^2\\0&-1\\0&\mu\end{bmatrix} \tag{11b}$$

$$\mathbf{E}(\mu)\equiv\begin{bmatrix}0&0&0&0&1&0&0&0\\0&0&0&0&0&\mu&0&0\\0&0&0&0&\mu&1&0&0\\0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&\mu\\0&0&0&0&0&0&\mu&1\end{bmatrix}. \tag{11c}$$

Clearly,

$$\mathbf{P}^T\mathbf{E}=\mathbf{0}, \tag{12a}$$

$$\mathbf{YE}=[\mathbf{0}_{4\times4},\mathbf{I}_4], \tag{12b}$$

where $\mathbf{0}_{n\times n}$, $\mathbf{I}_n$ denote, respectively, zero and identity matrices of dimension $n\times n$ and

$$\mathbf{Y}(\mu)\equiv\begin{bmatrix}1&0&0&0&0&0\\-\mu&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&-\mu&0&1\end{bmatrix}. \tag{12c}$$

The variables ϕ and θ must be related by the laminate constitutive equation, i.e., Eq. (1). Define the symmetric matrix

$$\mathbf{C}^*\equiv\begin{bmatrix}\mathbf{D}^*&\mathbf{B}^*\\\mathbf{B}^{*T}&-\mathbf{A}^*\end{bmatrix} \tag{13a}$$

where

$$\mathbf{A}^*\equiv\mathbf{A}^{-1}, \tag{13b}$$

$$\mathbf{B}^*\equiv\boldsymbol{\Lambda}\mathbf{B}\mathbf{A}^{-1}, \tag{13c}$$

$$\mathbf{D}^*\equiv\boldsymbol{\Lambda}(\mathbf{D}-\mathbf{B}\mathbf{A}^{-1}\mathbf{B})\boldsymbol{\Lambda}, \tag{13d}$$

$$\boldsymbol{\Lambda}\equiv\begin{bmatrix}0&1&0\\1&0&0\\0&0&-2\end{bmatrix}. \tag{13e}$$

Then Eq. (1) becomes

$$\boldsymbol{\theta}=\mathbf{C}^*\boldsymbol{\phi}, \tag{14}$$

and, using Eqs. (10*a*,*b*), one obtains

$$\mathbf{E}(\mu)\boldsymbol{\xi}=\mathbf{C}^*\boldsymbol{\Phi}(\mu)\boldsymbol{\xi}=\mathbf{C}^*\mathbf{P}(\mu)\boldsymbol{\eta}. \tag{15}$$

Comparing Eqs. (1) and (14), one finds that the roles of $\mathbf{m}$ and $\boldsymbol{\kappa}$ have been interchanged and their elements redefined and rearranged. Premultiplying the last equation by the matrices $\mathbf{Y}$ and $\mathbf{P}^T$, and using Eqs. (12*a*,*b*), one has

$$[\mathbf{0}_{4\times4},\mathbf{I}_{4\times4}]\xi=\{\xi_5,\xi_6,\xi_7,\xi_8\}^{\mathrm{T}}=\mathbf{Y}(\mu)\mathbf{C}^*\mathbf{P}(\mu)\boldsymbol{\eta}, \tag{16}$$

$$\mathbf{M}(\mu)\boldsymbol{\eta}=\mathbf{0}, \tag{17}$$

where $\mathbf{M}(\mu)$ is the 2×2 *symmetric* matrix defined by

$$\mathbf{M}(\mu)\equiv\mathbf{P}(\mu)^{\mathrm{T}}\mathbf{C}^*\mathbf{P}(\mu). \tag{18}$$

The components of the matrix $\mathbf{M}(\mu)$ are quartic functions of μ, i.e.,

$$\begin{aligned}M_{11}(\mu)&=\{-\mu^2,-1,\mu\}\mathbf{D}^*\{-\mu^2,-1,\mu\}^T.\\ M_{12}(\mu)&=M_{21}(\mu)=\{-\mu^2,-1,\mu\}\mathbf{B}^*\{-\mu^2,-1,\mu\}^T,\\ M_{22}(\mu)&\equiv\{-\mu^2,-1,\mu\}(-\mathbf{A}^*)\{-\mu^2,-1,\mu\}^T.\end{aligned} \tag{19}$$

Combining Eqs. (8) and (16), one obtains an expression of the eight-dimensional vector $\boldsymbol{\xi}$ in terms of $\boldsymbol{\eta}=\{\xi_2,\xi_4\}$:

$$\xi=\mathbf{J}(\mu)\boldsymbol{\eta}, \tag{20}$$

where

$$\mathbf{J}(\mu)\equiv\begin{bmatrix}\mathbf{J}_1(\mu)\\ \mathbf{J}_2(\mu)\end{bmatrix},\quad \mathbf{J}_1(\mu)\equiv\begin{bmatrix}-\mu & 0\\ 1 & 0\\ 0 & -\mu\\ 0 & 1\end{bmatrix},$$

$$\mathbf{J}_2(\mu)\equiv\mathbf{Y}(\mu)\mathbf{C}^*\mathbf{P}(\mu). \tag{21}$$

Equation (17) has a nontrivial solution η if and only if the determinant of $\mathbf{M}(\mu)$ vanishes. This yields the characteristic equation

$$\delta(\mu)\equiv M_{11}(\mu)M_{22}(\mu)-[M_{12}(\mu)]^2=0. \tag{22}$$

It will be shown later that, if the strain energy of the laminate is positive definite for arbitrary combinations of stretching and bending deformations, then the polynomial Eq. (22) has no real roots. The strain energy of the laminate is given by

$$\begin{aligned}U(\boldsymbol{\kappa},\boldsymbol{\epsilon})&=1/2(\boldsymbol{\kappa}^T\mathbf{m}+\boldsymbol{\epsilon}^T\mathbf{n})=1/2(\boldsymbol{\kappa}^T\mathbf{D}\boldsymbol{\kappa}+2\boldsymbol{\kappa}^T\mathbf{B}\boldsymbol{\epsilon}+\boldsymbol{\epsilon}^T\mathbf{A}\boldsymbol{\epsilon})\\ &=1/2\{\hat{\boldsymbol{\kappa}}^T\mathbf{D}^*\hat{\boldsymbol{\kappa}}+\mathbf{n}^T\mathbf{A}^*\mathbf{n}),\end{aligned} \tag{23a}$$

where

$$\hat{\boldsymbol{\kappa}}\equiv\{w_{,yy},w_{,xx},-w_{,xy}\}. \tag{23b}$$

It is interesting to notice that the last expression of Eq. (23*a*) does not involve $\mathbf{B}^*$.

For the function U to be positive definite under pure bending and pure stretching states, respectively, the stiffness matrices $\mathbf{D}$ and $\mathbf{A}$ must be positive definite. Positive definiteness of U under more general states of deformation requires, in addition, positive definiteness of $\mathbf{D}^*=\Lambda(\mathbf{D}-\mathbf{B}\mathbf{A}^{-1}\mathbf{B})\Lambda$ or, equivalently, of $\mathbf{D}-\mathbf{B}\mathbf{A}^{-1}\mathbf{B}$. Laminates with a $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}$ large or comparable in magnitude to $\mathbf{D}$ may violate this condition, making the state $\boldsymbol{\epsilon}=\boldsymbol{\kappa}=\mathbf{0}$ precarious (a state of saddle-point equilibrium). This cannot happen if all layers are bonded together without initial stress, because any deformation from the initial state increases the strain energy in every part. But it is conceivable to produce a laminate by prestraining certain layers in tension and others in compression before bonding the layers, so that a slight disturbance from the reference state $\mathbf{n}=\boldsymbol{\kappa}=\mathbf{0}$ will cause the laminate to curl up and to drastically reduce the stresses in severely pre-strained layers, similar to the release of subterranean stress during earthquakes. Laminated plate theory is broad enough to encompass laminates with differently prestressed layers. For such laminates $\boldsymbol{\kappa}=\mathbf{0}$ still refers to the flat state, but the state $\boldsymbol{\epsilon}=\mathbf{0}$ cannot be identified a priori except as the state corresponding to $\mathbf{n}=\mathbf{0}$. Be it so, it is not the principal reason for adopting the constitutive relation of Eq. (14) in the present theory, rather than the customary form of Eq. (1). The present choice is dictated essentially by the mathematical structure of the problem, and by the resulting expressions of the general solutions of laminated plates. These general solutions are obtained analytically in terms of the elements of the matrices $\mathbf{A}^*$, $\mathbf{B}^*$, and $\mathbf{D}^*$. To convert the solutions into expressions involving the elements of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ would make the expressions unduly complicated.

Under the assumption of a positive-definite U, the eigenvalues are not real and they occur in conjugate pairs since the complex conjugate of Eq. (22) is $M_{11}(\bar{\mu})M_{22}(\bar{\mu})-M_{12}(\bar{\mu})^2=0$, where $\bar{\mu}$ denotes the complex conjugate of μ. For each root μ_0, there is at least one nontrivial solution $\boldsymbol{\eta}$ of Eq. (17). Equations (20) and (21) yield the eight-dimensional constant vector $\boldsymbol{\xi}$, and Eq. (7) gives a solution of the laminated plate containing an arbitrary analytic function f of the complex variable $x+\mu y$. A root μ of the characteristic equation will be called an *eigenvalue*, and the corresponding vectors $\boldsymbol{\xi}$ and $\boldsymbol{\chi}$ will be called the *eigenvector* and *eigensolution* associated with μ. Since the elements of $\mathbf{M}(\mu)$ and $\mathbf{J}(\mu)$ are polynomials with real coefficients, it follows that if $\boldsymbol{\xi}$ is an eigenvector associated with μ, then the complex conjugate vector $\bar{\boldsymbol{\xi}}$ is an eigenvector associated with the conjugate eigenvalue $\bar{\mu}$. $\bar{\mu}$ and $\bar{\boldsymbol{\xi}}$ determine another eigensolution involving an arbitrary complex function $g(x+\bar{\mu}y)$. The sum of the two eigensolutions yield real values of $\boldsymbol{\chi}$ if and only if the functions f and g are related by

$$g(x+\bar{\mu}y)\equiv\overline{f(x+\mu y)}. \tag{24}$$

When this is the case, Eqs. (4)–(6) yield real values of the forces and moments. We will assume that the eigensolutions associated with a complex conjugate pair of eigenvalues are always combined in this way to yield real-valued physical quantities.

Consider the adjoint matrix of $\mathbf{M}$

$$\mathbf{W}(\mu)=\begin{bmatrix}M_{22} & -M_{12}\\ -M_{12} & M_{11}\end{bmatrix}=\mathbf{P}(\mu)^T\cdot\begin{bmatrix}-\mathbf{A}^* & -\mathbf{B}^*\\ -\mathbf{B}^{*T} & \mathbf{D}^*\end{bmatrix}\mathbf{P}(\mu), \tag{25}$$

which satisfies the identity

$$\mathbf{W}(\mu)\mathbf{M}(\mu)=\mathbf{M}(\mu)\mathbf{W}(\mu)=\delta(\mu)\mathbf{I}. \tag{26}$$

An eigenvalue μ_0 will be called *normal* if it is *not* a common root of all three equations $M_{11}(\mu)=M_{12}(\mu)=M_{22}(\mu)=0$, i.e., if $\mathbf{M}(\mu_0)$, and hence $\mathbf{W}(\mu_0)$, are not zero matrices. Otherwise μ_0 will be called *abnormal*. For a normal eigenvalue μ_0, $\mathbf{W}(\mu_0)$ is a matrix of rank one. Therefore, it has one and only one independent column vector. Then it has at least one nonvanishing diagonal element. Let $\boldsymbol{\eta}$ be the column containing the first nonvanishing diagonal element of $\mathbf{W}(\mu_0)$. Equation (26) with $\delta(\mu_0)=0$ ensures that $\boldsymbol{\eta}$ satisfies (17), i.e., $\mathbf{M}(\mu_0)\boldsymbol{\eta}=\mathbf{0}$. Therefore, $\boldsymbol{\xi}=\mathbf{J}(\mu_0)\boldsymbol{\eta}$ is an eigenvector associated with μ_0, and Eq. (7) gives the eigensolution $\boldsymbol{\chi}$.

If μ_0 is an *abnormal* eigenvalue, then $\mathbf{M}(\mu_0)$ and $\mathbf{W}(\mu_0)$ are zero matrices. Since μ_0 is a common root of all elements of $M_{11}(\mu)=M_{12}(\mu)=M_{22}(\mu)=0$, it must be a multiple root of $\delta(\mu)=0$. Equation (17) is trivially satisfied by an arbitrary $\boldsymbol{\eta}$. Two independent eigenvectors are given by $\boldsymbol{\xi}=\mathbf{J}(\mu_0)\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ may be chosen as $\{1,0\}^T$ and $\{0,1\}^T$. This choice yields the two columns of $\mathbf{J}(\mu_0)$ as the eigenvectors, and each may be multiplied by an arbitrary analytic function of $x+\mu_0 y$ to obtain a corresponding eigensolution.

3 Higher-Order Eigenvectors and Eigensolutions

Equations (17) and (20) yield one independent eigenvector for a normal eigenvalue and two independent eigenvectors for an abnormal one. If these numbers are smaller than the multiplicity of the eigenvalue, then the two equations in conjunction with (7) do not yield the complete set of eigensolutions (i.e., four complex conjugate pairs). In such cases, additional independent solutions must be found. Such additional solutions are determined by relations different from those governing the preceding eigensolutions, and they have more complicated forms of representation involving successive eigenvectors of higher orders (also called ''generalized eigenvectors'' in the literature on anisotropic elasticity). As independent solutions, the higher-order eigensolutions are not secondary in status. They are equally indispensable to the general solution of the laminate, in the same way that a second analytical function is indispensable to the general solution of the biharmonic equation in plane isotropic elasticity.

If an eigenvalue is of multiplicity p, and it has only r independent zeroth-order eigensolutions, then $p-r$ higher-order independent solutions will be sought. We consider an Nth-order eigensolution $(1\leqslant N\leqslant p-1)$ having the following expression:

$$\boldsymbol{\chi}^{[N]}=\sum_{0\leqslant j\leqslant N}(N,j)y^j f^{(j)}(x+\mu y)\boldsymbol{\xi}^{[N-j]}, \tag{27}$$

where $(N,j)\equiv N!/(N-j)!j!$, $f^{(j)}$ denotes the jth derivative of the complex analytic function f, and $\boldsymbol{\xi}^{[1]}$, $\boldsymbol{\xi}^{[2]}$, $\ldots$, $\boldsymbol{\xi}^{[N]}$ is a set of complex constant vectors of dimension eight. Differentiating the expression and regrouping the resulting terms, one obtains

$$\begin{aligned}\boldsymbol{\phi}^{[N]}&=y^N f^{(N+1)}\boldsymbol{\Phi}\boldsymbol{\xi}^{[0]}+\sum_{0\leqslant j\leqslant N-1}(N,j)y^j f^{(j+1)}\{\boldsymbol{\Phi}\boldsymbol{\xi}^{[N-j]}\\&\quad+(N-j)\boldsymbol{\Phi}'\boldsymbol{\xi}^{[N-j-1]},\end{aligned} \tag{28}$$

$$\begin{aligned}\boldsymbol{\theta}^{[N]}&=y^N f^{(N+1)}\mathbf{E}\boldsymbol{\xi}^{[0]}+\sum_{0\leqslant j\leqslant N-1}(N,j)y^j f^{(j+1)}\{\mathbf{E}\boldsymbol{\xi}^{[N-j]}\\&\quad+(N-j)\mathbf{E}'\boldsymbol{\xi}^{[N-j-1]},\end{aligned} \tag{29}$$

where $\boldsymbol{\Phi}(\mu)$ and $\mathbf{E}(\mu)$ are as given by Eqs. (11*a*,*c*). Let

$$\mathbf{K}(\mu)\equiv\begin{bmatrix}1&\mu&0&0&0&0&0&0\\0&0&1&\mu&0&0&0&0\end{bmatrix} \tag{30}$$

Then the relations $w_{,xy}=w_{,yx}$ and $F_{,xy}=F_{,yx}$ imply

$$\begin{aligned}&\sum_{0\leqslant j\leqslant N}(N,j)y^j f^{(j)}(x+\mu y)\mathbf{K}(\mu)\boldsymbol{\xi}^{[N-j]}\\&\quad+\sum_{0\leqslant j\leqslant N}(N,j)jy^{j-1}f^{(j)}(x+\mu y)\mathbf{K}'\boldsymbol{\xi}^{[N-j]}\\&\quad=\sum_{0\leqslant j\leqslant N}(N,j)y^j f^{(j+1)}(x+\mu y)\{\\&\qquad\times\mathbf{K}(\mu)\boldsymbol{\xi}^{[N-j]}+(N-j)\mathbf{K}'\boldsymbol{\xi}^{[N-j-1]}\}=\mathbf{0}.\end{aligned}$$

It is understood that $\boldsymbol{\xi}^{[j]}=\mathbf{0}$ if j is a negative integer. Let $z=x+\mu y$. Then $y=(z-\overline{z})/(\mu-\overline{\mu})$, so that the last equation involves a polynomial function of the complex variable $\overline{z}$. For this equation to be valid in a region of the $x-y$ plane, it is necessary that the coefficients of the various powers of $\overline{z}$ all vanish. Hence,

$$\mathbf{K}(\mu)\boldsymbol{\xi}^{[0]}=\mathbf{0}, \tag{31a}$$

$$\mathbf{K}(\mu)\boldsymbol{\xi}^{[j]}+j\mathbf{K}'\boldsymbol{\xi}^{[j-1]}=\mathbf{0},\quad(1\leqslant j\leqslant N). \tag{31b}$$

We now define $\boldsymbol{\eta}^{[j]}$ as the two-dimensional vector consisting of the second and fourth elements of $\boldsymbol{\xi}^{[j]}$, i.e.,

$$\boldsymbol{\eta}^{[j]}=\mathbf{K}'\boldsymbol{\xi}^{[j]}=\begin{bmatrix}0&1&0&0&0&0&0&0\\0&0&0&1&0&0&0&0\end{bmatrix}\boldsymbol{\xi}^{[j]}.$$

The last equation and (31*a*,*b*) may be combined into a single expression

$$\begin{bmatrix}\mathbf{K}'\\\mathbf{K}\end{bmatrix}\boldsymbol{\xi}^{[j]}=\begin{bmatrix}\boldsymbol{\eta}^{[j]}\\-j\mathbf{K}'\boldsymbol{\xi}^{[j-1]}\end{bmatrix},\quad(j=0,1,\ldots,N). \tag{32}$$

This equation, after premultiplication by

$$\begin{bmatrix}-\mu&0&1&0\\1&0&0&0\\0&-\mu&0&1\\0&1&0&0\end{bmatrix}$$

and repeated substitutions, yields the following expression for the first four elements of $\boldsymbol{\xi}^{[j]}$:

$$[\mathbf{I}_4,\mathbf{0}_{4\times4}]\boldsymbol{\xi}^{[j]}=\mathbf{J}_1(\mu)\boldsymbol{\eta}^{[j]}+j\mathbf{J}_1'\boldsymbol{\eta}^{[j-1]},\quad(j=1,2,\ldots,N) \tag{33}$$

where $\mathbf{J}_1(\mu)$ was given by the second equation of (21).

We now require that the kinetic and kinematical variables of the Nth-order generalized eigensolution be related by the laminate constitutive relation (14), i.e.,

$$\boldsymbol{\theta}^{[N]}=\mathbf{C}^*\boldsymbol{\phi}^{[N]}. \tag{34}$$

Premultiplying (34) by $\mathbf{Y}(\mu)$, and using Eqs. (12*b*), (13), (21), (28), (29) and $y=(z-\overline{z})/(\mu-\overline{\mu})$, one obtains an expression that may be separated into various powers of $\overline{z}$:

$$\begin{aligned}[\mathbf{0}_{4\times4},\mathbf{I}_{4\times4}]\boldsymbol{\xi}^{[j]}&=j\mathbf{Y}'\mathbf{E}(\mu)\boldsymbol{\xi}^{[j-1]}+j\mathbf{Y}(\mu)\mathbf{C}^*\mathbf{P}'(\mu)\boldsymbol{\eta}^{[j-1]}\\&\quad+(j,2)\mathbf{Y}(\mu)\mathbf{C}^*\mathbf{P}''\boldsymbol{\eta}^{[j-2]}\\&=\mathbf{J}_2(\mu)\boldsymbol{\eta}^{[j]}+(j,1)\mathbf{J}_2'(\mu)\boldsymbol{\eta}^{[j-1]}\\&\quad+(j,2)\mathbf{J}_2''(\mu)\boldsymbol{\eta}^{[j-2]}+(j,3)\mathbf{J}_2'''(\mu)\boldsymbol{\eta}^{[j-3]}\\&\quad\times(j=0,1,\ldots,N).\end{aligned}$$

Combining (33) with the last equation, one has

$$\begin{aligned}\boldsymbol{\xi}^{[j]}&=\mathbf{J}(\mu)\boldsymbol{\eta}^{[j]}+(j,1)\mathbf{J}'(\mu)\boldsymbol{\eta}^{[j-1]}+(j,2)\mathbf{J}''(\mu)\boldsymbol{\eta}^{[j-2]}\\&\quad+(j,3)\mathbf{J}'''(\mu)\boldsymbol{\eta}^{[j-3]}.\end{aligned} \tag{35}$$

Again, it is understood that $\boldsymbol{\eta}^{[j]}=\mathbf{0}$ for negative j. Hence an Nth-order eigenvector $\boldsymbol{\xi}^{[N]}$ may be expressed in terms of the two-dimensional vectors $\boldsymbol{\eta}^{[0]}$, $\boldsymbol{\eta}^{[1]}$, $\ldots$ $\boldsymbol{\eta}^{[N]}$.

Premultiplying (34) by $\mathbf{P}(\mu)^T$ and using Eqs. (12*a*), (28), (29), (35), and (18), one obtains the following governing equation for the eigenvectors of various orders after a sequence of algebraic manipulations:

$$\sum_{0\leqslant j\leqslant N}(N,j)\mathbf{M}^{(j)}(\mu)\boldsymbol{\eta}^{[N-j]}=\mathbf{0},\quad(N=0,1,\ldots,p-1) \tag{36}$$

where $\mathbf{M}^{(j)}$ denotes the jth derivative of the matrix defined by Eq. (18).

Equation (36) may be solved explicitly for $\boldsymbol{\eta}^{[j]}(1\leqslant j\leqslant N\leqslant p-1)$, but the form of the solution depends on whether μ is a normal or abnormal eigenvalue. If μ is normal, then $\mathbf{M}(\mu)$ and its adjoint matrix $\mathbf{W}(\mu)$ are both symmetric matrices of rank one. Hence $\mathbf{W}(\mu)$ has at least one nonzero diagonal element and we define $\boldsymbol{\rho}$ to be the vector $\{1,0\}^T$ if $|W_{11}(\mu)|\geqslant|W_{22}(\mu)|$ and to be $\{0,1\}^T$ if otherwise. Then $\mathbf{W}(\mu)\boldsymbol{\rho}$ is a nonvanishing vector. Repeated differentiation of Eq. (26) yields

$$\sum_{0\leqslant j\leqslant N}(N,j)\mathbf{M}^{(j)}\mathbf{W}^{(N-j)}(\mu)=\boldsymbol{\delta}^{(N)}(\mu)\mathbf{I}\quad(N=0,1,\ldots,L). \tag{37}$$

An Nth-order eigenvector is required only if p is equal to or greater than $N+1$. In that case $\boldsymbol{\delta}^{(j)}(\mu)=0$ for $0\leqslant j\leqslant N\leqslant p-1$. Hence the right-hand side of Eq. (37) vanishes. We choose $\boldsymbol{\eta}^{[0]}=\mathbf{W}(\mu)\boldsymbol{\rho}$, and take each higher order $\boldsymbol{\eta}^{[j]}$ to be the corresponding column of $\mathbf{W}^{(j)}(\mu)$, $1\leqslant j\leqslant N$. Then all $p-1$ equations of (36) are satisfied. Hence $\boldsymbol{\eta}^{[0]}$, $\boldsymbol{\eta}^{[1]}$, $\ldots$, $\boldsymbol{\eta}^{[p-1]}$ and Eq. (35) determine the successive eigenvectors of increasing orders:

$$\begin{aligned}\boldsymbol{\xi}^{[0]}&=\mathbf{JW}\boldsymbol{\rho},\quad \boldsymbol{\xi}^{[1]}=(\mathbf{JW}'+\mathbf{J}'\mathbf{W})\boldsymbol{\rho},\\ \boldsymbol{\xi}^{[2]}&=(\mathbf{JW}''+2\mathbf{J}'\mathbf{W}'+\mathbf{J}''\mathbf{W})\boldsymbol{\rho},\\ \boldsymbol{\xi}^{[3]}&=(\mathbf{JW}'''+3\mathbf{J}'\mathbf{W}''+3\mathbf{J}''\mathbf{W}'+\mathbf{J}'''\mathbf{W})\boldsymbol{\rho}.\end{aligned}\tag{38}$$

Equation (27) gives the corresponding eigensolutions

$$\begin{aligned}\boldsymbol{\chi}^{[0]}&=f\mathbf{JW}\boldsymbol{\rho},\quad \boldsymbol{\chi}^{[1]}=f(\mathbf{JW}'+\mathbf{J}'\mathbf{W})\boldsymbol{\rho}+yf'\mathbf{JW}\boldsymbol{\rho},\\ \boldsymbol{\chi}^{[2]}&=f(\mathbf{JW}''+2\mathbf{J}'\mathbf{W}'+\mathbf{J}''\mathbf{W})\boldsymbol{\rho}+2yf'(\mathbf{JW}'+\mathbf{J}'\mathbf{W})\boldsymbol{\rho}\\ &\quad+y^2f''\mathbf{JW}\boldsymbol{\rho},\\ \boldsymbol{\chi}^{[3]}&=f(\mathbf{JW}'''+3\mathbf{J}'\mathbf{W}''+3\mathbf{J}''\mathbf{W}'+\mathbf{J}'''\mathbf{W})\boldsymbol{\rho}+3yf'(\mathbf{JW}''+2\mathbf{J}'\mathbf{W}'\\ &\quad+\mathbf{J}''\mathbf{W})\boldsymbol{\rho}+3y^2f''(\mathbf{JW}'+\mathbf{J}'\mathbf{W})\boldsymbol{\rho}+y^3f'''\mathbf{JW}\boldsymbol{\rho}.\end{aligned}\tag{39}$$

Clearly, the zeroth-order eigenvector $\boldsymbol{\xi}^{[0]}$ and eigensolution $\boldsymbol{\chi}^{[0]}$ are identical to those determined in the previous section from Eqs. (17) and (20).

4 High-Order Eigensolutions Associated With Abnormal and Superabnormal Eigenvalues

Next we consider an *abnormal* eigenvalue μ_0 with the multiplicity $p\geqslant 3$, which has two independent zeroth-order eigenvectors given by the two columns of $\mathbf{J}(\mu_0)$ and requires, in addition, one or two higher-order eigenvectors for the cases $p=3$ and $p=4$, respectively.

The higher-order eigensolutions for a *normal* eigenvalue, obtained in the last section, were based on choosing $\boldsymbol{\eta}^{[j]}$ to be the column of $\mathbf{W}^{(j)}(\mu_0)$ which corresponds to a nonvanishing column of $\mathbf{W}(\mu_0)$, $1\leqslant j\leqslant L$. This procedure fails for an abnormal or superabnormal eigenvalue, for which $\mathbf{W}(\mu_0)$ has no nonvanishing columns. The proper nontrivial solutions of $\boldsymbol{\eta}^{[j]}$ then depend on whether or not $\mathbf{M}'(\mu_0)$ is also the zero matrix. If it is not, then it is of rank one and so is its adjoint matrix $\mathbf{W}'(\mu_0)$. Being a symmetric 2×2 matrix of rank one, at least one of the two diagonal elements of $\mathbf{W}'(\mu_0)$ is not zero. We now define $\hat{\boldsymbol{\rho}}$ to be the vector $\{1,0\}^T$ if $|W'_{11}(\mu)|\geqslant|W'_{22}(\mu)|$ and to be $\{0,1\}^T$ if otherwise. Then $W\equiv\hat{\boldsymbol{\rho}}^T\mathbf{W}'(\mu)\hat{\boldsymbol{\rho}}$ does not vanish. We then choose

$$\boldsymbol{\eta}^{[j]}=\mathbf{W}^{(j)}(\mu_0)\hat{\boldsymbol{\rho}},\quad (j=1,\ldots,p-1).\tag{40}$$

Equation (37), in conjunction with $\delta^{(j)}(\mu)=0$ for $0\leqslant j\leqslant p-1$, still ensure that $\boldsymbol{\eta}^{[0]}=\mathbf{0}$, $\boldsymbol{\eta}^{[1]}$, $\ldots$ and $\boldsymbol{\eta}^{[p-1]}$ satisfy all equations of (36). Therefore, they determine the eigenvectors and eigensolutions of the required orders via Eqs. (27) and (35). Substitution of (40) into (27) and (35) yields expressions of $\boldsymbol{\xi}^{[j]}$ and $\boldsymbol{\chi}^{[j]}$ that differ formally from those of a normal eigenvalue, Eqs. (39) and (38), only in replacing the vector $\boldsymbol{\rho}$ by $\hat{\boldsymbol{\rho}}$, i.e., in possibly making a different choice of the column from the 8×2 matrix $\mathbf{JW}$ and its μ-derivatives to ensure nontrivial results. But the new expressions for an abnormal eigenvalue also contain fewer terms because all terms involving the zero matrix $\mathbf{W}$ vanish. Furthermore, the contribution of the term $f\,\mathbf{JW}'''$ to $\boldsymbol{\chi}^{[3]}$ in the last equation of (39) to the general solution is not different from that of a zeroth-order eigensolution with the eigenvector $\boldsymbol{\xi}^{[0]}=\mathbf{J}\boldsymbol{\eta}^{[0]}=\mathbf{JW}'''$ (Since $\mathbf{M}(\mu_0)=\mathbf{0}$, Eq. (17) is satisfied by an *arbitrary* $\boldsymbol{\eta}^{[0]}$, in particular by $\boldsymbol{\eta}^{[0]}=\mathbf{JW}'''$). Hence the term involving $\mathbf{JW}'''$ in the expressions of $\boldsymbol{\xi}^{[3]}$ and $\boldsymbol{\chi}^{[3]}$ may be absorbed into the zeroth-order eigensolutions. In fact, according to Eq. (36), $\boldsymbol{\eta}^{[3]}$ is determined by the following governing equation:

$$\mathbf{M}'''(\mu_0)\boldsymbol{\eta}^{[0]}+3\mathbf{M}''(\mu_0)\boldsymbol{\eta}^{[1]}+3\mathbf{M}'(\mu_0)\boldsymbol{\eta}^{[2]}+\mathbf{M}(\mu_0)\boldsymbol{\eta}^{[3]}=\mathbf{0}.$$

But the last term on the left-hand side vanishes because $\mathbf{M}(\mu_0)$ is the zero matrix. Hence the equation imposes no restriction at all on $\boldsymbol{\eta}^{[3]}$, which may therefore be set to zero without loss of generality, and that in effect throws out the term $f\,\mathbf{JW}'''$ from $\boldsymbol{\chi}^{[3]}$. Although one has the option of discarding or keeping the term f $\mathbf{JW}'''$, it will be retained in the following expressions of $\boldsymbol{\xi}^{[3]}$ and $\boldsymbol{\chi}^{[3]}$ to facilitate future analysis of the algebraic structure of the solution spaces. Hence we have

$$\boldsymbol{\xi}^{[1]}=\mathbf{JW}'\hat{\boldsymbol{\rho}},\tag{41a}$$

$$\boldsymbol{\xi}^{[2]}=(\mathbf{JW}''+2\mathbf{J}'\mathbf{W}')\hat{\boldsymbol{\rho}},\tag{41b}$$

$$\boldsymbol{\xi}^{[3]}=(\mathbf{JW}'''+3\mathbf{J}'\mathbf{W}''+3\mathbf{J}''\mathbf{W}')\hat{\boldsymbol{\rho}},\tag{41c}$$

$$\boldsymbol{\chi}^{[1]}=\mathrm{f}_1\mathbf{JW}'\hat{\boldsymbol{\rho}},\tag{42a}$$

$$\boldsymbol{\chi}^{[2]}=f_2(\mathbf{JW}''+2\mathbf{J}'\mathbf{W}')\hat{\boldsymbol{\rho}}+2yf_2'\mathbf{JW}'\hat{\boldsymbol{\rho}},\tag{42b}$$

$$\begin{aligned}\boldsymbol{\chi}^{[3]}&=f_3(\mathbf{JW}'''+3\mathbf{J}'\mathbf{W}''+3\mathbf{J}''\mathbf{W}')\hat{\boldsymbol{\rho}}\\ &\quad+3yf_3'(\mathbf{JW}''+2\mathbf{J}'\mathbf{W}')\hat{\boldsymbol{\rho}}+3y^2f_3''\mathbf{JW}'\hat{\boldsymbol{\rho}},\end{aligned}\tag{42c}$$

where f_1, f_2, and f_3 are arbitrary analytic functions of $x+\mu_0 y$. The preceding list must be supplemented by a zeroth-order eigenvector $\boldsymbol{\xi}^{[0]}$ chosen from the two columns of $\mathbf{J}(\mu_0)$, and by the corresponding eigensolution $\boldsymbol{\chi}^{[0]}$. The choice must be made in such a way as to ensure that $\boldsymbol{\xi}^{[0]}$ and $\boldsymbol{\xi}^{[1]}$ are linearly independent. We therefore take

$$\boldsymbol{\xi}^{[0]}=\mathbf{J}(\mu_0)\begin{bmatrix}0&1\\1&0\end{bmatrix}\hat{\boldsymbol{\rho}},\tag{41d}$$

$$\boldsymbol{\chi}^{[0]}=\mathrm{f}_0(\mathrm{x}+\mu_0)\boldsymbol{\xi}^{[0]}.\tag{42d}$$

Besides having to replace the vector $\boldsymbol{\rho}$ by $\hat{\boldsymbol{\rho}}$, another important consequence of abnormality upon the structure of the general solutions is that, with $\mathbf{JW}\hat{\boldsymbol{\rho}}=\mathbf{0}$, the term involving the highest derivative $y^Nf^{[N]}(x+\mu y)$ in Eq. (27) vanishes. Therefore, an eigensolution originally sought as of order N in the scheme of Eq. (27) turns out to be effectively one order lower. Thus the factors y^3f''', y^2f'' and yf', which belong to a third, second, and first-order eigensolution, respectively, no longer appear in the expressions of $\boldsymbol{\chi}^{[3]}$, $\boldsymbol{\chi}^{[2]}$, and $\boldsymbol{\chi}^{[1]}$, Eqs. (42*a*,*b*,*c*). In contrast to a normal quadruple eigenvalue, which has one eigensolution of every order from zero to three, an abnormal quadruple eigenvalue has two zeroth-order eigensolutions, one eigensolution $\boldsymbol{\chi}^{[2]}$ effectively of the first order, and another $\boldsymbol{\chi}^{[3]}$ effectively of the second order. The effective order refers to the highest-order derivative of the arbitrary function f that appears in the expression of an eigensolution.

For an abnormal eigenvalue, $\boldsymbol{\chi}^{[1]}=f\mathbf{JW}'\hat{\boldsymbol{\rho}}$ is effectively a zeroth-order eigensolution. It is redundant and, in the case of a abnormal *double* eigenvalue, it has been discarded in favor of the two zeroth-order solutions whose eigenvectors are the two columns of $\mathbf{J}(\mu_0)$, because the latter are simpler in form. However, for abnormal eigenvalues of multiplicity three or four, the eigenvector $\boldsymbol{\xi}^{[1]}=\mathbf{JW}'\hat{\boldsymbol{\rho}}$ is needed in the expressions of the higher-order eigensolutions, as seen in the last terms of the expressions of $\boldsymbol{\chi}^{[2]}$ and $\boldsymbol{\chi}^{[3]}$ in Eqs. (42*b*,*c*). For this reason, $\boldsymbol{\xi}^{[1]}$ is included in Eq. (41), while only one of the two columns of $\mathbf{J}(\mu_0)$ is chosen in Eq. (41*d*) to complete the set of independent eigenvectors. Equations (42*a*,*b*,*c*,*d*) may be written in the matrix form

$$\{\boldsymbol{\chi}^{[0]},\boldsymbol{\chi}^{[1]},\boldsymbol{\chi}^{[2]},\boldsymbol{\chi}^{[3]}\}=\{\boldsymbol{\xi}^{[0]},\boldsymbol{\xi}^{[1]},\boldsymbol{\xi}^{[2]},\boldsymbol{\xi}^{[3]}\}\times\begin{bmatrix} f_0 & 0 & 0 & 0 \\ 0 & f_1 & 2yf_2' & 3y^2f_3'' \\ 0 & 0 & f_2 & 3yf_3' \\ 0 & 0 & 0 & f_3 \end{bmatrix}. \quad (43)$$

The complete list of three eigenvectors and eigensolutions associated with an abnormal *triple* eigenvalue are also given by Eqs. (41) and (42) when the expressions of $\boldsymbol{\xi}^{[3]}$ and $\boldsymbol{\chi}^{[3]}$ are removed.

An eigenvalue μ_0 is called superabnormal if both $\mathbf{M}(\mu_0)$ and $\mathbf{M}'(\mu_0)$ are zero matrices. Such an eigenvalue must be a double root of each scalar component of the matrix equation $\mathbf{M}(\mu)=\mathbf{0}$, and so must be $\overline{\mu_0}$, and their multiplicity must be four. Then $\delta(\mu_0)=\delta'(\mu_0)=\delta''(\mu_0)=\delta'''(\mu_0)=0$ and $\mathbf{M}(\mu)$ has the form

$$\mathbf{M}(\mu)=(\mu-\mu_0)^2(\mu-\overline{\mu_0})^2\mathbf{C},$$

where $\mathbf{C}$ is a 2×2 constant matrix. Again, two independent eigenvectors are given by the two columns of $\mathbf{J}(\mu_0)$. To obtain first-order eigensolutions, we notice that for $N=1$ and $\mathbf{M}(\mu_0)=\mathbf{M}'(\mu_0)=\mathbf{0}$, Eq. (36) is trivially satisfied regardless of $\boldsymbol{\eta}^{[0]}$ and $\boldsymbol{\eta}^{[1]}$. Setting $\boldsymbol{\eta}^{[1]}=\mathbf{0}$ and choosing $\boldsymbol{\eta}^{[0]}$ to be, successively, the two column of the identity matrix $\mathbf{I}_2$, then Eq. (35) with $j=1$ yields two independent first-order eigenvectors given by the two columns of the matrix $\mathbf{J}'(\mu_0)$. Equation (27) yields the corresponding first-order eigensolutions given by the two columns of the following matrix:

$$\mathbf{J}'(\mu)\begin{bmatrix} f_1 & 0 \\ 0 & f_2 \end{bmatrix}+\mathbf{J}(\mu)\begin{bmatrix} yf_1' & 0 \\ 0 & yf_2' \end{bmatrix}, \quad (44)$$

where f_1 and f_2 are two arbitrary analytic functions.

For each eigensolution obtained in this and the previous section, the expressions for the midplane strains, bending and twisting curvatures and stress and moment resultants may be obtained from Eqs. (28) and (29).

5 Classification of Eigenvalues and Proof of the Complexity

The preceding results imply the following classification of eigenvalues into eight types. Each type is marked by a number showing its multiplicity p, and abnormal eigenvalues are marked with the letter A. The symbol AA refers to a superabnormal eigenvalue μ_0, for which both $\mathbf{M}(\mu_0)$ and $\mathbf{M}'(\mu_0)$ are zero matrices, and therefore μ_0 must be a quadruple eigenvalue.

(A) Normal Eigenvalues

Type 1—A simple eigenvalue
Type 2—Normal, double eigenvalue
Type 3—Normal, triple eigenvalue
Type 4—Normal, quadruple eigenvalue

Normal eigenvalues with multiplicity p have eigenvectors and eigensolutions of orders zero through $p-1$ given by Eqs. (38) and (39).

(B) Abnormal Eigenvalues

Type 2A—Abnormal, $p=2$. μ has two independent zeroth-order eigenvectors given by the two columns of the matrix $\mathbf{J}(\mu)$, each yielding an independent eigensolution when multiplied by an arbitrary analytic function of $x+\mu y$.

Type 3A—Abnormal, $p=3$. Two eigenvectors are given by Eqs. (41a,b) and the corresponding eigensolutions are given by Eqs. (42a,b). In addition, one zeroth-order eigenvector $\boldsymbol{\xi}^{[0]}$ and eigensolution $\boldsymbol{\chi}^{[0]}$ are given by Eqs. (41*d*) and (42*d*), respectively, to ensure independence of $\boldsymbol{\xi}^{[0]}$ and $\boldsymbol{\xi}^{[1]}$.

Type 4A—An abnormal quadruple eigenvalue μ with $\mathbf{M}'(\mu)\neq\mathbf{0}$. Eigenvectors and eigensolutions are given, respectively, by the full systems of Eqs. (41) and (42).

(C) Superabnormal Eigenvalue

Type 4AA—The eigenvectors are given by the two columns of $\mathbf{J}(\mu)$ and the two columns of $\mathbf{J}'(\mu)$. The eigensolutions are given by the two columns of $\mathbf{J}(\mu)\langle f_3(\mu),f_4(\mu)\rangle$ and the two columns of Eq. (44). All isotropic plates have superabnormal eigenvalues $\pm i$, which are the double roots of $M_{11}(\mu)=0$ and also of $M_{22}(\mu)=0$, and for which M_{12} vanishes identically.

We now give a proof that, if the strain energy function $U=1/2(\boldsymbol{\kappa}^T\mathbf{m}+\boldsymbol{\epsilon}^T\mathbf{n})$ is positive definite, then the characteristic Eq. (22) has no real roots. Consider $\boldsymbol{\kappa}$ and $\mathbf{n}$ of the form

$$\{w_{,yy},w_{,xx},-w_{,xy}\}^T=\{-\mu^2,-1,\mu\}^T\boldsymbol{\kappa}_0,$$
$$\mathbf{n}=\{-\mu^2,-1,\mu\}^Tn_0.$$

Substituting into Eqs. (23*a*,*b*), and using Eq. (19), one obtains

$$2U=\boldsymbol{\kappa}^T\mathbf{m}+\mathbf{n}^T\boldsymbol{\epsilon}=\{\kappa_0,n_0\}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\mathbf{M}(\mu)\begin{Bmatrix} \kappa_0 \\ n_0 \end{Bmatrix}. \quad (45)$$

If the Eq. (22) has a real root μ, then $\mathbf{M}(\mu)\eta=\mathbf{0}$ has a nontrivial real solution $\boldsymbol{\eta}$. Let $\{\kappa_0,n_0\}^T=\boldsymbol{\eta}$. Then U vanishes for a nontrivial state. This contradicts the assumption that U is positive definite.

Applying the preceding arguments to the special cases where either κ_0 or n_0 but not both vanish, one also finds that $M_{11}(\mu)=0$ has no real roots provided that $\mathbf{D}$ is positive definite, and $M_{22}(\mu)=0$ has no real roots provided that $\mathbf{A}$ is positive definite.

Hence an anisotropic laminate has four complex conjugate pairs of eigenvalues, and they may or may not be all distinct. The multiplicity of the eigenvalues, and whether they are normal, abnormal, or superabnormal, determine the types of eigenvectors and eigensolutions belonging to the laminate. A laminate is *non-degenerate* if it has four complex conjugate pairs of *independent* zeroth-order eigenvectors. Laminates are called degenerate, extra-degenerate, and ultra-degenerate, respectively, if they require one, two, or three conjugate pairs of higher-order eigensolutions to supplement the zeroth-order eigensolutions.

The eight different types of eigenvalues and the corresponding sets of eigensolutions logically imply a classification of all anisotropic laminates into 11 mutually exclusive types, each with a distinctive representation of its general solution. To solve an equilibrium problem of a laminate using the general solution, one must first determine the type to which the laminate belongs, and use the representation of the general solution appropriate to that type.

Three of the 11 types are nondegenerate, three degenerate, four extra-degenerate, and one ultra-degenerate. Each type is characterized by a distinctive combination of eigenvalues, and only those eigenvalues with positive imaginary parts need be mentioned. The nondegenerate types include laminates with (a) four simple eigenvalues, (b) two simple and one double abnormal, and (c) two double abnormal eigenvalues. The degenerate types include those with (a) two simple and one double normal, (b) one double normal and one double abnormal and (c) one simple and one triple abnormal eigenvalue. The extra-degenerate types include laminates with (a) one simple and one triple normal, (b) two double normal, (c) one quadruple abnormal and (d) one quadruple superabnormal eigenvalue. Only laminates with a quadruple normal eigenvalue are ultra-degenerate. For each type of laminates, the complete sets of eigenvectors and eigensolutions are obtained by combining those associated with the various eigenvalues of the laminate, as described at the beginning of this section.

6 The Derivative Rule

The expressions of the higher-order eigenvectors and eigensolutions, as derived through an algebraic analysis in the present work, may be obtained in a formalistic way by differentiating appropriate *analytical expressions* of the lower-order eigenvectors and eigensolutions with respect to μ, which is regarded provisionally as a variable prior to evaluation at a specific multiple root of

the characteristic equation. This rule, which may be called the *derivative rule*, always yields correct results when applied properly.

The zeroth-order eigenvector associated with an eigenvalue μ_0 are determined by Eqs. (17) and (20). These equations yield two or one independent eigenvectors of the zeroth order, depending on whether μ_0 is an abnormal or a normal eigenvalue.

While Eqs. (7), (17), and (20) determine an eigenvector and an eigensolution only when μ is equated to a root of the characteristic equation, we shall temporarily suppress the identification of μ with a specific eigenvalue, and regard it instead as a variable. We then differentiate these equations repeatedly with respect to the variable μ to obtain

$$\boldsymbol{\chi}^{(N)}=\sum_{0\leqslant j\leqslant N}(N,j)y^j f^{(j)}(x+\mu y)\boldsymbol{\xi}^{(N-j)}, \tag{46a}$$

$$\boldsymbol{\xi}^{(j)}=\mathbf{J}(\mu)\boldsymbol{\eta}^{(j)}+(j,1)\mathbf{J}'(\mu)\boldsymbol{\eta}^{(j-1)}+(j,2)\mathbf{J}''(\mu)\boldsymbol{\eta}^{(j-2)}+(j,3)\mathbf{J}'''(\mu)\boldsymbol{\eta}^{(j-3)}, \tag{46b}$$

$$\sum_{0\leqslant j\leqslant N}(N,j)\mathbf{M}^{(j)}(\mu)\boldsymbol{\eta}^{(N-j)}=\mathbf{0}. \tag{46c}$$

The superscripts of $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, and $\boldsymbol{\chi}$, which previously appeared in bracket symbols in Eqs. (27), (35), and (36) to identify the *order* of higher eigenvectors and eigensolutions, now appear in parenthesis and denote the order of differentiation with respect to the variable μ. Otherwise, Eqs. (46a,b,c) are formally identical to Eqs. (27), (35), and (36), respectively. Therefore, in a formalistic way, repeated differentiation of Eqs. (7), (20), and (17) yields Eqs. (46*a*,*b*,*c*), which may be converted into correct expressions of the higher-order eigenvectors and eigensolutions, i.e., Eqs. (27), (35) and the solution of Eq. (36), by merely changing the *k*th derivatives of $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, and $\boldsymbol{\chi}$ into the *k*th-order vectors $\boldsymbol{\eta}^{[k]}$ and $\boldsymbol{\xi}^{[k]}$ and eigensolution $\boldsymbol{\chi}^{[k]}$. In other words, higher-order eigenvectors and eigensolutions may be obtained from the analytical expressions of the zeroth-order eigenvectors and eigensolutions, in which μ appears as a variable, by repeated differentiation, followed by evaluation of μ at the specific multiple eigenvalue. This derivative rule presents an exceedingly simple and formal routine for generating higher-order eigenvectors and eigensolutions.

Notice that the vectors $\boldsymbol{\rho}$ and $\hat{\boldsymbol{\rho}}$ are defined for the cases of normal and abnormal eigenvalues, respectively, to ensure a nonvanishing eigenvector $\mathbf{JW}\boldsymbol{\rho}$ in the first case and $\mathbf{JW}'\hat{\boldsymbol{\rho}}$ in the second case. With the exceptions of a double abnormal eigenvalue which has the two columns of $\mathbf{J}$ as eigenvectors, and a superabnormal eigenvalue which has the two columns of $\mathbf{J}'$ as additional eigenvectors, in all other cases the eigenvectors of various orders are given by $\mathbf{Jw}\boldsymbol{\rho}$ or $\mathbf{JW}\hat{\boldsymbol{\rho}}$ and their μ-derivatives. Differentiation is to be applied to the *unnormalized* zeroth-order eigenvector. Normalization, as is often done in the Stroh formalism of anisotropic elasticity, [4], yields complicated analytical expressions of eigenvectors, making the implementation of the derivative rule unduly cumbersome.

A trivial remark from a mathematical viewpoint is that the various μ-derivatives of $\mathbf{Jw}\rho$ or $\mathbf{Jw}\hat{\rho}$ must be evaluated at the specific eigenvalue *after* performing differentiation, not before. Thus the derivative rule cannot be applied to the first and second-order expressions of Eqs. (41) and (42) because the terms involving the vanishing matrix $\mathbf{W}(\mu)$ have already been discarded, so that further differentiation of these expressions is no longer legitimate.

In light of the derivative rule, the general solutions of the various types of anisotropic laminates may be written in concise forms. Let $\mathbf{Z}^+$ denote the 8×4 matrix containing the four eigenvectors $\boldsymbol{\xi}_k$ associated, respectively, with the four eigenvalues μ_k having positive real parts, $k=1, 2, 3, 4$. The eigenvalues are arranged so that simple eigenvalues precede multiple ones, and all eigenvectors associated with a common multiple eigenvalue are arranged in ascending orders. Let

$$\mathbf{f}=\{f_1(x+\mu_1 y),f_2(x+\mu_2 y),f_3(x+\mu_3 y),f_4(x+\mu_4 y)\}^T, \tag{47}$$

where the four arbitrary functions $f_1,\ldots,f_4$ are generally distinct even if their arguments involve a common eigenvalue μ_k. The general solution for all types of laminates may be expressed as follows:

$$\boldsymbol{\chi}=2\ \mathrm{Re}[\mathbf{Z}^+\mathbf{Df}], \tag{48}$$

where $\mathbf{D}$ is the identity matrix if the laminate is nondegenerate. For degenerate laminates (a) and (b), one has

$$\mathbf{D}=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&d/d\mu\\0&0&0&1\end{bmatrix}. \tag{49}$$

In the degenerate case (c), one has to replace the off-diagonal element $d/d\mu$ in Eq. (49) by $2\,d/d\mu$. In the extra-degenerate case (a), the $\mathbf{D}$-operator is given by Eq. (50*a*) below. In the extra-degenerate cases (b) and (d) it is given by Eq. (50*b*).

$$\mathbf{D}=\begin{bmatrix}1&0&0&0\\0&1&d/d\mu&d^2/d\mu^2\\0&0&1&2d/d\mu\\0&0&0&1\end{bmatrix} \tag{50a}$$

$$\mathbf{D}=\begin{bmatrix}1&d/d\mu&0&0\\0&1&0&0\\0&0&1&d/d\mu\\0&0&0&1\end{bmatrix}. \tag{50b}$$

For ultra-degenerate laminates one has

$$\mathbf{D}=\begin{bmatrix}1&d/d\mu&d^2/d\mu^2&d^3/d\mu^3\\0&1&2d/d\mu&3d^2/d\mu^2\\0&0&1&3d/d\mu\\0&0&0&1\end{bmatrix}. \tag{51}$$

Replacing the first row of the matrix in Eq. (51) by {1,0,0,0}, one obtains the $\mathbf{D}$-operator of the extra-degenerate case (c).

7 An Example: Infinite Plate With an Elliptical Hole

Consider an infinite laminate with an elliptical hole defined by the boundary curve $(x/a)^2+(y/b)^2=1$. The laminate is subject to stress and moment resultants which approach constant limits $\{F_{,yy},F_{,xx},-F_{,xy}\}=\{N_{11},N_{22},N_{12}\}$ and $\{M_y,M_x,-2M_{xy}\}=\{M_{22},M_{11},-2M_{12}\}$ at infinity. The boundary of the hole is load-free. This problem was solved by Lu and Mahrenholtz [12] for anisotropic laminates whose eigenvalues are all distinct. The solution is obtained by combining a uniform solution with the preceding limiting state of forces and moments, and a complementary solution. The force and moment components of the complementary solution decay to zero at infinity, while on the boundary of the hole they nullify the boundary forces and moments of the uniform solution. The boundary condition of the complementary problem may be given for the stress and moment potentials by integrating the uniform field of forces and moments followed by reversal of the algebraic signs. One has

$$\{F_{,y},-F_{,x},\Psi_1,\Psi_2\}^T=-y\mathbf{t}_1+x\mathbf{t}_2, \tag{52a}$$

$$\mathbf{t}_1\equiv\{N_{11},N_{12},-M_{12},M_{11}\}^T,\quad \mathbf{t}_2\equiv\{N_{12},N_{22},-M_{22},M_{12}\}^T. \tag{52b}$$

We consider the complex analytic function $\zeta=f(z)$ defined by the inverse expression

$$2z=(a+i\mu b)\zeta+(a-i\mu b)\zeta^{-1}, \tag{53}$$

where the parameter μ will be identified subsequently with an eigenvalue. Equation (53) is essentially identically to the mapping used by Lekhnitskii [2] (p. 155). Substituting

$$\zeta=e^{-i\theta} \tag{54}$$

into Eq. (53), one obtains

$$z=x+\mu y=a\cos\theta+\mu b\sin\theta. \tag{55}$$

Hence the unit circle in the ζ-plane is mapped by Eq. (53) into the boundary of the elliptical hole:

$$x=a\cos\theta, \quad y=b\sin\theta. \tag{56}$$

Differentiating Eq. (53) with respect to μ, one has

$$2b\sin\theta=ib(\zeta-\zeta^{-1})+\{(a+i\mu b)-(a-i\mu b)\zeta^{-2}\}(d\zeta/d\mu).$$

Substitution of Eq. (54) into the last equation yields $d\zeta/d\mu=0$ on the elliptical boundary. If the function implicitly defined by Eq. (53) is written as $\zeta=f(\mu,z)$, then

$$df/d\mu=\partial f/\partial\mu+y\partial f/\partial z=0 \tag{57}$$

on the boundary. Notice that

$$\partial f/\partial\mu=-ib(\zeta-\zeta^{-1})\{(a+i\mu b)-(a-i\mu b)\zeta^{2}\}^{-1},$$

$$\partial f/\partial z=1/(\partial z/\partial\zeta)=2\{(a+i\mu b)-(a-i\mu b)\zeta^{2}\}^{-1}.$$

Repeated differentiation of Eq. (53), followed by evaluation on $\zeta=e^{-i\theta}$, shows that the higher derivatives of f with respect to μ also vanish on the hole boundary. This result is important for the solution of the present problem if the laminate has repeated eigenvalues. The expressions of the higher-order eigensolutions are complicated because they contain the derivatives of f with respect to μ. This generally makes it difficult to find the appropriate combination of eigensolutions to fit the boundary data analytically. Since the derivatives of f vanish on the hole boundary, the present problem is not burdened by this difficulty.

Let

$$\mathbf{f}=\langle\boldsymbol{\zeta}_1(x+\mu_1 y),\boldsymbol{\zeta}_2(x+\mu_2 y),\boldsymbol{\zeta}_3(x+\mu_3 y),\boldsymbol{\zeta}_4(x+\mu_4 y)\rangle\mathbf{c}, \tag{58}$$

where ζ_k is defined by Eq. (53), in which the parameter μ assumes the value μ_k, $\langle\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_4\rangle$ denotes the diagonal matrix having the diagonal elements $\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_4$, and $\mathbf{c}$ is an undetermined complex constant vector. On the elliptical boundary, $d\zeta/d\mu$ vanishes so that Eqs. (49)–(51) all reduce to $\mathbf{D}=\mathbf{I}_4$. Furthermore, on the same ellipse, all $\boldsymbol{\zeta}_k$ reduce to $e^{-i\theta}$ irrespective of μ_k. Let $\mathbf{V}$ denote the 4×4 matrix obtained by removing the first two and the last two rows from the matrix $\mathbf{Z}^{+}$, i.e.,

$$\mathbf{V}=[\mathbf{0}_{4\times2},\mathbf{I}_4,\mathbf{0}_{4\times2}]\mathbf{Z}^{+}. \tag{59}$$

Then, on the hole boundary,

$$\mathbf{V}\mathbf{c}(\cos\theta-i\sin\theta)+\overline{\mathbf{V}\mathbf{c}}(\cos\theta+i\sin\theta)$$
$$=\{F_{,y},-F_{,x},\Psi_1,\Psi_2\}^{T}=-b\sin\theta\mathbf{t}_1+a\cos\theta\mathbf{t}_2.$$

One easily obtains

$$\mathbf{c}=1/2\mathbf{V}^{-1}(a\mathbf{t}_2-ib\mathbf{t}_1). \tag{60}$$

Hence the solution of the complementary problem is given by

$$\{w_{,y},-w_{,x},F_{,y},-F_{,x},\Psi_1,\Psi_2,-u,-v\}^{T}$$
$$=\mathrm{Re}[\mathbf{Z}^{+}(\mathbf{D}\langle\zeta_1,\zeta_2,\zeta_3,\zeta_4\rangle)\mathbf{V}^{-1}(a\mathbf{t}_2-ib\mathbf{t}_1)]. \tag{61}$$

Let $\mathbf{Z}_2$ and $\mathbf{Z}_3$ be the 6×4 matrices whose kth columns are obtained by premultiplying the kth column of $\mathbf{Z}^{+}$ by $\Phi(\mu_k)$ and $\mathbf{E}(\mu_k)$, respectively, $k=1,\ldots,4$ (see Eqs. (11*a*,*c*) for the definitions of Φ and $\mathbf{E}$), and let

$$\Gamma\equiv\langle\partial\zeta_1/\partial z,\partial\zeta_2/\partial z,\partial\zeta_3/\partial z,\partial\zeta_4/\partial z\rangle, \tag{62a}$$

$$\partial\zeta_k/\partial z=2\{(a+i\mu_k b)-(a-i\mu_k b)\zeta_k^{-2}\}^{-1}. \tag{62b}$$

Then, for the complementary solution,

$$\{w_{,yy},w_{,xx},-w_{,xy},F_{,yy},F_{,xx},-F_{,xy}\}^{T}$$
$$=\mathrm{Re}[\mathbf{Z}_2(\mathbf{D}\Gamma)\mathbf{V}^{-1}(a\mathbf{t}_2-ib\mathbf{t}_1)], \tag{63a}$$

$$\{M_y,M_x,-2M_{xy},-\epsilon_x,-\epsilon_y,-2\epsilon_{xy}\}^{T}$$
$$=\mathrm{Re}[\mathbf{Z}_3(\mathbf{D}\Gamma)\mathbf{V}^{-1}(a\mathbf{t}_2-ib\mathbf{t}_1)]. \tag{63b}$$

Notice that, for various degenerate laminates, the differential operator $\mathbf{D}$ reduces to the identity matrix $\mathbf{I}_4$ only on the elliptical boundary but not in the interior region of the plate. Hence the interior solutions of degenerate laminates have more complicated forms. Notice also that Eq. (53) is a quadratic equation for $\boldsymbol{\zeta}$ which can be solved explicitly in terms of μ and z. Therefore, the solutions given by Eqs. (61) and (63*a*,*b*) are completely explicit. Finally, the differential operator $\mathbf{D}$ for generating higher-order derivatives involves the *total* derivatives of the functions $\boldsymbol{\zeta}_k$ associated with multiple eigenvalues, and these total derivatives have vanishing boundary values for the present problem. In contrast, the diagonal matrix $\boldsymbol{\Gamma}$ in Eqs. (62) and (63) involves *partial* derivatives $\partial\boldsymbol{\zeta}_k/\partial z$, which do not vanish on the boundary.

In the solution of Ref. [12], the boundary conditions were satisfied by fitting the in-plane normal and shearing forces, the normal moment and the out-of-plane shearing force. This involves expressions requiring the spatial derivatives of the functions $\boldsymbol{\zeta}_k$. The task is considerably simplified in the present analysis by fitting the moment potentials and the two components of ∇F, which require only the boundary values of $\boldsymbol{\zeta}_k$ but not their normal derivatives.

For laminates with a vanishing coupling stiffness matrix $\mathbf{B}$ (and therefore $\mathbf{B}^{*}=\mathbf{0}$), the bending solution can be obtained separately from the in-plane solution. Consider the latter problem. Equation (17) reduces to the scalar equation $M_{22}(\mu)=0$, which has either two distinct pairs of complex conjugate roots or one conjugate pair of double roots. For the first case, there are two independent eigenvectors whose first two components $\{F_{,y},-F_{,x}\}$ are given by $\{-\mu_1,1\}$ and $\{-\mu_2,1\}$, respectively. For the second (degenerate) case, they are given instead by $\{-\mu_1,1\}$ and $\{-1,0\}$. Equation (61) yields the solution of the complementary problem, which assumes the following form for the nondegenerate and degenerate cases, respectively,

$$\begin{Bmatrix}F_{,y}\\-F_{,x}\end{Bmatrix}=\mathrm{Re}\left[(\mu_2-\mu_1)^{-1}\begin{bmatrix}-\mu_1&-\mu_2\\1&1\end{bmatrix}\langle\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2\rangle\begin{bmatrix}1&\mu_2\\-1&-\mu_1\end{bmatrix}\times\left(a\begin{Bmatrix}N_{12}\\N_{22}\end{Bmatrix}-ib\begin{Bmatrix}N_{11}\\N_{12}\end{Bmatrix}\right)\right] \tag{64a}$$

$$\begin{Bmatrix}F_{,y}\\-F_{,x}\end{Bmatrix}=\mathrm{Re}\left[\begin{bmatrix}-\mu_1&-1\\1&0\end{bmatrix}\begin{bmatrix}\zeta_1&d\boldsymbol{\zeta}_1/d\mu\\0&\boldsymbol{\zeta}_1\end{bmatrix}\begin{bmatrix}0&1\\-1&-\mu_1\end{bmatrix}\times\left(a\begin{Bmatrix}N_{12}\\N_{22}\end{Bmatrix}-ib\begin{Bmatrix}N_{11}\\N_{12}\end{Bmatrix}\right)\right]. \tag{64b}$$

It is interesting that this solution depends on the elastic properties of the laminate through the eigenvalues only. In particular, for all isotropic plates, Eq. (64*b*) with $\mu_1=\mu_2=i$ yields stress solutions that are independent of the elastic moduli, a fact which is known to be valid if the traction data on the hole boundary have vanishing resultant forces and moments.

8 Summary and Concluding Remarks

The problem of obtaining general equilibrium solutions of anisotropic laminated plates is completely solved through reduction to an eigenvalue problem associated with a 2×2 matrix function $\mathbf{M}(\mu)$. A fundamental difference that sets the present analytical formulation apart from the Stroh-type formalism of anisotropic

elasticity is the choice of the primary set of unknown variables. In contrast to the stress-function-based approach of Lekhnitskii and Muskhelishvili, the Stroh formalism starts from the equilibrium equations governing the displacement functions, [13]. Many fundamental results have been obtained, historically, through the use of this formalism. Yet the expressions of the eigenvectors, which were obtained by the Lekhnitskii formalism in such simple and explicit forms, become considerably more complicated in the Stroh formalism. The reason is simple. The differential equations governing the stress functions have coefficients that are the elastic compliances, which are taken as the constitutive parameters in the Lekhnitskii formalism. In contrast, the Stroh formalism uses the elastic stiffness. While the eigenvectors have simple analytical expressions in terms of the compliance coefficients, they become unduly complicated in terms of the stiffness parameters.

The algebraic complexity of the stiffness-based formulation becomes more acute in the degenerate, extra-degenerate, and ultra-degenerate cases, where the zeroth-order eigensolutions must be supplemented by higher-order ones, and the latter have more complicated expressions involving the lower-order eigenvectors as well. A comparison of Section 2 of this paper with Sections 3 and 4 shows that the determination of the general solution is much simpler for nondegenerate laminates treated in existing studies, than for the various degenerate laminates that are investigated fully in this paper. It is shown that the latter results may be obtained from the former in a formalistic way by proper use of the derivative rule. But the implementation of this rule becomes more cumbersome, due to the complexity of the analytical expression of the zeroth-order eigenvectors in the Stroh formalism. Although many important findings concerning the algebraic structure of the solutions of anisotropic elasticity were made in the context of the Stroh formalism, most of these same results can be obtained in a simpler way, and expressed in simpler forms, by using the compliance-based formalism, [11].

The present analysis of anisotropic laminates is a mixed formulation using the curvatures and the in-plane forces as the primary variables. The moments and in-plane displacements or strains are treated as secondary unknowns related to ϕ through the constitutive Eq. (14), which involves the new elasticity matrices $\mathbf{A}^*$, $\mathbf{B}^*$, and $\mathbf{D}^*$. The reason for this choice is that the first four elements of $\boldsymbol{\chi}$ are the components of gradients of w and F. They imply Eq. (8) which, together with the constitutive relation, reduce the determination of eigensolutions to a two-dimensional eigenvalue problem associated with a 2×2 symmetric matrix function $\mathbf{M}(\mu)$. Cases involving multiple and abnormal eigenvalues are treated with some modification in the analysis.

If a different choice of the primary unknown variables is made, then one or both relations of Eq. (8) cannot be used to reduce the dimension of the eigenvalue problem to two. The Stroh formalism has been used to derive an eigenvalue problem associated with a 6×6 matrix function, [12]. The eigenvectors were not given in explicit analytical form, due to the intrinsic complexity of the expressions in terms of the conventional stiffness matrices. Although they may be obtained numerically, the resulting general solution is valid for a nondegenerate laminate but not for degenerate, extradegenerate and ultra-degenerate laminates. An extension of that formalism to the latter cases would entail algebraic operations which are ultimately equivalent to applying the derivative rule to the adjoint matrix of a 4×4 matrix function. But the extension is achieved much more easily in the present formulation. Finally, Eqs. (23*a*,*b*) show that the strain energy function of laminated plates assumes a particularly simple form in terms of the matrices $\mathbf{A}^*$ and $\mathbf{D}^*$. It does not involve $\mathbf{B}^*$ (but includes the bending-stretching effect through $\mathbf{D}^*$), and it contains the second derivatives of the two scalar functions $w(x,y)$ and $F(x,y)$ as the arguments.

Anisotropic laminates are classified into 11 distinct types corresponding to the various forms of the general solution. The classification is based on the multiplicity of the eigenvalues and whether a multiple eigenvalue μ is normal ($\mathbf{M}(\mu)\neq\mathbf{0}$), abnormal ($\mathbf{M}(\mu)=\mathbf{0}$) or superabnormal ($\mathbf{M}(\mu)=\mathbf{M}'(\mu)=\mathbf{0}$). The elements of $\mathbf{M}(\mu)$ are defined directly in terms of the elasticity matrices. $\mathbf{A}^*$, $\mathbf{B}^*$, and $\mathbf{D}^*$ according to Eq. (19). If the conventional stiffness matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$, or another set of elasticity matrices is used instead, then the conditions of abnormality and superabnormality assume more complex expressions. This may lead to indirect characterizations of the various types of laminates, more complicated representations of solutions, and a less transparent classification.

Because the classification of the laminates is based purely on mathematical criteria, each class does not correspond to a peculiar set of physical characteristics. For example, all homogeneous isotropic laminates belong to, but form only a small subset of, the class of extra-degenerate laminates that have a quadruple, superabnormal eigenvalues μ_0 (this is one of the 11 classes listed at the end Section 5 and designated as extra-degenerate (d)). This class is defined only by the conditions $\mathbf{M}(\mu_0)=\mathbf{M}'(\mu_0)=\mathbf{0}$ on the matrix function $\mathbf{M}(\mu)$ of Eq. (18), where μ_0 need not be equal to $\pm i$. Even in the subclass where the superabnormal eigenvalues are $\pm i$, the preceding conditions impose only two restrictions on the nine elements of the coupling matrix $\mathbf{B}^*$:

$$B_{11}^*+B_{22}^*-(B_{66}^*+B_{12}^*+B_{21}^*)=0, \quad B_{23}^*+B_{32}^*-(B_{13}^*+B_{31}^*)=0,$$

and similar restrictions on the symmetric matrices $\mathbf{A}^*$ and $\mathbf{D}^*$. Isotropic laminates are defined by much more stringent conditions on the elasticity matrices.

Despite their lack of association with specific types of material symmetry properties and other physical characteristics, the various classes of laminates and their peculiar eigenvalues and eigensolutions are needed for correct representations of general solutions, Green's functions, and the integrals arising in boundary element methods. A general computational code for anisotropic laminates must include the proper representations of solutions if it is expected to work for nondegenerate as well as degenerate laminates (including isotropic and quasi-isotropic ones), because in the latter cases the number of independent zeroth-order eigensolutions is insufficient and higher-order eigensolutions must be used to provide a complete representation.

A detailed study of the algebraic structure of the solution space of anisotropic laminates, including orthogonal eigenspaces, pseudometrics, invariant tensors, and isomorphisms is presented in a companion paper, [14].

References

[1] Lekhnitskii, S. G., 1968, *Anisotropic Plates*, Gordon and Breach, New York.
[2] Lekhnitskii, S. G., 1963, *Theory of Elasticity of an Anisotropic Body*, Holden Day, San Francisco.
[3] Muskhelishvili, N. I., 1954, *Some Basic Problems of the Mathematical Theory of Elasticity*, Nordhoff, Groningen.
[4] Ting, T. C. T., 1966, *Anisotropic Elasticity—Theory and Applications*, Oxford University Press, New York.
[5] Yin, W.-L., 1997, "A General Analysis Method for Singularities in Composite Structures," *Proc. AIAA/ASME/ASCE/AHS/ASC 38th SDM Conference*, Apr. 7–10, Kissimee, FL., AIAA, Washington, DC, pp. 2238–2246.
[6] Yin, W.-L., 2001, "Anisotropic Elasticity and Multi-Material Singularities," *Contemporary Research in Engineering Mechanics*, G. A. Kardomateas and V. Birman, eds., ASME, New York, ASME-AMD-249, pp. 117–128.
[7] Banerjee, P. K., and Butterfield, R., 1981, *Boundary Element Methods in Engineering Science*, McGraw-Hill, London.
[8] Rizzom, F. H., and Shippy, D. J., 1970, "A Method for Stress Determination in Plane Anisotropic Bodies," J. Compos. Mater., **4**, pp. 36–61.
[9] Tsai, S. W., and Hahn, H. T., 1980, *Introduction to Composite Materials*, Technomic, Lancaster, PA.
[10] Yin, W.-L., 2000, "Deconstructing Plane Anisotropic Elasticity, Part I: The Latent Structure of Lekhnitskii's Formalism," Int. J. Solids Struct., **37**, pp. 5257–5276.
[11] Yin, W.-L., 2000, "Deconstructing Plane Anisotropic Elasticity, Part II: Stroh's Formalism Sans Frills," Int. J. Solids Struct., **37**, pp. 5277–5296.
[12] 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.
[13] Stroh, A. N., 1962, "Steady State Problems in Anisotropic Elasticity," J. Math. Phys., **41**, pp. 77–103.
[14] Yin, W.-L., 2003, "Structure and Properties of the Solution Space of General Anisotropic Laminates," Int. J. Solids Struct., **40**, pp. 1825–1852.

Concepts of Separated *J*-Integrals, Separated Energy Release Rates, and the Component Separation Method of the *J*-Integral for Interfacial Fracture Mechanics

T. Nishioka
Professor,
Mem. ASME
e-mail: nishioka@cc.kshosen.ac.jp

S. Syano
Graduate Student

T. Fujimoto
Lecturer

Department of Ocean Mechanical Engineering,
Kobe Universityof Mercantile Marine,
5-1-1 Fukae Minamimachi,
Higashinada-ku,
Kobe 658-0022, Japan

First, this paper presents the concepts of separated J-integrals and separated energy release rates. The path-independent separated J-integrals have the physical significance of energy flows into an interfacial crack tip from adjacent individual material sides or, equivalently, separated energy release rates. Thus, the J-integral and the energy release rate can be evaluated by the sum of the path-independent separated J-integrals. Second, the relations between the separated J-integrals and the stress intensity factors are derived. Third, the component separation method of the J-integral is extended for interfacial crack problems to allow accurate evaluation of the stress intensity factors. Finally, pertinent numerical analyses are carried out to demonstrate the usefulness of the separated J-integrals and the component separation method. [DOI: 10.1115/1.1576803]

1 Introduction

The establishment of interfacial fracture mechanics is of great importance, because in heterogeneous materials containing distinct interfaces most failures occur at these interfaces. For this reason, extensive research has been done on interfacial fracture problems and continues up to the present day. Interfacial fracture mechanics is founded mainly on the solutions for interfacial cracks. Because of the mismatch between the materials on the neighboring sides of interfaces, interfacial cracks inherently produce mixed-mode states. Thus, in interfacial fracture mechanics the evaluation of mixed-mode states is mandatory.

Nishioka and Yasin [1] recently proposed the concept of separated dynamic *J*-integrals (equivalent to separated dynamic energy release rates) for dynamic interfacial fracture mechanics. The separated dynamic *J*-integrals have the following salient features:

i. The separated dynamic *J*-integrals can be expressed in vector form, [1]. Thus they can be decomposed into the components of any coordinate system.
ii. The separated dynamic *J*-integrals can be expressed in path independent forms, [1,2].
iii. The components of the separated dynamic *J*-integrals parallel to the crack direction have the physical significance of energy flows into a propagating interfacial crack tip from the individual material sides or, equivalently, the separated dynamic energy release rates, [1].
iv. The sum of the separated dynamic *J* integrals corresponds to the dynamic *J*-integral derived by Nishioka and Atluri [2], which has the physical significance of the dynamic energy release rate, [3].

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, May 31, 2001; final revision, Dec. 19, 2002. Associate Editor: B. M. Moran. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

In the consideration of static interfacial fracture mechanics presented in this paper, the concepts of the separated (static) *J*-integral and separated (static) energy release rate are readily developed from the separated dynamic *J*-integral, because under static conditions the dynamic *J*-integral naturally reduces to the static *J*-integral derived by Rice [4]. Furthermore, the relations between the separated *J*-integrals and the stress intensity factors are obtained using the analytical solutions for the stress and displacement fields of an interfacial crack, which were derived by Sun and Jih [5].

In addition to the *J*-integral and the energy release rate, the stress intensity factor is also widely used in fracture mechanics. In order to accurately evaluate mixed-mode stress intensity factors, Nishioka and co-workers, [6–8], developed the component separation method of the dynamic *J*-integral for various crack problems in homogeneous materials. In this paper, the component separation method of the *J*-integral is developed to accurately and conveniently evaluate mixed-mode stress intensity factors for static interfacial cracks. The component separation method of the dynamic *J*-integral for evaluating mixed-mode stress intensity factors of dynamic interfacial cracks will be reported elsewhere, [9].

An alternative formulation to that proposed here was put forward by Yau and Wang [10], who proposed the M_1 integral method for evaluating the mixed-mode stress intensity factors of an interfacial crack. In this method, the analytical solution for an appropriate auxiliary problem is necessary. However, the auxiliary solution field is often difficult or impossible to construct for complicated interface problems such as curved interfaces. In contrast, the component separation method does not require fictitious auxiliary solution fields, which represents a great advantage of the present method over the M_1 integral method.

To demonstrate the applicability of the separated *J*-integral and the component separation method, numerical analyses for pertinent interfacial crack problems are also carried out.

2 Separated *J*-Integrals

Now we consider a crack along a bimaterial interface (see Fig. 1), this interface may be curved or straight. The local coordinate

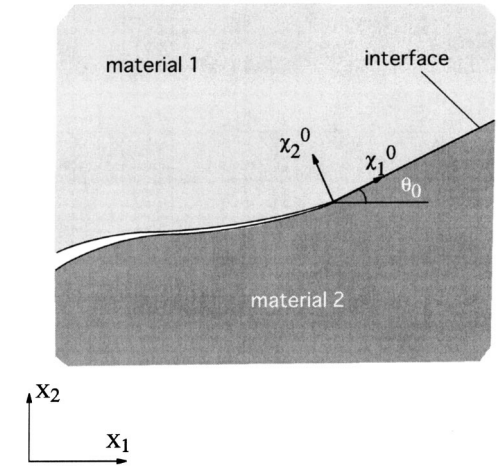


Fig. 1 An interfacial crack in a nonhomogeneous material

system x_k^0 is considered at the interfacial crack tip. The angle θ_0 measured from the global X_1-axis denotes the direction of the interfacial crack.

For a static interfacial crack, the separated (static) *J*-integrals can be derived by eliminating all inertia effects from the separated dynamic *J*-integrals, [1], to give

$$J_k^{(m)} = \lim_{\Gamma_\varepsilon^{(m)} \to 0} \int_{\Gamma_\varepsilon^{(m)}} [W n_k - t_i u_{i,k}] dS \tag{1a}$$

$$= \int_{\Gamma^{(m)} + \Gamma_c^{(m)} + \Gamma_I^{(m)}} [W n_k - t_i u_{i,k}] dS \quad (m = 1,2), \tag{1b}$$

where the superscript (m) denotes the material number, W is the strain energy density, and $(\,)_{,k} = \partial(\,)/\partial X_k$. The integral paths are defined in Fig. 2. $\Gamma_\varepsilon^{(m)}$, $\Gamma^{(m)}$, $\Gamma_c^{(m)}$, and $\Gamma_I^{(m)}$ are the separated paths for the near-field, far-field, crack surface, and interface integrals, respectively. In Eq. (1), the separated *J*-integrals are expressed in the global coordinate system.

The crack-axis components of the separated *J*-integral $J_k^{0(m)}$ can be evaluated by the coordinate transformation

$$J_k^{0(m)} = \alpha_{kl} J_l^{(m)} \quad (m = 1,2), \tag{2}$$

where α_{kl} is the coordinate transformation tensor. The crack-axis components of the separated *J*-integrals $J_k^{0(m)}$ $(m = 1,2)$ can also

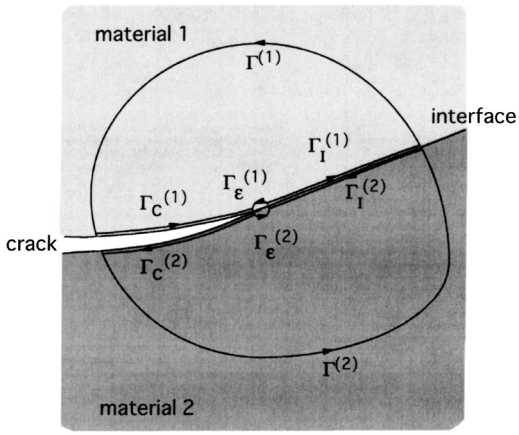


Fig. 2 Definition of integral paths for an interfacial crack

be evaluated directly by the following equation, when all components in the integrands of Eqs. (1*a*) and (1*b*) are expressed in the crack-tip coordinate system x_k^0, or when $\theta_0 = 0$:

$$J_k^{0(m)} = \lim_{\Gamma_\varepsilon^{(m)} \to 0} \int_{\Gamma_\varepsilon^{(m)}} [W n_k - t_i u_{i,k}] dS \tag{3a}$$

$$= \int_{\Gamma^{(m)} + \Gamma_c^{(m)} + \Gamma_I^{(m)}} [W n_k - t_i u_{i,k}] dS \quad (m = 1,2). \tag{3b}$$

Similarly to the separated dynamic *J*-integral, [1], the crack-axis component of the separated *J*-integral $J_1^{0(m)}$ has the physical significance of an energy flow into the interfacial crack tip from the material *m*, passing through the separated near-field path $\Gamma_\varepsilon^{(m)}$. In other words, $J_1^{0(m)}$ is equivalent to the separated energy release rate $G^{(m)}$ from the material *m*.

The *J*-integral and the energy release rate can easily be obtained by the sum of the separated *J*-integrals:

$$J_k'^0 = \sum_{m=1}^{2} J_k'^{0(m)}. \tag{4}$$

In numerical analyses, the following far-field expression for the separated *J*-integrals is convenient after taking the limit of $\Gamma_\varepsilon \to 0$.

$$J_k^{(m)} = \int_{\Gamma^{(m)} + \Gamma_c^{(m)} + \Gamma_I^{(m)}} [W n_k - t_i u_{i,k}] dS, \quad (m = 1,2). \tag{5}$$

3 Separated Energy Release Rates

Using the concept of the crack closure integral, the separated energy release rates can also be derived by

$$G^{(m)} = \lim_{\Delta a \to 0} \left\{ \frac{1}{2\Delta a} \int_0^{\Delta a} [t_i^{(m)}(a)(u_i^{(m)}(a + \Delta a) - u_i^c(a + \Delta a))|_{x_2^0 = 0}] dS \right\} \quad (m = 1,2), \tag{6}$$

where $t_i^{(m)}$ is the traction acting on the material-*m* side along the interface line of $a \leqslant x_1^0 \leqslant a + \Delta a$, $u_i^{(m)}$ is the crack-face displacement of the material m $(m = 1,2)$, and u_i^c is the x_i^0 displacement component at the crack tip.

Furthermore, the separated energy release rates $G^{(m)}$ $(m = 1,2)$ can be evaluated from the energy balance (see Fig. 3), as

$$G^{(m)} = \frac{1}{B} \frac{dF^{(m)}}{da} = \frac{1}{B} \left(\frac{d\overline{P}^{(m)}}{da} + \frac{d\overline{P}_I^{(m)}}{da} - \frac{d\overline{W}^{(m)}}{da} \right) \quad (m = 1,2), \tag{7}$$

where B is the thickness of the body, and $F^{(m)}$ and $\overline{W}^{(m)}$ are the fracture energy provided from material *m* and the strain energy in material *m*, respectively. $\overline{P}_I^{(m)}$ is the input energy from the applied loads at the material-*m* side and $\overline{P}_I^{(m)}$ is the work done by the traction of the other side material through the interface.

The rates of these input energies to the material *m* can be evaluated by

$$\frac{d\overline{P}^{(m)}}{da} = \int_{\partial V^{(m)} - S_I^{(m)}} t_i \frac{du_i}{da} dS \tag{8}$$

and

$$\frac{d\overline{P}_I^{(m)}}{da} = \int_{S_I^{(m)}} t_i \frac{du_i}{da} dS \tag{9}$$

where $\partial V^{(m)}$ $(= S_t^{(m)} + S_u^{(m)} + S_c^{(m)} + S_I^{(m)})$ denotes the entire boundary of the region occupied by the material *m*, and $S_I^{(m)}$ de-

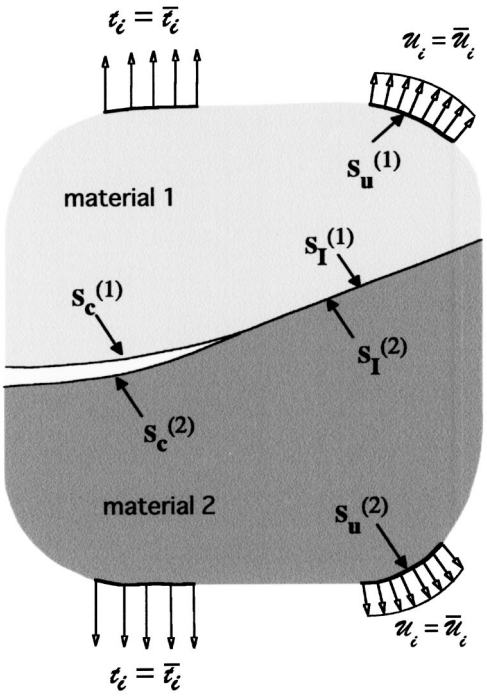


Fig. 3 A bimaterial with an interfacial crack

notes the entire uncracked interface at the material-*m* side. These integral paths are defined in Fig. 3. Since the continuity conditions along the uncracked interface can be written as

$$u_i^{(1)} = u_i^{(2)} \tag{10a}$$

and

$$t_i^{(1)} + t_i^{(2)} = 0, \tag{10b}$$

the following relation holds:

$$\frac{d\bar{P}_I^{(1)}}{da} + \frac{d\bar{P}_I^{(2)}}{da} = 0. \tag{11}$$

Therefore the sum of the separated energy release rates naturally becomes the total energy release rate G:

$$G^{(1)} + G^{(2)} = \frac{1}{B}\left\{\frac{d(\bar{P}^{(1)} + \bar{P}^{(2)})}{da} - \frac{d(\bar{W}^{(1)} + \bar{W}^{(2)})}{da}\right\} = G. \tag{12}$$

4 Near-Tip Field of an Interfacial Crack

In the subsequent sections we derive the relations between the separated *J*-integrals and the stress intensity factors. However, in the literature there are several definitions of the stress intensity factors for an interfacial crack tip. In this section we therefore summarize the near-tip stress and displacement fields and the definition of stress intensity factors used in this study.

In this paper, the following definitions, [11,12], of the stress intensity factors K_1 and K_2 are employed:

$$[\sigma_{22} + i\sigma_{12}]_{\theta=0} = \frac{K_1 + iK_2}{\sqrt{2\pi r}}\left(\frac{r}{l}\right)^{i\varepsilon}, \tag{13}$$

where the polar coordinate system (r, θ) is defined in Fig. 4. The parameter l is the characteristic length that normalizes the oscillatory singular term, and is usually taken as the entire crack length $(l = 2a)$. The parameter ε is the bimaterial constant, given by

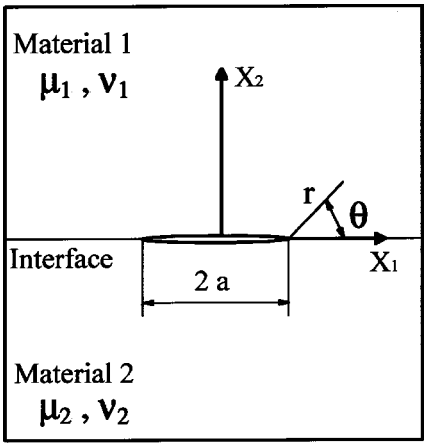


Fig. 4 Interfacial crack

$$\varepsilon = \frac{1}{2\pi}\ln\left\{\frac{\kappa^{(1)}/\mu^{(1)} + 1/\mu^{(2)}}{\kappa^{(2)}/\mu^{(2)} + 1/\mu^{(1)}}\right\}, \tag{14}$$

$$\kappa^{(m)} = \left\{\begin{array}{ll} 3 - 4\nu^{(m)} & (\text{plane strain}) \\ (3 - \nu^{(m)})/(1 + \nu^{(m)}) & (\text{plane stress}) \end{array} \quad (m = 1,2)\right\}, \tag{15}$$

where $\nu^{(m)}$ and $\mu^{(m)}$ denote Poisson's ratio and the shear modulus of the material *m*, respectively.

The explicit near-tip stress and displacement expressions for an interfacial crack were derived by Sun and Jih [5]. However, their definitions of the stress intensity factors differ from Eq. (13). Using the stress intensity factors defined by Eq. (13), the asymptotic in-plane stress components for the material *m* side can be expressed as follows:

$$\sigma_{ij}^{(m)} = \frac{1}{2\sqrt{2\pi r}\cosh(\pi\varepsilon)}\left\{K_1 f_{1_{ij}}^{(m)}\left(\theta, \ln\frac{r}{l}, \varepsilon\right) + K_2 f_{2_{ij}}^{(m)}\left(\theta, \ln\frac{r}{l}, \varepsilon\right)\right\}, \quad (ij = 11,22,12). \tag{16}$$

Detailed expressions of the stress components are given in the Appendix. Similarly the in-plane displacement components for the material m side are expressed by

$$u_i^{(m)} = \frac{\sqrt{2\pi r}}{4\pi\mu^{(m)}\cosh(\pi\varepsilon)}\left\{K_1 g_{1_i}^{(m)}\left(\theta, \ln\frac{r}{l}, \varepsilon, \kappa^{(m)}\right) + K_2 g_{2_i}^{(m)}\left(\theta, \ln\frac{r}{l}, \varepsilon, \kappa^{(m)}\right)\right\}, \quad (i = 1,2). \tag{17}$$

Detailed expressions of the displacement components are also given in the Appendix.

On the other hand, the out-of-plane stress components are given in Ref. [13] as

$$\sigma_{13} = -\frac{K_3}{\sqrt{2\pi r}}\sin\frac{\theta}{2}, \tag{18a}$$

$$\sigma_{23} = +\frac{K_3}{\sqrt{2\pi r}}\cos\frac{\theta}{2}, \tag{18b}$$

for both material sides, where K_3 is the stress intensity factor for out-of-plane deformation. Similarly the out-of-plane displacement is expressed by

$$u_3^{(m)}=\frac{2K_3}{\mu^{(m)}}\sqrt{\frac{r}{2\pi}}\sin\frac{\theta}{2}\quad (m=1,2). \tag{19}$$

It is noted that, because of the oscillatory singularity in the asymptotic stress field, the in-plane stress intensity factors K_1 and K_2 cannot be interpreted as the mode I and mode II stress intensity factors K_{I} and K_{II} in a homogeneous material. In contrast to this, K_3 is equivalent with the mode III stress intensity factor, in the case of an isotropic bimaterial.

5 Relations Between the Separated Energy Release Rates and the Stress Intensity Factors

For the in-plane deformation, Eq. (6) can be rewritten as

$$G^{(1)}=+\lim_{\Delta a\to 0}\frac{1}{2\Delta a}\int_0^{\Delta a}\{\sigma_{22}^{(1)}|_{\theta=0}(x_1^0)\cdot u_2^{(1)}|_{\theta=+\pi}(\Delta a-x_1^0)+\sigma_{12}^{(1)}|_{\theta=0}(x_1^0)\cdot u_1^{(1)}|_{\theta=+\pi}(\Delta a-x_1^0)\}dx_1^0, \tag{20a}$$

$$G^{(2)}=-\lim_{\Delta a\to 0}\frac{1}{2\Delta a}\int_0^{\Delta a}\{\sigma_{22}^{(2)}|_{\theta=0}(x_1^0)\cdot u_2^{(2)}|_{\theta=-\pi}(\Delta a-x_1^0)+\sigma_{12}^{(2)}|_{\theta=0}(x_1^0)\cdot u_1^{(2)}|_{\theta=-\pi}(\Delta a-x_1^0)\}dx_1^0. \tag{20b}$$

And for the out-of-plane (mode III) deformation, Eq. (6) can be rewritten as

$$G^{(1)}=+\lim_{\Delta a\to 0}\frac{1}{2\Delta a}\int_0^{\Delta a}\sigma_{23}^{(1)}|_{\theta=0}(x_1^0)\cdot u_3^{(1)}|_{\theta=+\pi}(\Delta a-x_1^0)dx_1^0, \tag{21a}$$

$$G^{(2)}=-\lim_{\Delta a\to 0}\frac{1}{2\Delta a}\int_0^{\Delta a}\sigma_{23}^{(2)}|_{\theta=0}(x_1^0)\cdot u_3^{(2)}|_{\theta=-\pi}(\Delta a-x_1^0)dx_1^0. \tag{21b}$$

Then, substituting Eqs. (16) and (17) into Eqs. (20) and (21), the following relations between the separated energy release rates and the stress intensity factors are obtained: For in-plane deformation,

$$G^{(m)}=\frac{e^{\pi\varepsilon}(1+\kappa^{(m)})(K_1^2+K_2^2)}{8\mu^{(m)}(1+e^{2\pi\varepsilon})\cosh(\pi\varepsilon)}\quad (m=1,2), \tag{22}$$

and for out-of-plane deformation,

$$G^{(m)}=\frac{K_3^2}{4\mu^{(m)}}\quad (m=1,2). \tag{23}$$

Thus, the sum of the separated energy release rates for in-plane deformation (see Eq. (22)) becomes

$$G=G^{(1)}+G^{(2)}=\frac{1}{16\cosh^2(\pi\varepsilon)}\left\{\frac{\kappa^{(1)}+1}{\mu^{(1)}}+\frac{\kappa^{(2)}+1}{\mu^{(2)}}\right\}(K_1^2+K_2^2). \tag{24}$$

This agrees with the relation between the (total) energy release rate and stress intensity factors obtained by Malyshev and Salganik [13]. Furthermore, the sum of the separated energy release rates for out-of-plane deformation (see Eq. (23)) is

$$G=G^{(1)}+G^{(2)}=\frac{K_3^2}{4}\left(\frac{1}{\mu^{(1)}}+\frac{1}{\mu^{(2)}}\right). \tag{25}$$

This is in agreement with the relation between the (total) energy release rate and the stress intensity factor obtained by Willis [14].

6 Relations Between the Separated *J*-Integrals and the Stress Intensity Factors

Using the asymptotic near-field solutions given by Eqs. (16)–(19), the separated *J*-integrals can be related to the stress intensity factors. If we use a circular path for the evaluation of the near-field separated *J*-integrals, as shown in Fig. 5, Eq. (1*a*) can be rewritten as

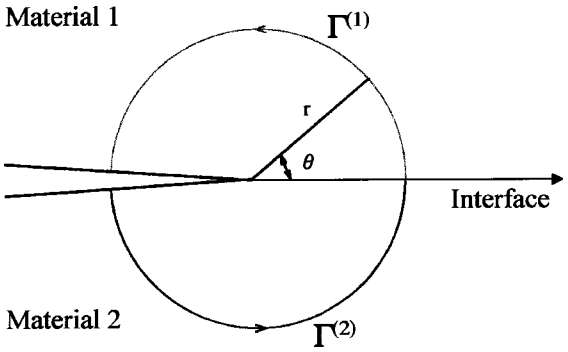


Fig. 5 Semicircular near-field paths for the separated *J*-integrals

$$J_k^{0(1)}=\lim_{r\to 0}\int_0^{\pi}[Wn_k-t_iu_{i,k}]rd\theta, \tag{26a}$$

$$J_k^{0(2)}=\lim_{r\to 0}\int_{-\pi}^{0}[Wn_k-t_iu_{i,k}]rd\theta. \tag{26b}$$

6.1 Tangential Component of the Separated *J*-Integral $J_1^{0(m)}$. Using the asymptotic solutions for the in-plane deformation, i.e., Eqs. (16) and (17) in Eqs. (26*a*) and (26*b*) with $k=1$, the relations between the separated *J*-integrals and the stress intensity factors K_1 and K_2 can be derived as

$$J_1^{0(1)}=\frac{(1+\kappa^{(1)})(K_1^2+K_2^2)}{16\mu^{(1)}\cosh^2(\pi\varepsilon)}, \tag{27a}$$

$$J_1^{0(2)}=\frac{(1+\kappa^{(2)})(K_1^2+K_2^2)}{16\mu^{(2)}\cosh^2(\pi\varepsilon)}. \tag{27b}$$

These agree with the relations between the separated energy release rates and stress intensity factors given by Eq. (22). Similarly, for the out-of-plane deformation, substituting Eqs. (18) and (19) into Eqs. (26*a*) and (26*b*) gives

$$J_1^{(1)}=\frac{K_3^2}{4\mu^{(1)}}, \tag{28a}$$

$$J_1^{(2)}=\frac{K_3^2}{4\mu^{(2)}}. \tag{28b}$$

These agree with the relations between the separated energy release rates and stress intensity factors given by Eq. (23).

Since the separated *J*-integrals and the separated energy release rates represent the energies supplied from the individual material sides to the tip of the actually extending crack or virtually extending crack, the sum of the separated quantities becomes the *J*-integral or the total energy release rate. Thus, we have

$$J_1^0=J_1^{0(1)}+J_1^{0(2)}=(J_1^{(1)}+J_1^{(2)})\cos\theta_0+(J_2^{(1)}+J_2^{(2)})\sin\theta_0=G=G^{(1)}+G^{(2)}. \tag{29}$$

For in-plane deformation, Eqs. (27*a*) and (27*b*) can be used to obtain the *J*-integral as

$$J_1^{(0)}=J_1^{0(1)}+J_1^{0(2)}=\frac{1}{16\cosh^2(\pi\varepsilon)}\left(\frac{\kappa^{(1)}+1}{\mu^{(1)}}+\frac{\kappa^{(2)}+1}{\mu^{(2)}}\right)\times(K_1^2+K_2^2). \tag{30}$$

This agrees with Eq. (24). Thus, we can confirm the relations given in Eq. (29), and the relation between the energy release rate and the stress intensity factors, as obtained by Malyshev and Salgnik [13] (see Eq. (24)). Furthermore, for out-of-plane deformation, the J-integral from Eqs. (28a) and (28b) is expressed by

$$J_1^{(0)} = J_1^{0(1)} + J_1^{0(2)} = \frac{K_3^2}{4}\left(\frac{1}{\mu^{(1)}} + \frac{1}{\mu^{(2)}}\right). \quad (31)$$

This agrees with both Eq. (25) and the relation between the energy release rate and the stress intensity factor obtained by Willis [14].

The ratio of the energies supplied to the crack tip from the individual material sides seems to be very useful in interfacial fracture mechanics. This ratio can be evaluated from the separated J-integrals. For the in-plane deformation, the ratios of the separated quantities to their total can be expressed by

$$\frac{J_1^{0(m)}}{J_1^0} = \frac{G^{(m)}}{G} = \left(\frac{\kappa^{(m)}+1}{\mu^{(m)}}\right) \Big/ \left(\frac{\kappa^{(1)}+1}{\mu^{(1)}} + \frac{\kappa^{(2)}+1}{\mu^{(2)}}\right) \quad (m=1,2). \quad (32)$$

Consequently, the ratio of the separated J-integrals or separated energy release rates can be expressed by

$$\frac{J_1^{0(1)}}{J_1^{0(2)}} = \frac{G^{(1)}}{G^{(2)}} = \frac{(\kappa^{(1)}+1)\mu^{(2)}}{(\kappa^{(2)}+1)\mu^{(1)}} = \begin{cases} \dfrac{E^{(2)}}{E^{(1)}} & \text{(plane stress)} \\ \dfrac{(1-\nu^{(1)^2})}{(1-\nu^{(2)^2})}\dfrac{E^{(2)}}{E^{(1)}} & \text{(plane strain)} \end{cases}. \quad (33)$$

Similarly, for out-of-plane deformation, the ratios of the separated quantities to their total can be expressed by

$$\frac{J_1^{0(m)}}{J_1^0} = \frac{G^{(m)}}{G} = \left(\frac{1}{\mu^{(m)}}\right) \Big/ \left(\frac{1}{\mu^{(1)}} + \frac{1}{\mu^{(2)}}\right) \quad (m=1,2). \quad (34)$$

Then, the ratio of the separated J-integrals or separated energy release rates is given by

$$\frac{J_1^{0(1)}}{J_1^{0(2)}} = \frac{G^{(1)}}{G^{(2)}} = \frac{\mu^{(2)}}{\mu^{(1)}}. \quad (35)$$

It is interesting to see in Eqs. (33) and (35) that the ratios of the separated J-integrals or separated energy release rates are proportional to the inverse of the mismatch ratio of the shear moduli or Young's moduli. Thus, the compliant material side supplies more fracture energy to the interfacial crack tip. These ratios can be used to quantify the fracture energy supply mechanism to the interfacial crack tip.

6.2 Vertical Component of the Separated J-integral $J_2^{0(m)}$. For in-plane deformation, using Eqs. (16) and (17) in Eqs. (26a) and (26b) with $k=2$, the vertical components of the separated J-integrals $J_2^{0(m)}$ $(m=1,2)$ are evaluated as

$$J_2^{0(m)} = -\lim_{r\to 0}\left\{\frac{e^{\pi\varepsilon(2m-3)}\tanh(\pi\varepsilon)\left[K_1^2\sin\left(2\varepsilon\ln\frac{r}{l}\right)+2K_1K_2\cos\left(2\varepsilon\ln\frac{r}{l}\right)-K_2^2\sin\left(2\varepsilon\ln\frac{r}{l}\right)\right](1+\kappa^{(m)})}{16\mu^{(m)}\pi\varepsilon\cosh(\pi\varepsilon)}\right\} \quad (m=1,2). \quad (36)$$

Furthermore, from Eq. (36), the vertical component of the J-integral for the interfacial crack is obtained as

$$J_2^0 = -\lim_{r\to 0}\left\{\frac{\tanh(\pi\varepsilon)\left[K_1^2\sin\left(2\varepsilon\ln\frac{r}{l}\right)+2K_1K_2\cos\left(2\varepsilon\ln\frac{r}{l}\right)-K_2^2\sin\left(2\varepsilon\ln\frac{r}{l}\right)\right]}{16\pi\varepsilon\cosh(\pi\varepsilon)}\left[\frac{(1+\kappa^{(1)})}{\mu^{(1)}}e^{-\pi\varepsilon}+\frac{(1+\kappa^{(2)})}{\mu^{(2)}}e^{+\pi\varepsilon}\right]\right\}. \quad (37)$$

Because of the trigonometric functions containing the logarithmic singularity, the vertical components of the separated J-integrals and the J-integral (see Eqs. (36) and (37)) oscillate when the near-field path shrinks to the crack tip $(r\to 0)$. Therefore, the vertical components of the separated J-integrals and the J-integral cannot be related to the stress intensity factors for the interfacial crack tip.

However, if we consider a crack in a homogeneous material, the logarithmic singularity terms vanish because $\varepsilon=0$. Consequently, the stress intensity factors K_1 and K_2 reduce to the mode I and mode II stress intensity factors K_{I} and K_{II}, respectively, and Eq. (36) reduces to

$$J_2^{0(m)} = -\frac{K_{\mathrm{I}}K_{\mathrm{II}}(1+\kappa)}{8\mu} \quad (m=1,2), \quad (38)$$

where $\mu=\mu^{(1)}=\mu^{(2)}$ and $\kappa=\kappa^{(1)}=\kappa^{(2)}$. Thus, the vertical component of the J-integral for the homogeneous case is given by

$$J_2^0 = -\frac{K_{\mathrm{I}}K_{\mathrm{II}}(1+\kappa)}{4\mu}. \quad (39)$$

This agrees with the relation given by Nishioka and Atluri [2].

For out-of-plane deformation, using Eqs. (18) and (19) in Eqs. (26a) and (26b) with $k=2$, the vertical components of the separated J-integrals $J_2^{0(m)}$ $(m=1,2)$ are evaluated as

$$J_2^{0(m)} = 0 \quad (m=1,2). \quad (40)$$

In addition, from Eq. (38), the vertical component of the J-integral becomes zero:

$$J_2^0 = 0. \quad (41)$$

7 The Component Separation Method of the J-Integral for Extracting the Mixed-Mode Stress Intensity Factors of an Interfacial Crack Tip

To extract mixed-mode stress intensity factors from the dynamic J-integral and the dynamic energy release rate, Nishioka

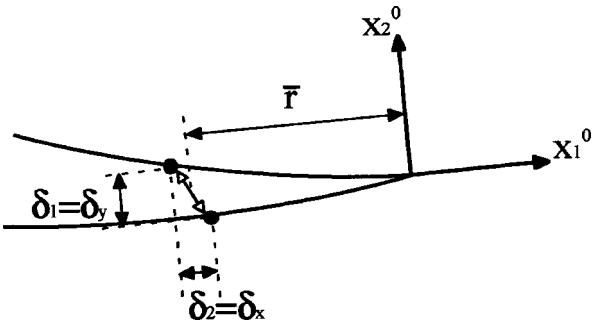


Fig. 6 Crack opening displacements

and coworkers developed the component separation method, [6]. In this paper, the component separation method, [6], is extended to the case of static interfacial fracture mechanics.

Using the ratio of the stress intensity factors ($\alpha = K_2/K_1$), the tangential component of the *J*-integral, i.e., Eq. (30), can be rewritten as

$$J_1^0 = \Lambda(1+\alpha^2)K_1^2, \tag{42}$$

and

$$\Lambda = \frac{1}{16\cosh^2(\varepsilon\pi)}\left(\frac{\kappa^{(1)}+1}{\mu^{(1)}}+\frac{\kappa^{(2)}+1}{\mu^{(2)}}\right). \tag{43}$$

Using Eqs. (17*a*) and (17*b*), the crack opening displacements in the $x_1^{(0)}$ and $x_2^{(0)}$-directions behind the crack tip, δ_x and δ_y, are obtained as

$$\delta_y + i\,\delta_x = \frac{(K_1+iK_2)}{2(1+2i\varepsilon)\cosh(\pi\varepsilon)} \times\left(\frac{\kappa^{(1)}+1}{\mu^{(1)}}+\frac{\kappa^{(2)}+1}{\mu^{(2)}}\right)\sqrt{\frac{r}{2\pi}}\left(\frac{r}{l}\right)^{i\varepsilon}. \tag{44}$$

Then, from Eq. (44), the ratio of the stress intensity factors can be related to the ratio of crack opening displacements as follows [15]:

$$\alpha = K_2/K_1 = \lim_{r\to 0}(1 - S\,\delta_y/\delta_x)/(\delta_y/\delta_x + S), \tag{45}$$

$$S = (\tan Q - 2\varepsilon)/(1+2\varepsilon\tan Q), \tag{46}$$

and

$$Q = \varepsilon\ln(r/l). \tag{47}$$

To calculate the ratio of the stress intensity factors it is necessary to take the limit $r\to 0$. However, it is difficult to obtain accurate numerical results in this limit because the quantity *S* has logarithmic and oscillatory singular terms. For this reason, and to derive explicit formulas for the component separation method, we eliminate $S\,(S=0)$ by taking $\tan Q = 2\varepsilon$ in Eq. (46). This can be achieved by choosing the following special characteristic length:

$$\bar{l} = \bar{r}/e^{\varepsilon^{-1}\tan^{-1}(2\varepsilon)}, \tag{48}$$

where $\bar{r}$ is the location at which the crack opening displacements are evaluated (see Fig. 6). The ratio of the stress intensity factors can then be accurately evaluated by

$$\alpha = \delta_x/\delta_y. \tag{49}$$

Using Eq. (49) in Eq. (42), the explicit formulas for the component separation method can be derived as

$$\bar{K}_k = \delta_k\sqrt{\frac{J_1^0}{\Lambda(\delta_1^2+\delta_2^2)}} = \delta_k\sqrt{\frac{J_1^{0(1)}+J_1^{0(2)}}{\Lambda(\delta_1^2+\delta_2^2)}} = \delta_k\sqrt{\frac{G}{\Lambda(\delta_1^2+\delta_2^2)}}, \quad (k=1,2), \tag{50}$$

where $\delta_1 = \delta_y$ and $\delta_2 = \delta_x$ (see Fig. 6).

The transformation to the stress intensity factors with the characteristic length $l = 2a$, or to those with a desired characteristic length *l*, can simply be conducted using the following equation, [16],

$$\begin{Bmatrix} K_1 \\ K_2 \end{Bmatrix} = \begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix}\begin{Bmatrix} \bar{K}_1 \\ \bar{K}_2 \end{Bmatrix}, \tag{51a}$$

$$\omega = \varepsilon\ln(l/\bar{l}). \tag{51b}$$

The features of the component separation method can be summarized as follows:

i. It can be expressed by explicit formulas.
ii. It does not require any auxiliary solution field.
iii. It is applicable using the path-independent separated *J*-integrals, the *J*-integral, or the energy release rate.
iv. The signs of the stress intensity factors are automatically determined by the signs of the corresponding crack opening displacements.
v. Since its formulas do not include the oscillatory and logarithmic singular terms, the numerical results for the stress intensity factors are stable and accurate.

In previous studies on extracting mixed-mode stress intensity factors for interfacial cracks, Yau and Wang's M_1 integral method, [10], has been commonly used. However, it is sometimes difficult to set up the auxiliary solution field that is necessary in the application of their method. For some complicated conditions, such as crack kinking and branching, it is hard to obtain the auxiliary solution. The component separation method developed here therefore has great advantages over the M_1 integral method, because no auxiliary solution field is needed.

8 Numerical Analyses of Interfacial Crack Problems

In this paper, all the stress intensity factors for interfacial cracks are expressed using the characteristic length $l = 2a$. Stress intensity factors with a different characteristic length can be obtained using the transformation given in Eq. (50).

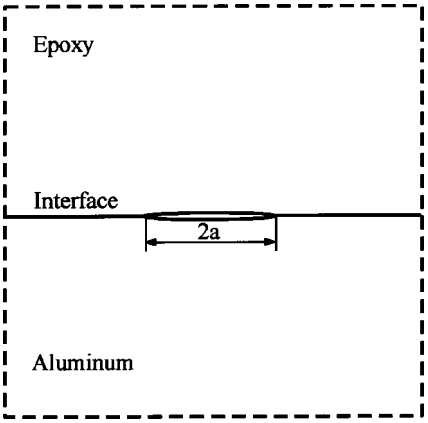


Fig. 7 An interfacial crack in an infinite bimaterial plate

Table 1 Material properties of the bimaterial plate

Material	Young's Modulus (GPa)	Poisson's Ratio
Epoxy	2.3	0.38
Aluminum Alloy	72.0	0.32

8.1 An Interfacial Crack in an Infinite Bimaterial Plate. To verify the validity of the separated J integrals and the component separation method, we considered an interfacial crack in an infinite bimaterial plate, as depicted in Fig. 7. The bimaterial consists of epoxy (material 1) and aluminum alloy (material 2). The properties of these materials are listed in Table 1. The crack length was set as $2a=60$ mm.

The theoretical near-tip displacement field under the plane stress condition was produced using Eqs. (17*a*) and (17*b*) with $K_1=K_2=0.18\,\mathrm{MPa}\cdot\mathrm{m}^{1/2}$. The theoretical displacement components were used as the corresponding nodal displacements in the finite element model (see Fig. 8), which consists of the eight-noded isoparametric elements. The stresses and strains were calculated using the finite element method. The separated *J*-integrals were then calculated along the paths depicted in Fig. 8. Excellent path-independence of the separated *J*-integrals was observed. The *J*-integral value was obtained by summing the separated *J*-integrals.

The stress intensity factors were converted from the *J*-integral value using the component separation method (see Eqs. (50) and (51)), and are listed in Table 2. The ratio of the crack opening displacements was evaluated at the corner nodes nearest to the crack tip. The numerical solutions agree excellently with the theoretical values.

8.2 A Central Crack in a Finite Bimaterial Plate. Now, we considered a central crack in a finite bimaterial plate subject to a uniform applied stress $\bar{\sigma}_{22}$, as depicted in Fig. 9. The finite element mesh with the eight-noded isoparametric elements is shown in Fig. 10(*a*). The separated *J*-integrals were evaluated using five paths around the crack tip, as shown in Fig. 10(*b*). The separated *J*-integrals showed excellent path independence. The *J*-integral values were evaluated by summing the separated *J*-integrals. Then, the stress intensity factors were obtained using the component separation method (see Eqs. (50) and (51)). The stress intensity factors were normalized by

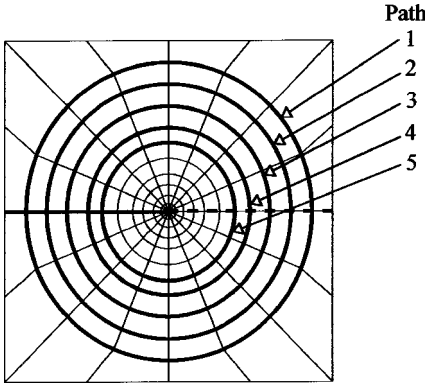


Fig. 8 Finite element mesh pattern around the crack tip

Table 2 Stress intensity factors for an interfacial crack in an infinite plate

	Theoretical Value	Component Separation Method
K_1 [MPa·m$^{1/2}$]	0.1800	0.1795 (−0.27%)
K_2 [MPa·m$^{1/2}$]	0.1800	0.1795 (−0.27%)

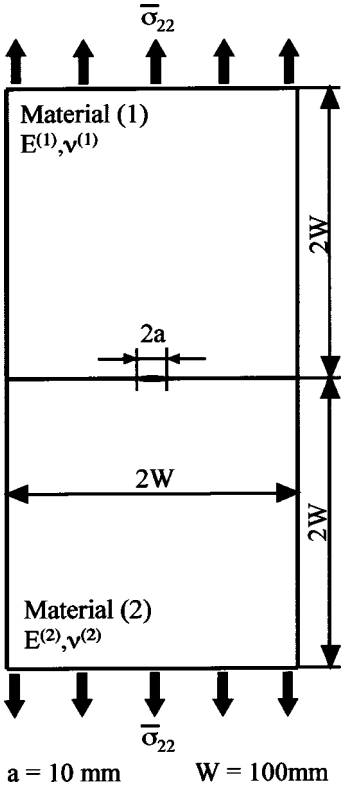
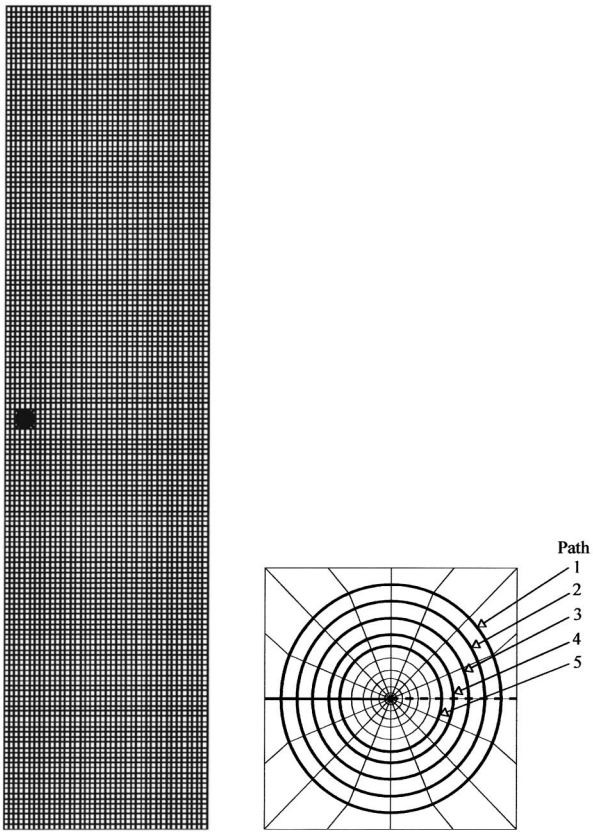


Fig. 9 A central interface crack in a bimaterial plate



(a) Finite element mesh pattern (a/W =0.1) (b) Integral paths

Fig. 10 Finite element model for a finite bimaterial plate

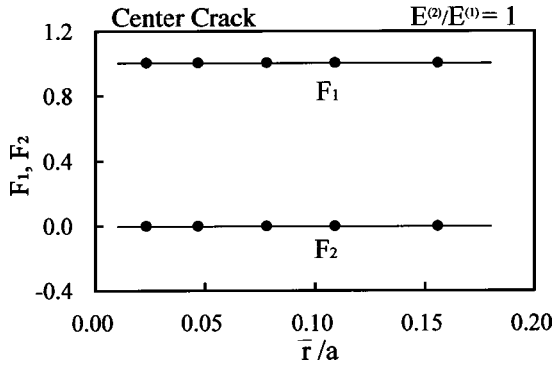


Fig. 11 Variation of normalized stress intensity factor with evaluating location (homogeneous plate)

$$F_k = K_k/(\bar{\sigma}_{22}\sqrt{\pi a}), \quad (k=1,2). \tag{52}$$

In order to check the effect of the location at which the crack opening displacements (see $\bar{r}$ in Fig. 6) are evaluated, we first solved the problem depicted in Fig. 9 for $E^{(2)}/E^{(1)}=1$ and 10. The normalized stress intensity factors obtained by the component separation method are plotted against the distance $\bar{r}$ in Figs. 11 and 12 for $E^{(2)}/E^{(1)}=1$ and 10, respectively. In these cases, the nearest five-corner nodes were used for the evaluation of the crack opening displacements. The solid lines are the average values of the five results. The results show excellent independence from the distance $\bar{r}$. Thus, extrapolation to the crack tip is not necessary when the component separation method is used.

Next, we systematically changed the crack length and mismatch ratio. The normalized stress intensity factors obtained by the component separation method are summarized in Table 3 and compared with the numerical solutions of Yuuki and Cho [17] using the boundary element method. The difference between the present solutions and their solutions is defined as

$$\frac{F_k - F_{k\ \mathrm{BEM}}}{\sqrt{F_{1\ \mathrm{BEM}}^2 + F_{2\ \mathrm{BEM}}^2}} \quad (k=1,2), \tag{53}$$

where $F_{k\ \mathrm{BEM}}\,(k=1,2)$ represents the normalized stress intensity factors obtained by Yuuki and Cho [17]. It is seen that the present solutions are very close to their solutions. In fact, the difference is at most only 0.46%.

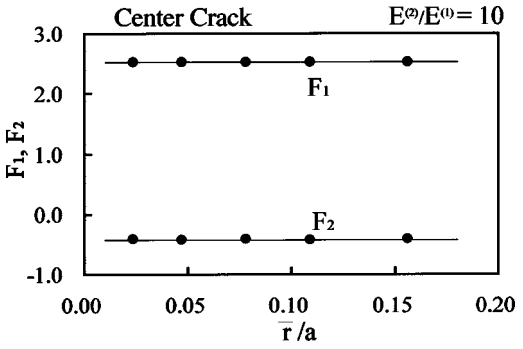


Fig. 12 Variation of normalized stress intensity factor with evaluating location (bimaterial plate)

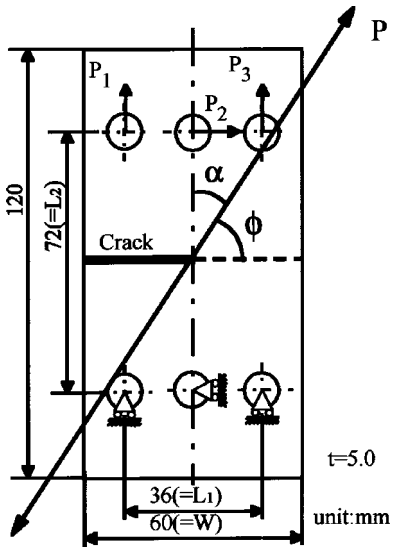


Fig. 13 Compact normal and shear specimen

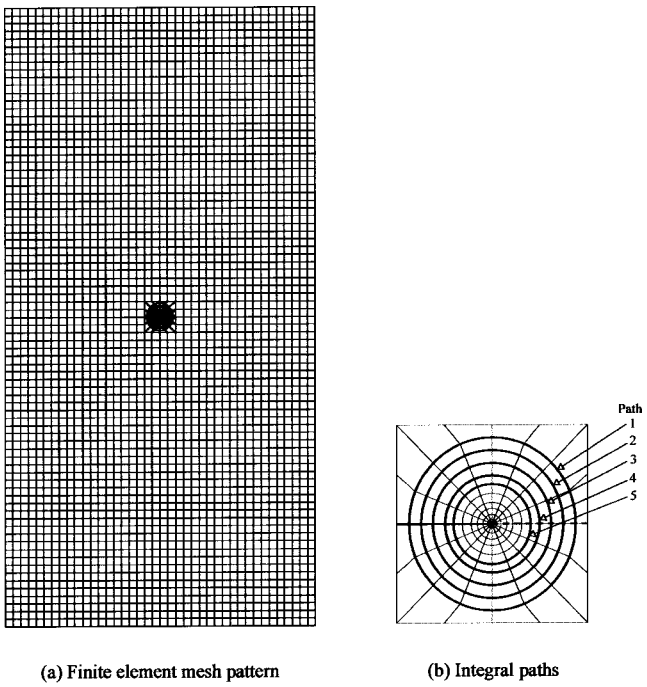
8.3 Compact Normal and Shear Specimen. Here we consider the compact normal and shear (CNS) specimen devised by Richard and Benitz [18], as shown in Fig. 13. The loading direction can be systematically changed from a loading angle of $\phi=0°$ to 180°. The finite element mesh pattern is shown in Fig. 14(*a*). The paths for the separated *J*-integral are shown in Fig. 14(*b*). The loading and constraint conditions are indicated in Fig. 13. The nodal forces P_1, P_2, and P_3 were determined by the following equilibrium conditions for forces and moments [19]:

$$P_2 = P\cos\phi \tag{54}$$

$$P_1 + P_3 = P\sin\phi \tag{55}$$

Table 3 Normalized stress intensity factors for a central interface crack in a finite bimaterial plate under uniform tension ($l=2a$)

	Upper: F_1 (difference %), Lower: F_2 (difference %),				
a/W $E^{(1)}/E^{(2)}$	0.1	0.2	0.3	0.4	0.5
1.0	1.0035(−0.25)	1.0226(−0.19)	1.0561(−0.15)	1.1079(−0.14)	1.1852(−0.13)
	0.0000(0.00)	0.0000(0.00)	0.0000(0.00)	0.0000(0.00)	0.0000(0.00)
2.0	0.9978(0.14)	1.0169(−0.16)	1.0501(−0.23)	1.1013(−0.21)	1.1774(−0.25)
	−0.0747(−0.29)	−0.0742(−0.28)	−0.0748(−0.31)	−0.0771(−0.39)	−0.0819(−0.34)
3.0	0.9903(0.12)	1.0094(−0.20)	1.0423(−0.22)	1.0926(−0.26)	1.1673(−0.25)
	−0.1112(−0.34)	−0.1101(−0.38)	−0.1108(−0.40)	−0.1139(−0.39)	−0.1206(−0.40)
4.0	0.9840(0.14)	1.0031(−0.19)	1.0357(−0.25)	1.0854(−0.26)	1.1588(−0.27)
	−0.1326(−0.40)	−0.1310(−0.43)	−0.1315(−0.46)	−0.1349(−0.39)	−0.1425(−0.46)
10.0	0.9644(0.20)	0.9836(−0.18)	1.0152(−0.24)	1.0627(−0.20)	1.1322(−0.19)
	−0.1774(−0.42)	−0.1741(−0.36)	−0.1736(−0.26)	−0.1768(−0.30)	−0.1851(−0.22)
100.0	0.9426(0.28)	0.9618(−0.05)	0.9922(−0.17)	1.0372(−0.10)	1.1023(−0.15)
	−0.2079(−0.31)	−0.2027(−0.16)	−0.2008(0.00)	−0.2028(0.07)	−0.2104(0.04)



(a) Finite element mesh pattern (b) Integral paths

Fig. 14 Finite element model for a CNS specimen

$$P_1L_1+P_2L_2=P_3L_1. \tag{56}$$

The stress intensity factors were normalized as follows:

$$F_k=K_k/(\bar{\sigma}\sqrt{\pi a}) \quad (k=1,2), \tag{57}$$

where $\bar{\sigma}=P/(W\cdot t)$, and W and t are the width and thickness of the specimen, respectively.

First, a homogeneous aluminum alloy specimen is considered. In Fig. 15 the normalized stress intensity factors obtained using the component separation method are plotted against the evaluating location $\bar{r}$. Under the pure tension loading with $\phi=90$ deg, the F_2 value is exactly zero, whereas under the pure shear loading with $\phi=0$ deg, the F_1 value is exactly zero. For both values of the loading angle, $\phi=0$ deg and 90 deg, the results show excellent independence from $\bar{r}$. Thus, the stress intensity factors can accurately be evaluated using the component separation method.

The normalized stress intensity factors obtained using the component separation method of the *J*-integral are plotted in Fig. 16 against the loading angle. In this case, the normalized stress intensity factors F_1 and F_2 correspond to the mode I and mode II

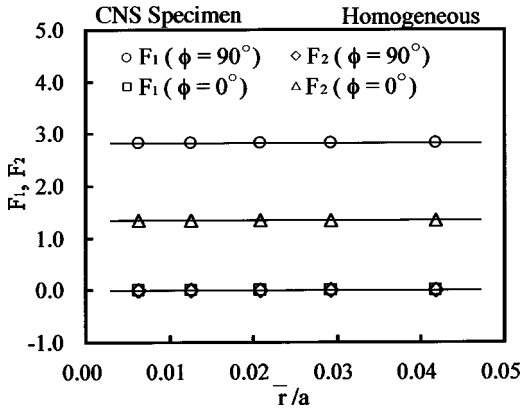


Fig. 15 Variation of normalized stress intensity factor with evaluating location (homogeneous CNS specimen)

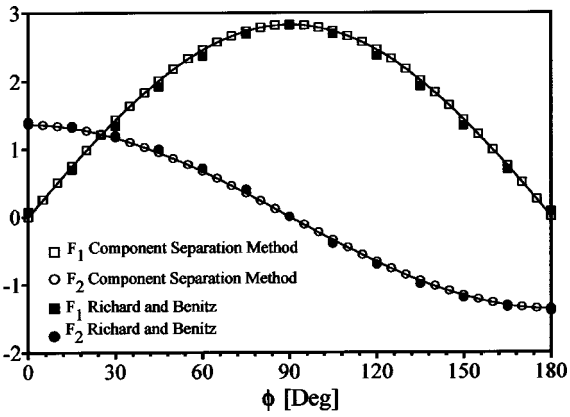


Fig. 16 Stress intensity factors versus loading angle (homogeneous material)

correction factors F_{I} and F_{II}, respectively. In addition to the pure modes under $\phi=0$ deg and 90 deg, the F_1 value under another pure shearing load, $\phi=180$ deg, is also exactly zero. Furthermore, as can also be predicted theoretically, the F_1 results are exactly symmetric with respect to $\phi=90$ deg, whereas the F_2 results are exactly antisymmetric. The present results agree excellently with the results obtained by Richard and Benitz [18].

Next we consider a bimaterial specimen of epoxy and aluminum alloy. The properties used are the same as those listed in Table 1. Figure 17 shows the variation in the normalized stress intensity factor with the evaluating location. Once again, the results show excellent independence from the evaluating location.

Figure 18 shows the variations of the normalized stress intensity factors for various loading angles. In contrast to the homogeneous case, the mismatch of the Young's moduli causes the F_1 values to be nonzero under the pure shear loadings with $\phi=0$ deg and 180 deg, and the F_2 value is also nonzero under the pure tension loading with $\phi=90$ deg.

8.4 A Curved Interfacial Crack in a Composite Material Now we demonstrate the applicability of the separated *J*-integrals and the component separation method to the problem of a curved interfacial crack. For a curved interfacial crack, the M_1 integral method is difficult to apply because of the difficulty of constructing an appropriate auxiliary solution field.

We consider a fiber-reinforced composite material that consists of SiC fibers and aluminum base material, [20], as shown in Fig. 19. Using periodicity conditions, only a unit cell (see Fig. 20) was analyzed under the plane strain condition. The Young's modulus and Poisson's ratio of SiC (material 2) are assumed to be

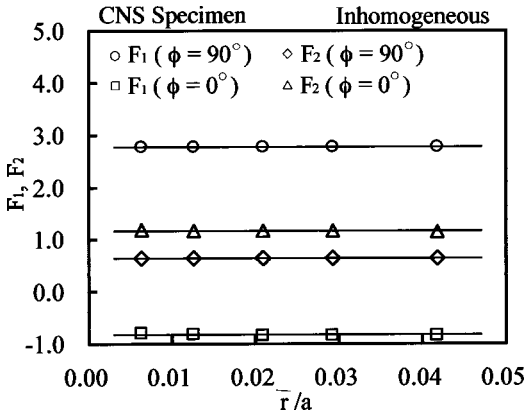


Fig. 17 Variation of normalized stress intensity factor with evaluating location (bimaterial CNS specimen)

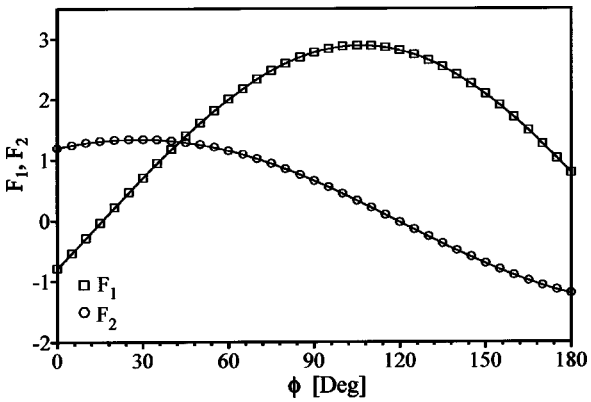


Fig. 18 Stress intensity factors versus loading angle (nonhomogeneous material)

$E^{(2)}=450$ Gpa and $\nu^{(2)}=0.170$, respectively, whereas those of aluminum (material 1) are $E^{(1)}=69$ Gpa and $\nu^{(1)}=0.333$. In this problem, two symmetrical curved interfacial cracks are considered.

Figure 21 shows the finite element mesh pattern used for the analysis. Due to the symmetry conditions, only one-quarter of the cell was analyzed. The unit cell was subjected to uniform displacements of $\bar{u}=0.1\,\mu$m at the upper and lower boundaries. The separated *J*-integrals were evaluated for five paths around the crack tip, as shown in Fig. 22.

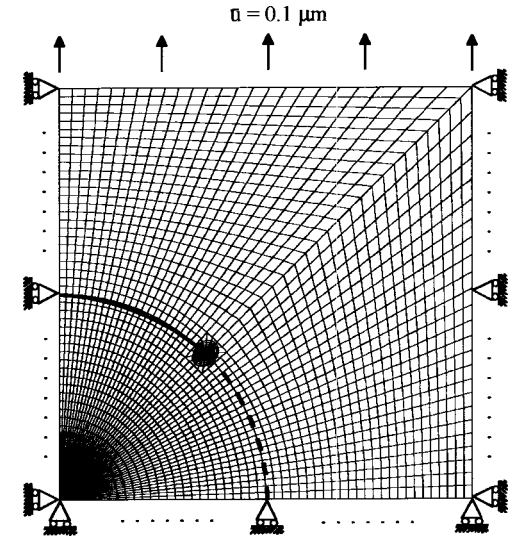


Fig. 21 Finite element mesh pattern for the fiber-reinforced composite material

The separated *J*-integrals and the *J*-integral are plotted in Fig. 23 against the path number. All of them are excellently path-independent. The separated *J*-integral of the matrix is much higher than that of the fiber. In other words, the separated energy release rate of the matrix is much higher than that of the fiber. This im-

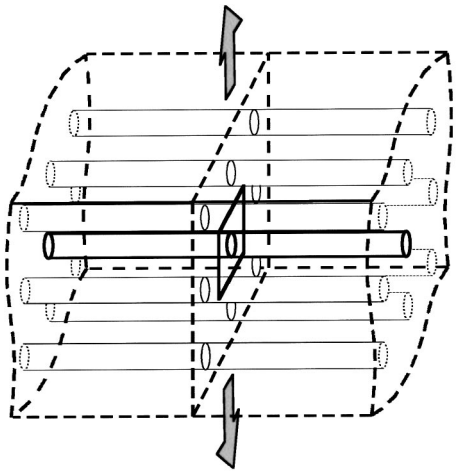


Fig. 19 Fiber-reinforced composite material

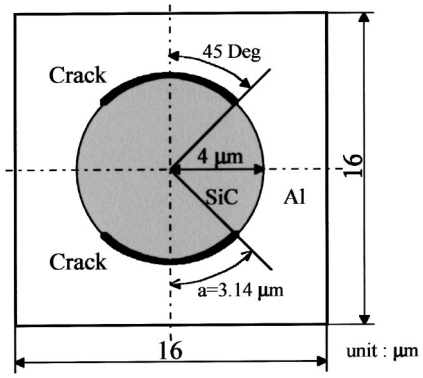


Fig. 20 Curved interfacial cracks around the SiC fiber

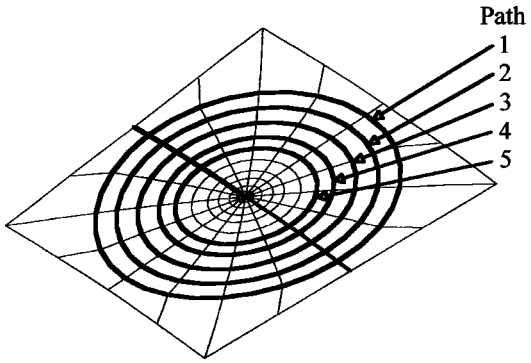


Fig. 22 Mesh pattern around the crack tip

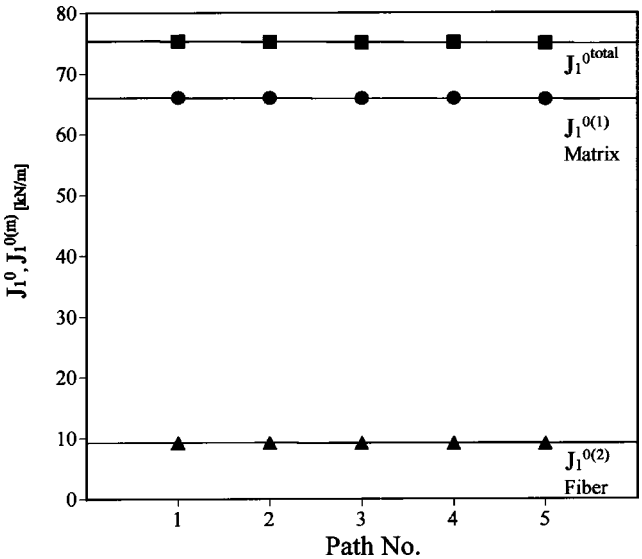


Fig. 23 Path independence of the separated *J*-integrals

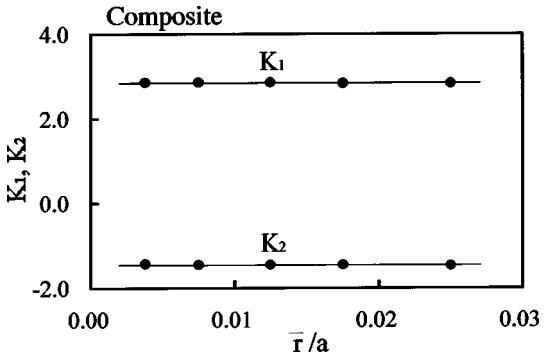


Fig. 24 Variation of stress intensity factor with evaluating location (fiber-reinforced composite)

plies that the compliant base metal provides a much larger fracture energy to the interfacial crack tip than does the stiffer fiber.

The stress intensity factors were obtained using the separated *J*-integrals and the *J*-integral in the formulas of the component separation method (see Eqs. (50) and (51)). The results for the stress intensity factors are plotted against the evaluating location. (See Fig. 24.) Excellent location independence is found, even for the curved crack. Thus, the results should be very accurate. For this case, the average stress intensity factors were $K_1 = 90.520\ \mathrm{MPa}\cdot\mathrm{m}^{1/2}$ and $K_2 = -45.277\ \mathrm{MPa}\cdot\mathrm{m}^{1/2}$.

9 Conclusions

In this paper, we first presented the concepts of path-independent separated *J*-integrals and separated energy release rates. We then derived the relations between the separated *J*-integrals and the stress intensity factors using the asymptotic solutions for an interfacial crack tip. It was shown that the path-independent separated *J*-integrals have the physical significance of energy flows into an interfacial crack tip from the adjacent individual material sides or, equivalently, the separated energy release rates.

To accurately evaluate the stress intensity factors, the component separation method of the *J*-integral was extended for interfacial crack problems. The component separation method has great advantages over the M_1 integral method that is often used in interfacial fracture mechanics problems.

Finally, pertinent numerical analyses were carried out to demonstrate the usefulness of the separated *J*-integrals and the component separation method. In particular, it was shown from the separated *J*-integrals for curved interfacial cracks in a SiC fiber-reinforced aluminum base composite that the compliant base metal provides a much larger fracture energy to the interfacial crack tip. In addition, the component separation method demonstrated accurate extraction of the stress intensity factors from the path independent separated *J*-integrals, without constructing any auxiliary solution field.

Acknowledgments

This study was supported by the natural science grant from the Mitsubishi Foundation. The authors gratefully acknowledge Mr. T. Kagimoto for his assistance.

Appendix

On the basis of the stress intensity factors defined by Eq. (13), the explicit near-tip stress and displacement expressions for an interfacial crack can be summarized as follows.

The asymptotic in-plane stress components for the material *m* side can be expressed by

$$\sigma_{11}^{(m)}=\frac{K_1}{2\sqrt{2\pi r}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\left\{3\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)+2\varepsilon\sin\theta\cos\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)-\sin\theta\sin\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}\right.$$
$$\left.-e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\cos\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right]-\frac{K_2}{2\sqrt{2\pi r}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\left\{3\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right.\right.$$
$$\left.\left.+2\varepsilon\sin\theta\sin\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)+\sin\theta\cos\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}+e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\sin\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right], \tag{A1}$$

$$\sigma_{22}^{(m)}=\frac{K_1}{2\sqrt{2\pi r}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\left\{\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)-2\varepsilon\sin\theta\cos\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)+\sin\theta\sin\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}\right.$$
$$\left.+e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\cos\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right]-\frac{K_2}{2\sqrt{2\pi r}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\left\{\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right.\right.$$
$$\left.\left.-2\varepsilon\sin\theta\sin\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)-\sin\theta\cos\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}-e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\sin\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right], \tag{A2}$$

$$\sigma_{12}^{(m)}=\frac{K_1}{2\sqrt{2\pi r}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\left\{\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)+2\varepsilon\sin\theta\sin\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)+\sin\theta\cos\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}\right.$$
$$\left.-e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\sin\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right]-\frac{K_2}{2\sqrt{2\pi r}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\left\{-\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right.\right.$$
$$\left.\left.-2\varepsilon\sin\theta\cos\left(\frac{3\theta}{2}+\epsilon\ln\frac{r}{l}\right)+\sin\theta\sin\left(\frac{3\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}-e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\cos\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right]. \tag{A3}$$

The in-plane displacement components for the material *m* side are expressed by

$$u_1^{(m)}=\frac{K_1\sqrt{2\pi r}}{4\pi\mu^{(m)}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\frac{\kappa^{(m)}}{1+4\varepsilon^2}\left\{\cos\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)-2\varepsilon\sin\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right\}\right.$$

$$-e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\frac{1}{1+4\varepsilon^2}\left\{\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)+2\varepsilon\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}+e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\sin\theta\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)$$

$$+\frac{K_2\sqrt{2\pi r}}{4\pi\mu^{(m)}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\frac{\kappa^{(m)}}{1+4\varepsilon^2}\left\{\sin\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)+2\varepsilon\cos\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right\}\right]$$

$$\left.-e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\frac{1}{1+4\varepsilon^2}\left\{-\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)+2\varepsilon\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}+e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\sin\theta\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right],\qquad(A4)$$

$$u_2^{(m)}=\frac{K_1\sqrt{2\pi r}}{4\pi\mu^{(m)}\cosh(\pi\varepsilon)}\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\frac{\kappa^{(m)}}{1+4\varepsilon^2}\left\{\sin\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)+2\varepsilon\cos\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right\}-e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\right.$$

$$\times\frac{1}{1+4\varepsilon^2}\left\{-\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)+2\varepsilon\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}-e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\sin\theta\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\Big]+\frac{K_2\sqrt{2\pi r}}{4\pi\mu^{(m)}\cosh(\pi\varepsilon)}$$

$$\times\left[e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\frac{-\kappa^{(m)}}{1+4\varepsilon^2}\left\{\cos\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)-2\varepsilon\sin\left(\frac{\theta}{2}-\varepsilon\ln\frac{r}{l}\right)\right\}+e^{-\varepsilon\theta+(3-2m)\varepsilon\pi}\frac{1}{1+4\varepsilon^2}\left\{\cos\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right.\right.$$

$$\left.\left.+2\varepsilon\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right\}+e^{+\varepsilon\theta-(3-2m)\varepsilon\pi}\sin\theta\sin\left(\frac{\theta}{2}+\varepsilon\ln\frac{r}{l}\right)\right].\qquad(A5)$$

References

[1] Nishioka, T., and Yasin, A., 1999, "The Dynamic J Integral, Separated Dynamic J Integrals and Moving Finite Element Simulation, for Subsonic, Transonic and Supersonic Interfacial Crack Propagation," JSME Int. J., Ser. A, **42**, pp. 25–39.

[2] Nishioka, T., and Atluri, S. N., 1983, "Path-Independent Integrals, Energy Release Rates, and General Solutions of Near-Tip Fields in Mixed-Mode Dynamic Fracture Mechanics," Eng. Fract. Mech., **18**, pp. 1–22.

[3] Nishioka, T., and Yasin, A., 1999, "Finite Element Simulations of Impact Interfacial Fracture Problems," *Impact Response of Materials and Structures*, V. P. W. Shim, S. Tanimura, and C. T. Lim, Eds., Oxford University Press, New York, pp. 426–431.

[4] Rice, J. R., 1968, "A Path Independent Integral and Approximate Analysis of Strain Concentration by Notches and Cracks," ASME J. Appl. Mech., **35**, pp. 379–386.

[5] Sun, C. T., and Jih, C. J., 1987, "On Strain Energy Release Rates for Interfacial Cracks in Bi-Material Media," Eng. Fract. Mech., **28**, pp. 13–20.

[6] Nishioka, T., Murakami, R., and Takemoto, Y., 1990, "The Use of the Dynamic J Integral (J′) in Finite Element Simulation of Mode I and Mixed-Mode Dynamic Crack Propagation," Int. J. Pressure Vessels Piping, **44**, pp. 329–352.

[7] Nishioka, T., 1994, "The State of the Art in Computational Dynamic Fracture Mechanics," JSME Int. J., Ser. A, **37**, pp. 313–333.

[8] Nishioka, T., 1998, "On the Dynamic J Integral in Dynamic Fracture Mechanics," *Fracture: A Topical Encyclopedia of Current Knowledge*, G. P. Cherepanov, ed., Krieger, Melbourne, FL, pp. 575–617.

[9] Nishioka, T., Hu, Q. H., and Fujimoto, T., 2001, "Numerical Simulations of Dynamic Interfacial Fracture Phenomena," *Proceedings of the 10th International Conference on Fracture*, Honolulu, HI, Dec. 3–7, Elsevier, New York.

[10] Yau, J. F., and Wang, S. S., 1984, "An Analysis of Interface Cracks Between Dissimilar Isotropic Materials Using Conservation Integrals in Elasticity," Eng. Fract. Mech., **20**, pp. 423–432.

[11] Yuuki, R., and Cho, S. B., 1989, "Efficient Boundary Element Analysis of Stress Intensity Factors for Interface Cracks in Dissimilar Materials," Eng. Fract. Mech., **34**, pp. 179–188.

[12] Shih, C. F., and Asaro, R. J., 1988, "Elastic-Plastic Analysis of Cracks in Bimaterial Interface: Part I—Small Scale Yielding," ASME J. Appl. Mech., **55**, pp. 299–316.

[13] Malyshev, B., and Salganik, R., 1965, "The Strength of Adhesive Joints Using the Theory of Cracks," Int. J. Fract. Mech., **1**, pp. 114–119.

[14] Willis, J. R., 1971, "Fracture Mechanics of Interfacial Cracks," J. Mech. Phys. Solids, **19**, pp. 353–368.

[15] Yuuki, et al., 1993, *Mechanics of Interface*, Baifukan, pp. 103–105 (in Japanese).

[16] Ikeda, T., Miyazaki, M., Soda, T., and Munakata, T., 1992, "Mixed Mode Fracture Criteria of Interface Crack Between Dissimilar Materials," Trans. Jpn. Soc. Mech. Eng., **58A**, pp. 2080–2087.

[17] Yuuki, R., and Cho, S. B., 1989, "Efficient Boundary Element Analysis of Stress Intensity Factors for Interface Cracks in Dissimilar Materials," Eng. Fract. Mech., **34**, pp. 179–188.

[18] Richard, H. A., and Benitz, K., 1983, "A Loading Device for the Creation of Mixed Mode in Fracture Mechanics," Int. J. Fract., **22**, pp. R55–R59.

[19] Kishimoto, K., Fukano, H., Yoshida, T., and Aoki, S., 1990, "Elastic-Plastic Fracture Behavior of an Aluminum Alloy Under Mixed Mode Conditions," Trans. Jpn. Soc. Mech. Eng., **56A**, pp. 957–965.

[20] Fujimoto, T., and Tomita, Y., 1998, "Modeling and Simulation of Response of Particulate-Reinforced Composite Materials Under a Plane Strain Condition," Trans. Jpn. Soc. Mech. Eng., **64A**, pp. 2694–2700.

The Mode III Crack Problem in Microstructured Solids Governed by Dipolar Gradient Elasticity: Static and Dynamic Analysis

H. G. Georgiadis
Mechanics Division,
National Technical University of Athens,
1 Konitsis Street,
Zographou GR-15773, Greece
e-mail: georgiad@central.ntua.gr
Mem. ASME

This study aims at determining the elastic stress and displacement fields around a crack in a microstructured body under a remotely applied loading of the antiplane shear (mode III) type. The material microstructure is modeled through the Mindlin-Green-Rivlin dipolar gradient theory (or strain-gradient theory of grade two). A simple but yet rigorous version of this generalized continuum theory is taken here by considering an isotropic linear expression of the elastic strain-energy density in antiplane shearing that involves only two material constants (the shear modulus and the so-called gradient coefficient). In particular, the strain-energy density function, besides its dependence upon the standard strain terms, depends also on strain gradients. This expression derives from form II of Mindlin's theory, a form that is appropriate for a gradient formulation with no couple-stress effects (in this case the strain-energy density function does not contain any rotation gradients). Here, both the formulation of the problem and the solution method are exact and lead to results for the near-tip field showing significant departure from the predictions of the classical fracture mechanics. In view of these results, it seems that the conventional fracture mechanics is inadequate to analyze crack problems in microstructured materials. Indeed, the present results suggest that the stress distribution ahead of the tip exhibits a local maximum that is bounded. Therefore, this maximum value may serve as a measure of the critical stress level at which further advancement of the crack may occur. Also, in the vicinity of the crack tip, the crack-face displacement closes more smoothly as compared to the classical results. The latter can be explained physically since materials with microstructure behave in a more rigid way (having increased stiffness) as compared to materials without microstructure (i.e., materials governed by classical continuum mechanics). The new formulation of the crack problem required also new extended definitions for the J-integral and the energy release rate. It is shown that these quantities can be determined through the use of distribution (generalized function) theory. The boundary value problem was attacked by both the asymptotic Williams technique and the exact Wiener-Hopf technique. Both static and time-harmonic dynamic analyses are provided.
[DOI: 10.1115/1.1574061]

1 Introduction

The present work is concerned with the exact determination of mode III crack-tip fields within the framework of the dipolar gradient elasticity (or strain-gradient elasticity of grade two). This theory was introduced by Mindlin [1], Green and Rivlin [2], and Green [3] in an effort to model the mechanical response of materials with *microstructure*. The theory begins with the very general concept of a continuum containing elements or particles (called macromedia), which are in themselves *deformable* media. This behavior can easily be realized if such a macro-particle is viewed as a collection of smaller subparticles (called micromedia). In this way, each particle of the continuum is endowed with an *internal* displacement field, which is expanded as a power series in internal coordinate variables. Within the above context, the lowest-order theory (dipolar or grade-two theory) is the one obtained by retaining only the first (linear) term. Also, since these theories introduce dependence on strain and/or rotation gradients, the new material constants imply the presence of characteristic lengths in the material behavior, which allow the incorporation of size effects into stress analysis in a manner that the classical theory cannot afford.

The Mindlin-Green-Rivlin theory and related ideas, after a first development and some successful applications mainly on stress concentration problems during the sixties (see, e.g., Mindlin and Eshel [4], Weitsman [5], Day and Weitsman [6], Cook and Weitsman [7], Herrmann and Achenbach [8], and Achenbach et al. [9]), have also recently been employed to analyze complex problems in materials with microstructure (see, e.g., Vardoulakis and Sulem [10], Fleck et al. [11], Lakes [12], Vardoulakis and Georgiadis [13], Wei and Huthinson [14], Begley and Huthinson [15], Exadaktylos and Vardoulakis [16], Huang et al. [17], Zhang et al. [18], Chen et al. [19], Georgiadis and Vardoulakis [20], Georgiadis et al. [21,22], Georgiadis and Velgaki [23], and Amanatidou and Aravas [24]). More specifically, recent work by the author and co-workers [13,20–23], on wave-propagation problems showed that the gradient approach predicts types of elastic waves that are not predicted by the classical theory (SH and torsional *surface* waves in homogeneous materials) and also predicts *dispersion* of high-frequency Rayleigh waves (the classical elasticity fails to predict dispersion of these waves at *any* frequency). Notice that all these phenomena are observed in experiments and are also predicted by atomic-lattice analyses (see, e.g., Gazis et al. [25]).

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Apr. 28, 2002; final revision, Dec. 19, 2002. Associate Editor: B. M. Moran. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

Thus, based on existing gradient-type results, one may conclude that the Mindlin-Green-Rivlin theory extends the range of applicability of continuum theories in an effort towards bridging the gap between classical (monopolar or nongeneralized) theories of continua and theories of atomic lattices.

In the present work the concept adopted, following the aforementioned ideas, is to view the continuum as a periodic structure like that, e.g., of crystal lattices, crystallites of a polycrystal or grains of a granular material. The material is composed wholly of unit cells (micromedia) having the form of cubes with edges of size $2h$. This size is therefore an intrinsic material length. We further assume (and this is a rather standard assumption in studies applying the Mindlin-Green-Rivlin theory to practical problems) that the continuum is *homogeneous* in the sense that the relative deformation (i.e., the difference between the macrodisplacement gradient and the microdeformation—cf. Mindlin [1]) is zero and the microdensity does not differ from the macrodensity. Then, we formulate the mode III crack problem by considering an isotropic and linear expression of the strain-energy density W. This expression in antiplane shear and with respect to a Cartesian coordinate system $Ox_1x_2x_3$ reads $W=\mu\varepsilon_{p3}\varepsilon_{p3}+\mu c(\partial_s\varepsilon_{p3})(\partial_s\varepsilon_{p3})$, where the summation convention is understood over the Latin indices, which take the values 1 and 2 only, $(\varepsilon_{13},\varepsilon_{23})$ are the only identically nonvanishing components of the linear strain tensor, μ is the shear modulus, c is the gradient coefficient (a positive constant accounting for microstructural effects), and $\partial_s(\,)\equiv\partial(\,)/\partial x_s$. The problem is two-dimensional and is stated in the plane (x_1,x_2). The above strain-energy density function is the simplest possible form of case II in Mindlin's [1] theory and is appropriate for a gradient formulation with *no* couple-stress effects, because W is completely *independent* upon rotation gradients. Indeed, by referring to a strain-energy density function that depends upon strains and strain gradients in a three-dimensional body (the Latin indices now span the range (1,2,3)), i.e., a function of the form $W=(1/2)c_{pqsj}\varepsilon_{pq}\varepsilon_{sj}+(1/2)d_{pqsjlm}\kappa_{pqs}\kappa_{jlm}$ with (c_{pqsj},d_{pqsjlm}) being tensors of material constants and $\kappa_{pqs}=\partial_p\varepsilon_{qs}\equiv\partial_p\varepsilon_{sq}$, and by defining the Cauchy (in Mindlin's notation) stress tensor as $\tau_{pq}=\partial W/\partial\varepsilon_{pq}$ and the dipolar stress tensor (a third-rank tensor) as $m_{pqs}=\partial W/\partial(\partial_p\varepsilon_{qs})$, one may observe that the relations $m_{pqs}=m_{p(qs)}$ and $m_{p[qs]}=0$ hold, where $(\,)$ and $[\,]$ as subscripts denote the symmetric and antisymmetric parts of a tensor, respectively. Accordingly, couple stresses do not appear within the present formulation by assuming dipolar (internal) forces with vanishing antisymmetric part (more details on this are given in Section 2 below). A couple-stress, quasi-static solution of the mode-III crack problem was given earlier by Zhang et al. [18]. Note in passing that in the literature one may find mainly two types of approaches: In the first type (couple-stress case) the strain-energy density depends on rotation gradients and has no dependence upon strain gradients of the kind mentioned above (see, e.g., [11,17–19,23]), whereas in the second type the strain-energy density depends on strain gradients and has no dependence upon rotation gradients (see, e.g., [13,16,20–22]). Exceptions from this trend exist of course (see, e.g., [5–7]) and these works employ a more complicated formulation based on form III of Mindlin's theory, [1].

Here, in addition to the quasi-static case, we also treat the time-harmonic dynamical case, which is pertinent to the problem of stress-wave diffraction by a pre-existing crack in the body. In the latter case, besides the standard inertia term in the equation of motion, a micro-inertia term is also taken into account (in a consistent and rigorous manner by considering the proper kinetic-energy density) and this leads to an *explicit* appearance of the intrinsic material length h. We emphasize that quasi-static approaches cannot include explicitly the size of the material cell in their governing equations. In these approaches, rather, a characteristic length appears in the governing equations only through the gradient coefficient c (which has dimensions of $[\text{length}]^2$) in the gradient theory without couple-stress effects or the ratio (η/μ) (which again has dimensions of $[\text{length}]^2$) in the couple-stress theory without the effects of collinear dipolar forces, where η is the couple-stress modulus and μ is the shear modulus of the material. Of course, one of the quantities c or (η/μ) also appears within a dynamic analysis, which therefore may allow for an interrelation of the two different characteristic lengths (the one introduced in the strain energy and the other introduced in the kinetic energy—see relative works by Georgiadis et al. [22] and Georgiadis and Velgaki [23]). Indeed, by comparing the forms of dispersion curves of Rayleigh waves obtained by the dipolar (''pure'' gradient and couple-stress) approaches with the ones obtained by the atomic-lattice analysis of Gazis et al. [25], it can be estimated that c is of the order of $(0.1h)^2$, [22], and η is of the order of $0.1\mu h^2$, [23].

The mathematical analysis of the dynamical problem here presents some novel features related to the Wiener-Hopf technique not encountered in dealing with the static case. The Wiener-Hopf technique is employed to obtain exact solutions in both cases, and also the Williams technique is employed for an asymptotic determination of the near-tip fields. Also, since the gradient formulation exhibits a *singular-perturbation* character, the concept of a *boundary layer* is employed to accomplish the solution. On the other hand, the gradient formulation demands extended definitions of the J-integral and the energy release rate. It is further proved, by utilizing some theorems of distribution theory, that both energy quantities remain bounded despite the hypersingular behavior of the near-tip stress field. Finally, physical aspects of the solution are discussed with particular reference to the closure of the crack faces and the nature of cohesive tractions.

2 Fundamentals of the Dipolar Gradient Elasticity

A brief account of the Mindlin-Green-Rivlin theory, [1–3], pertaining to the elastodynamics of homogeneous and isotropic materials is given here. If a continuum with microstructure is viewed as a collection of subparticles (micromedia) having the form of unit cells (cubes), the following expression of the kinetic-energy density (kinetic energy per unit macrovolume) is obtained with respect to a Cartesian coordinate system $Ox_1x_2x_3$, [1],

$$T=\frac{1}{2}\rho\dot{u}_p\dot{u}_p+\frac{1}{6}\rho h^2(\partial_p\dot{u}_q)(\partial_p\dot{u}_q), \tag{1}$$

where ρ is the mass density, $2h$ is the size of the cube edges, u_p is the displacement vector, $\partial_p(\,)\equiv\partial(\,)/\partial x_p$, $\dot{(\,)}\equiv\partial(\,)/\partial t$ with t denoting the time, and the Latin indices span the range (1,2,3). We also notice that Georgiadis et al. [22] by using the concept of *internal* motions have obtained (1) in an alternative way to that by Mindlin [1]. In the RHS of Eq. (1), the second term representing the effects of velocity gradients (a term not encountered within classical continuum mechanics) reflects the greater detail with which the dipolar theory describes the motion.

Next, the following expression of the strain-energy density is postulated:

$$W=\frac{1}{2}c_{pqsj}\varepsilon_{pq}\varepsilon_{sj}+\frac{1}{2}d_{pqsjlm}\kappa_{pqs}\kappa_{jlm}, \tag{2}$$

where (c_{pqsj},d_{pqsjlm}) are tensors of material constants, $\varepsilon_{pq}=(1/2)(\partial_pu_q+\partial_qu_p)$ is the linear strain tensor, and $\kappa_{pqs}=\partial_p\varepsilon_{qs}$ is the strain gradient. Notice that in the tensors c_{pqsj} and d_{pqsjlm} (which are of even rank) the number of independent components can be reduced to yield isotropic constitutive relations. Such an isotropic behavior is considered here. Again, the form in (2) can be viewed as a more accurate description of the constitutive response than that provided by the classical elasticity, if one thinks of a series expansion for W containing higher-order strain gradients. Also, one may expect that the additional term (or terms) will be significant in the vicinity of stress-concentration points where the strain undergoes very steep variations.

Then, pertinent stress tensors can be defined by taking the variation of W

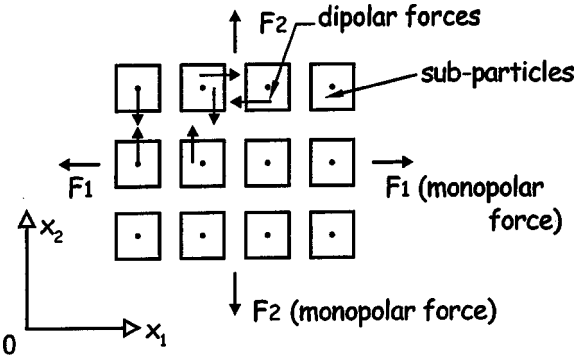


Fig. 1 Monopolar (external) and dipolar (internal) forces acting on an ensemble of subparticles in a material with microstructure

$$\tau_{pq}=\frac{\partial W}{\partial \varepsilon_{pq}}, \tag{3a}$$

$$m_{pqs}=\frac{\partial W}{\partial \kappa_{pqs}}\equiv\frac{\partial W}{\partial(\partial_p \varepsilon_{qs})}, \tag{3b}$$

where $\tau_{pq}=\tau_{qp}$ is the Cauchy (in Mindlin's notation) stress tensor and $m_{pqs}=m_{psq}$ is the dipolar (or double) stress tensor. The latter tensor follows from the notion of *multipolar* forces, which are antiparallel forces acting between the micro-media contained in the continuum with microstructure (see Fig. 1). As explained by Green and Rivlin [2] and Jaunzemis [26], the notion of multipolar forces arises rather naturally if one considers a series expansion for the mechanical power $\mathcal{M}$ containing higher-order velocity gradients, i.e., $\mathcal{M}=F_p\dot{u}_p+F_{pq}(\partial_p\dot{u}_q)+F_{pqs}(\partial_p\partial_q\dot{u}_s)+\ldots$, where F_p are the usual forces (monopolar forces) within classical continua and $(F_{pq},F_{pqs},\ldots)$ are the multipolar forces (dipolar or double forces, triple forces and so on) within generalized continua. In this way, the resultant force on an ensemble of subparticles can be viewed as being decomposed into *external* and *internal* forces with the latter ones being self-equilibrating (see Fig. 1). However, these self-equilibrating forces (which are multipolar forces) produce *nonvanishing* stresses, the multipolar stresses. Examples of force systems of the dipolar collinear or noncollinear type are given, e.g., in Jaunzemis [26] and Fung [27].

As for the notation of dipolar forces and stresses, the first index of the forces denotes the orientation of the lever arm between the forces and the second index the orientation of the pair of the forces; the same meaning is attached to the last two indices of the stresses, whereas the first index denotes the orientation of the normal to the surface on which the stress acts. The dipolar forces F_{pq} have dimensions of [force][length]; their diagonal terms are double forces without moment and their off-diagonal terms are double forces with moment. The antisymmetric part $F_{[pq]}=(1/2)(x_pF_q-x_qF_p)$ gives rise to couple stresses. Here, we do not consider couple-stress effects emphasizing that this is compatible with the particular choice of the form of W in (2), i.e., a form dependent upon the strain gradient but completely independent upon the rotation gradient.

Further, the equations of motion and the tractionboundary conditions along a smooth boundary can be obtained either from Hamilton's principle (Mindlin [1]) or from the momentum balance laws and their application on a material tetrahedron (Georgiadis et al. [22]):

$$\partial_p(\tau_{pq}-\partial_s m_{spq})=\rho\ddot{u}_q-\frac{\rho h^2}{3}(\partial_{pp}\ddot{u}_q), \tag{4}$$

$$n_q(\tau_{qs}-\partial_p m_{pqs})-D_q(n_p m_{pqs})+(D_l n_l)n_p n_q m_{pqs}+\frac{\rho h^2}{3}n_r(\partial_r\ddot{u}_s)=P_s^{(n)}, \tag{5a}$$

$$n_q n_r m_{qrs}=R_s^{(n)}, \tag{5b}$$

where body forces are absent, $D_p(\,)=\partial_p(\,)-n_pD(\,)$, $D(\,)=n_l\partial_l(\,)$, n_s is the unit outward-directed vector normal to the boundary, $P_s^{(n)}$ is the surface force per unit area (monopolar traction), and $R_s^{(n)}$ is the surface double force per unit area (dipolar traction).

Finally, it is convenient for calculations to introduce another quantity, which is a kind of "balance stress" (see Eq. (7) below), and is defined as

$$\sigma_{pq}=\tau_{pq}+\alpha_{pq}, \tag{6}$$

where $\alpha_{qs}=(\rho h^2/3)(\partial_q\ddot{u}_s)-\partial_p m_{pqs}$. With this definition, Eq. (4) takes the more familiar form

$$\partial_p\sigma_{pq}=\rho\ddot{u}_q. \tag{7}$$

Notice that σ_{pq} is not an objective quantity since it contains the acceleration terms $(\rho h^2/3)(\partial_q\ddot{u}_s)$. These micro-inertia terms also are responsible for the asymmetry of σ_{pq}. This, however, does not pose any inconsistency but reflects the role of micro-inertia and the nonstandard nature of the theory. In the quasi-static case, where the acceleration terms are absent, σ_{pq} is an objective tensor. On the other hand, the constitutive equations should definitely obey the principle of objectivity (cf. Eqs. (9) and (10) below).

Now, the simplest possible form of constitutive relations is obtained by taking an isotropic version of the expression in (2) involving only three material constants. This strain-energy density function reads

$$W=\frac{1}{2}\lambda\varepsilon_{pp}\varepsilon_{qq}+\mu\varepsilon_{pq}\varepsilon_{pq}+\frac{1}{2}\lambda c(\partial_s\varepsilon_{pp})(\partial_s\varepsilon_{qq})+\mu c(\partial_s\varepsilon_{pq})(\partial_s\varepsilon_{pq}), \tag{8}$$

and leads to the constitutive relations

$$\tau_{pq}=\lambda\delta_{pq}\varepsilon_{ss}+2\mu\varepsilon_{pq}, \tag{9}$$

$$m_{spq}=c\partial_s(\lambda\delta_{pq}\varepsilon_{jj}+2\mu\varepsilon_{pq}), \tag{10}$$

where (λ,μ) are the standard Lamé's constants, c is the gradient coefficient (material constant with dimensions of $[\text{length}]^2$), and δ_{pq} is the Kronecker delta. Equations (9) and (10) written for a general three-dimensional state will be employed below only for an antiplane shear state.

In summary, Eqs. (4), (5), (9), and (10) are the governing equations for the isotropic dipolar-gradient elasticity with no couple stresses. Combining (4), (9), and (10) leads to the field equation of the problem. Pertinent *uniqueness* theorems have been proved for various forms of the general theory (Mindlin and Eshel [4], Achenbach et al. [9], and Ignaczak [28]) on the basis of *positive definiteness* of the strain-energy density. The latter restriction requires, in turn, the following inequalities for the material constants appearing in the theory employed here (Georgiadis et al. [22]): $(3\lambda+2\mu)>0$, $\mu>0$, $c>0$. In addition, *stability* for the field equation in the general inertial case was proved in [22] and to accomplish this the condition $c>0$ is a necessary one (we notice incidentally that some heuristic gradient-like approaches not employing the rigorous Mindlin-Green-Rivlin theory appeared in the literature that take a negative c—their authors, unfortunately, do not realize that stability was lost in their field equation). Finally, the analysis in [22] provides the order-of-magnitude estimate $(0.1h)^2$ for the gradient coefficient c, in terms of the intrinsic material length h.

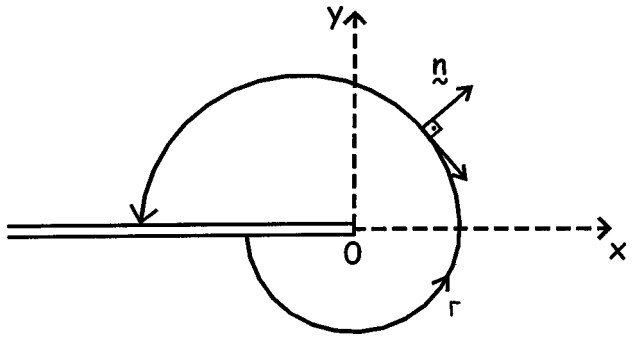


Fig. 2 A crack under a remotely applied antiplane shear loading. The contour Γ surrounding the crack tip serves for the definition of the *J*-integral.

3 Formulation of the Quasi-Static Mode III Crack Problem, the *J*-Integral, and the Energy Release Rate

Consider a crack in a body with microstructure under a quasi-static antiplane shear state (see Fig. 2). As will become clear in the next two sections, the semi-infinite crack model serves in a *boundary layer* type of analysis of any crack problem provided that the crack faces in the problem under consideration are traction free. It is assumed that the mechanical behavior of the body is determined by the Eqs. (4), (5), (9), and (10) of the previous section. An $Oxyz$ Cartesian coordinate system coincident with the system $Ox_1x_2x_3$ utilized previously is attached to that body, and an antiplane shear loading is taken in the direction of z-axis. Also, a *pure* antiplane shear state will be reached, if the body has the form of a thick slab in the z-direction. In such a case, the following two-dimensional field is generated:

$$u_x = u_y = 0, \tag{11a}$$

$$u_z \equiv w \neq 0, \tag{11b}$$

$$w \equiv w(x,y), \tag{11c}$$

and Eqs. (8)–(10) take the forms

$$W = \mu(\varepsilon_{xz}^2 + \varepsilon_{yz}^2) + \mu c\left[\left(\frac{\partial \varepsilon_{xz}}{\partial x}\right)^2 + \left(\frac{\partial \varepsilon_{xz}}{\partial y}\right)^2 + \left(\frac{\partial \varepsilon_{yz}}{\partial x}\right)^2 + \left(\frac{\partial \varepsilon_{yz}}{\partial y}\right)^2\right], \tag{12}$$

$$\tau_{xz} = \mu \frac{\partial w}{\partial x}, \tag{13a}$$

$$\tau_{yz} = \mu \frac{\partial w}{\partial y}, \tag{13b}$$

$$m_{xxz} = \mu c \frac{\partial^2 w}{\partial x^2}, \tag{14a}$$

$$m_{xyz} = \mu c \frac{\partial^2 w}{\partial x \partial y}, \tag{14b}$$

$$m_{yxz} = \mu c \frac{\partial^2 w}{\partial x \partial y}, \tag{14c}$$

$$m_{yyz} = \mu c \frac{\partial^2 w}{\partial y^2}. \tag{14d}$$

Further, (4) provides the equation of equilibrium

$$\frac{\partial}{\partial x}\left(\tau_{xz} - \frac{\partial m_{xxz}}{\partial x} - \frac{\partial m_{yxz}}{\partial y}\right) + \frac{\partial}{\partial y}\left(\tau_{yz} - \frac{\partial m_{xyz}}{\partial x} - \frac{\partial m_{yyz}}{\partial y}\right) = 0, \tag{15}$$

which along with (13) and (14) leads to the following field equation of the problem

$$c\nabla^4 w - \nabla^2 w = 0, \tag{16}$$

where $\nabla^2 = (\partial^2/\partial x^2) + (\partial^2/\partial y^2)$ and $\nabla^4 = \nabla^2\nabla^2$. Finally, one may utilize σ_{pq} defined in (6) for more economy in writing some equations in the ensuing analysis. The antiplane shear components of this quantity are as follows:

$$\sigma_{xz} = \mu\left(\frac{\partial w}{\partial x}\right) - \mu c \nabla^2\left(\frac{\partial w}{\partial x}\right), \tag{17a}$$

$$\sigma_{yz} = \mu\left(\frac{\partial w}{\partial y}\right) - \mu c \nabla^2\left(\frac{\partial w}{\partial y}\right). \tag{17b}$$

Assume now that the cracked body is under a *remotely* applied loading that is also *antisymmetric* about the x-axis (crack plane). Also, the crack faces are traction-free. Due to the antisymmetry of the problem, only the upper half of the cracked domain is considered. Then, the following conditions can be written along the plane $(-\infty < x < \infty, y = 0)$:

$$t_{yz} \equiv \tau_{yz} - \frac{\partial m_{xyz}}{\partial x} - \frac{\partial m_{yyz}}{\partial y} - \frac{\partial m_{yxz}}{\partial x} = 0 \quad \text{for } (-\infty < x < 0, y = 0), \tag{18}$$

$$m_{yyz} = 0 \quad \text{for } (-\infty < x < 0, y = 0), \tag{19}$$

$$w = 0 \quad \text{for } (0 < x < \infty, y = 0), \tag{20}$$

$$\frac{\partial^2 w}{\partial y^2} = 0 \quad \text{for } (0 < x < \infty, y = 0), \tag{21}$$

where (18) and (19) directly follow from Eqs. (5) (notice also that (18) can be written as $\sigma_{yz} - (\partial m_{yxz}/\partial x) = 0$ by using the σ_{pq} quantity), t_{yz} is defined as the *total monopolar* stress, and (20) together with (21) always guarantee an antisymmetric displacement field w.r.t. the line of the crack prolongation. The definition of the stress t_{yz} follows from (5*a*). The problem described by (11)–(21) will be considered by both the asymptotic Williams method and the exact Wiener-Hopf technique. Notice finally that no difficulty will arise by having zero boundary conditions along the crack faces since, eventually, the solution will be matched at regions where gradient effects are not dominant (i.e., for $x \gg c^{1/2}$) with the K_{III} field of the classical theory and in this way the remote loading will appear in the solution.

Next, we present the new extended definitions of the J-integral and the energy release rate G. These definitions of the energy quantities are pertinent to the present framework of dipolar gradient elasticity and to the aforementioned case of a crack in a quasi-static antiplane shear state. By following relative concepts from Rice [29,30], we first introduce the definition

$$J = \int_\Gamma \left(W dy - \overline{P}_z^{(n)} \frac{\partial w}{\partial x} d\Gamma - \overline{R}_z^{(n)} D\left(\frac{\partial w}{\partial x}\right) d\Gamma \right), \tag{22}$$

where Γ is a two-dimensional contour surrounding the crack tip (see Fig. 2), whereas the monopolar and dipolar tractions $\overline{P}_z^{(n)}$ and $\overline{R}_z^{(n)}$ on Γ are given as

$$\overline{P}_z^{(n)} = n_q(\tau_{qz} - \partial_p m_{pqz}) - D_q(n_p m_{pqz}) + (D_l n_l) n_p n_q m_{pqz}, \tag{23a}$$

$$\overline{R}_z^{(n)} = n_p n_q m_{pqz}. \tag{23b}$$

In the above expressions, n_p with components (n_x, n_y) is the unit outward-directed vector normal to Γ, the differential operators D and D_p were defined in Section 2, W is the strain-energy density function given by (12), and the indices (l,p,q) take the values x and y only.

Of course, the above expressions for the tractions on Γ are compatible with Eqs. (5). Further, it can be proved that the integral in (22) is path independent by following Rice's, [29], procedure. Path independence is of great utility since it permits alternate choices of integration paths that may lead to a direct

evaluation of J. We should mention at this point that (22) is quite novel within the present version of the gradient theory (i.e., a form *without* couple stresses), but expressions for J within the couple-stress theory were presented before by Atkinson and Leppington [31], Zhang et al. [18], and Lubarda and Markenscoff [32]. In particular, the latter work gives a systematic derivation of conservation integrals by the use of Noether's theorem. Finally, we notice that the way the J-integral will be evaluated below is quite different than that by Zhang et al. [18]. Indeed, use of the theory of distributions in the present work leads to a very simple way to evaluate J (see Section 7 below).

As for the energy release rate (ERR) now, we also modify the classical definition in order to take into account a higher-order term that is compatible with the present strain-gradient framework

$$G=\lim_{\Delta x\to 0}\frac{\displaystyle\int_0^{\Delta x}\left[t_{yz}(x,y=0)\cdot w(x,y=0)+m_{yyz}(x,y=0)\cdot\frac{\partial w(x,y=0)}{\partial y}\right]dx}{\Delta x}, \tag{24}$$

where Δx is the small distance of a crack advancement.

Of course, any meaningful crack-tip field given as solution to an associated mathematical problem, should result in a *finite* value for the energy quantities defined above. Despite the strong singularity of the stress field obtained in Sections 5 and 6, the results of Section 7 prove that J and G are indeed bounded.

4 Asymptotic Analysis by the Williams Method

As is well known, Williams [33,34] (see also Barber [35]) developed a method to explore the nature of the stress and displacement field near wedge corners and crack tips. This is accomplished by attaching a set of (r,θ) polar coordinates at the corner point and by expanding the stress field as an asymptotic series in powers of r. By following this method here we are concerned, in a way, only with the field components in the sharp crack at very small values of r, and hence we imagine looking at the tip region through a strong microscope so that situations like the ones, e.g., on the left of Fig. 3 (i.e., a finite length crack, an edge crack or a crack in a strip) appear to us like the semi-infinite crack on the right of this figure. The magnification is so large that the other surfaces of the body, including the loaded remote boundaries, appear enough far away for us to treat the body as an "infinite wedge" with "loading at infinity." The field is, of course, a complicated function of (r,θ) but near to the crack tip (i.e., as $r\to 0$) we seek to expand it as a series of separated variable terms, each of which satisfies the traction-free boundary conditions on the crack faces.

In view of the above, we consider the following separated form $w(r,\theta)=r^{\omega+1}u(\theta)$, where the displacement satisfies (16). Further, if only the dominant singular terms in (16) are retained, the PDE of the problem becomes $\nabla^4 w=0$, where $\nabla^4=\nabla^2\nabla^2=(\partial^2/\partial r^2+1/r\,\partial/\partial r+1/r^2\partial^2/\partial\theta^2)^2$. Also, in view of the definitions of stresses as combinations of derivatives of w and by retaining again only the dominant singular terms, the boundary conditions $t_{yz}(x,y=\pm 0)=0$ and $m_{yyz}(x,y=\pm 0)=0$ will give at $\theta=\pm\pi$

$$\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{1}{r^2}\right)\frac{\partial w}{\partial\theta}=0, \tag{25a}$$

$$\left(\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}\right)w=0. \tag{25b}$$

In addition, the pertinent antisymmetric solution (i.e., with odd behavior in θ) to the equation $\nabla^4 w=0$ has the following general form:

$$w=r^{\omega+1}(A_1\sin[(\omega+1)\theta]+A_2\sin[(\omega-1)\theta]), \tag{26}$$

where ω is (in general) a complex number and (A_1,A_2) are unknown constants. Now, (25) and (26) provide the *eigenvalue* problem

$$(\omega+1)\cos[(\omega+1)\pi]\cdot A_1-3(\omega-1)\cos[(\omega-1)\pi]\cdot A_2=0, \tag{27a}$$

$$(\omega+1)\sin[(\omega+1)\pi]\cdot A_1+(\omega-3)\sin[(\omega-1)\pi]\cdot A_2=0. \tag{27b}$$

For a nontrivial solution to exist, the determinant of the coefficients of (A_1,A_2) in the above system should vanish and this gives the result: $\sin(2\omega\pi)=0\Rightarrow\omega=0,1/2,1,3/2,2,\ldots$. Next, by observing from (12) that the strain-energy density W behaves at most as $(\partial^2 w/\partial r^2)$ or, by using the form $w(r,\theta)=r^{\omega+1}u(\theta)$, no worse than $r^{\omega-1}$, we conclude that the maximum eigenvalue allowed by the *integrability* condition of the strain-energy density is $\omega=1/2$.

The above analysis suggests that the general asymptotic solution is of the form $w(r,\theta)=r^{3/2}u(\theta)$, which by virtue of (26) and (27*b*) becomes

$$w(r,\theta)=Ar^{3/2}[3\sin(\theta/2)-5\sin(3\theta/2)], \tag{28}$$

where $A\equiv -A_1$ and the other constant in (26) is given by (27*b*) as $A_2=3A_1/5$. The constant A (amplitude of the field) is left unspecified by the Williams technique but still the nature of the near-tip field has been determined. Finally, the total monopolar stress has the following asymptotic behavior:

$$t_{yz}(x,y=0)=O(x^{-3/2})\quad\text{as}\quad x\to+0. \tag{29}$$

This asymptotic behavior will also be corroborated by the results of the exact analysis in the next section.

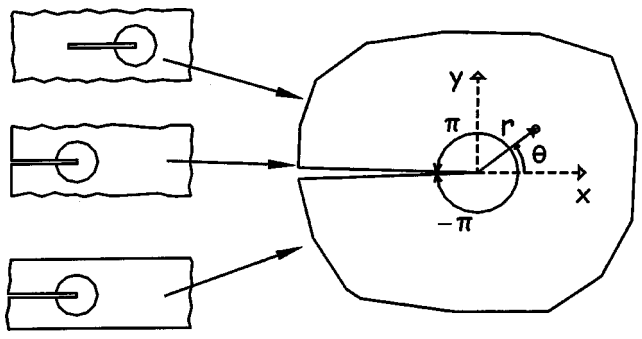


Fig. 3 William's method: the near-tip fields of (i) a finite length crack, (ii) an edge crack, and (iii) a cracked strip correspond to the field generated in a body with a semi-infinite crack

5 Exact Analysis by the Wiener-Hopf Method

An *exact* solution to the problem described by (11)–(21) will be obtained through two-sided Laplace transforms (see, e.g., van der Pol and Bremmer [36] and Carrier et al. [37]), the Wiener-

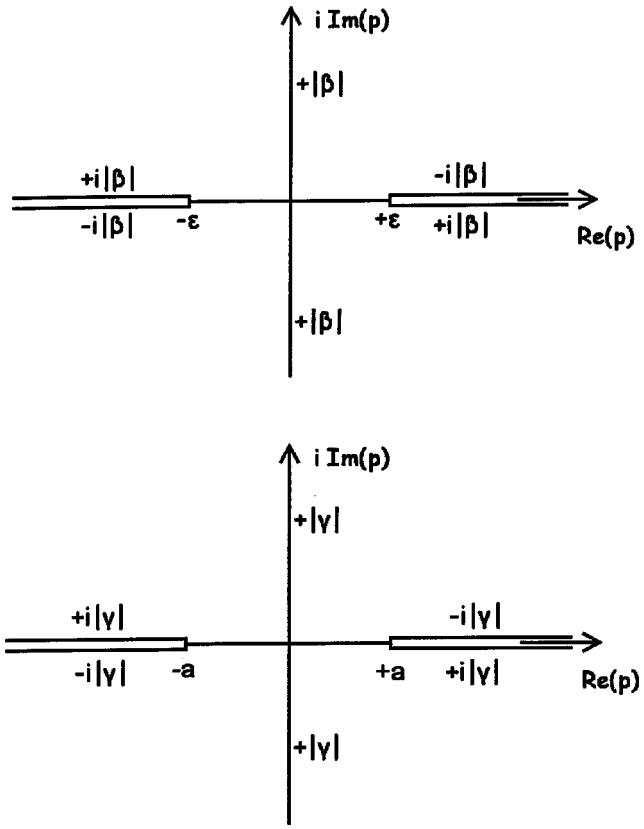


Fig. 4 Branch cuts for the functions (β,γ)

Hopf technique (see, e.g., Roos [38] and Mittra and Lee [39]) and certain results from the theory of distributions (see, e.g., Gel'fand and Shilov [40] and Lauwerier [41]).

The direct and inverse two-sided Laplace transforms are defined as

$$f^*(p,y)=\int_{-\infty}^{\infty} f(x,y)e^{-px}dx, \tag{30a}$$

$$f(x,y)=\frac{1}{2\pi i}\int_{Br} f^*(p,y)e^{px}dp, \tag{30b}$$

where *Br* denotes the Bromwich inversion path *within* the region of analyticity of the function $f^*(p,y)$ in the complex p-plane. Transforming (16) with (30*a*) gives the ODE

$$c\frac{d^4w^*}{dy^4}+(2cp^2-1)\frac{d^2w^*}{dy^2}+(cp^4-p^2)w^*=0. \tag{31}$$

The above equation has the following general solution that is bounded as $y\to+\infty$

$$w^*(p,y)=B(p)\cdot\exp(-\beta y)+C(p)\cdot\exp(-\gamma y)\quad\text{for}\quad y\geqslant 0, \tag{32}$$

where $B(p)$ and $C(p)$ are yet unknown functions, $\beta\equiv\beta(p)=(\varepsilon^2-p^2)^{1/2}$ with ε being a real number such that $\varepsilon\to+0$, and $\gamma\equiv\gamma(p)=[(1/c)-p^2]^{1/2}\equiv(a^2-p^2)^{1/2}$ with $a=(1/c)^{1/2}$. In fact, introducing ε facilitates the introduction of the branch cuts for $\beta=(-p^2)^{1/2}$-cf. [20] and [37] for this procedure as applied to related situations. To obtain a bounded solution as $y\to+\infty$, the p-plane should be cut in the way shown in Fig. 4. This introduction of branch cuts secures that the functions (β,γ) are single-valued and that $\mathrm{Re}(\beta)>0$ and $\mathrm{Re}(\gamma)>0$ along the Bromwich path.

The transformed expressions for the stresses that enter the boundary conditions are also quoted (for convenience, the σ_{yz} quantity is employed in the boundary conditions)

$$\sigma_{yz}^*(p,y)=-\mu B\beta e^{-\beta y}, \tag{33}$$

$$m_{yyz}^*(p,y)=\mu(Bc\beta^2e^{-\beta y}+Cc\gamma^2e^{-\gamma y}), \tag{34}$$

$$m_{yxz}^*(p,y)=-\mu p(Bc\beta e^{-\beta y}+Cc\gamma e^{-\gamma y}). \tag{35}$$

Next, in preparation for formulating the Wiener-Hopf equation, the one-sided Laplace transforms of the unknown total monopolar stress $t_{yz}(x>0,y=0)$ ahead of the crack tip and the unknown crack-face displacement $w(x<0,y=0)$ are defined

$$T^+(p)=\int_0^{\infty} t_{yz}(x,y=0)e^{-px}dx\equiv\int_0^{\infty}\left[\sigma_{yz}(x,y=0)-\frac{\partial m_{yxz}(x,y=0)}{\partial x}\right]e^{-px}dx, \tag{36}$$

$$W^-(p)=\int_{-\infty}^{0} w(x,y=0)e^{-px}dx. \tag{37}$$

Further, we assume the following *finiteness* conditions at $x\to\pm\infty$: $|t_{yz}(x,y=0)|<M\cdot\exp(-p_Tx)$ for $x\to+\infty$ and $|w(x,y=0)|<N\cdot\exp(p_Wx)$ for $x\to-\infty$, where (M,N,p_T,p_W) are positive constants. As a consequence, $T^+(p)$ is analytic and defined in the right half-plane $-p_T<\mathrm{Re}(p)$ (the "plus" half-plane), while $W^-(p)$ is analytic and defined in the left half-plane $\mathrm{Re}(p)<p_W$ (the "minus" half-plane).

Then, enforcement of boundary conditions results in the following equations:

$$T^+(p)=\sigma_{yz}^*(p,y=0)-p\cdot m_{yxz}^*(p,y=0), \tag{38}$$

$$W^-(p)=w^*(p,y=0). \tag{39}$$

The above equations along with the equation $\partial^2w^*(p,y=0)/\partial y^2=0$, Eqs. (33)–(35) and the general solution in (32) provide an algebraic system of three equations in four unknowns (the functions T^+, W^-, B, C). Finally, eliminating B and C in this system leads to the following Wiener-Hopf problem

$$\frac{T^+(p)}{(a+p)^{1/2}}=-\mu cp^2(a-p)^{1/2}\cdot L(p)\cdot W^-(p), \tag{40}$$

where the kernel function $L(p)$ is given as

$$L(p)=-cp^2\left[1+\frac{1-cp^2}{cp^2}\frac{(a^2-p^2)^{1/2}}{(\varepsilon^2-p^2)^{1/2}}\right]. \tag{41}$$

The next target will be to determine both T^+ and W^- from the single Eq. (40). This will be effected through the use of elements of the theories of complex variables, integral transforms, and distributions (theorem of analytic continuation, extended Liouville's theorem, Abel-Tauber asymptotic theorems, transforms of distributions). First, we check that the function $L(p)$ has no zeros in the complex plane. This was found independently by using both the principle of the argument, [37], and the program MATHEMATICA™. We notice that unlike the current static case, the counterpart kernel function in the dynamic case exhibits two (nonextraneous) zeros, a fact that modifies somehow the standard Wiener-Hopf method. Further, we find that the asymptotic behavior of the kernel is $\lim_{|p|\to\infty}L(p)=-3/2$ and this leads us to introduce a modified kernel given as $N(p)=-(2/3)\cdot L(p)$, which possesses the desired asymptotic property $\lim_{|p|\to\infty}N(p)=1$. Indeed, this new form of the kernel facilitates its *product splitting* by the use of Cauchy's integral theorem. The Wiener-Hopf equation takes now the form

$$\frac{T^+(p)}{(a+p)^{1/2}}=\left(-\frac{3}{2}\right)(-\mu c)p^2(a-p)^{1/2}N(p)\cdot W^-(p), \tag{42}$$

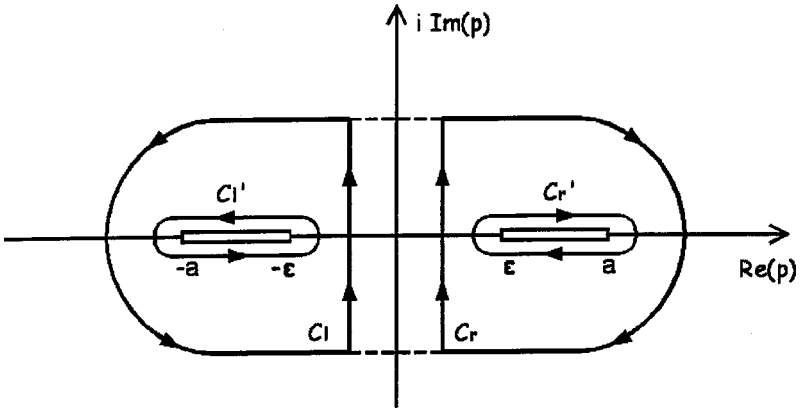


Fig. 5 Contour integrations for the factorization of the kernel function in Eq. (42)

and the kernel is written as the following product of two analytic and nonzero functions defined in pertinent half-plane domains of the complex plane, [38,39],

$$N(p)=N^{+}(p)\cdot N^{-}(p), \tag{43}$$

where

$$N^{+}(p)=\exp\left\{-\frac{1}{2\pi i}\int_{C_l}\frac{\ln[N(\zeta)]}{\zeta-p}d\zeta\right\}, \tag{44a}$$

$$N^{-}(p)=\exp\left\{\frac{1}{2\pi i}\int_{C_r}\frac{\ln[N(\zeta)]}{\zeta-p}d\zeta\right\}. \tag{44b}$$

The use of Cauchy's integral theorem is depicted in Fig. 5. $N^{+}(p)$ is analytic and nonzero in $\mathrm{Re}(p)>-\varepsilon$ and $N^{-}(p)$ is analytic and nonzero in $\mathrm{Re}(p)<\varepsilon$. The original integration paths (C_l,C_r) extend parallel to the imaginary axis in the complex ζ-plane. Finally, an alteration of the integration contour (also depicted in Fig. 5) along with use of Cauchy's theorem and Jordan's lemma allows taking as equivalent integration paths the (C_l',C_r') contours around the branch cuts extending along $-a<\zeta<-\varepsilon$ and $\varepsilon<\zeta<a$. This eventually leads to the following forms of the sectionally analytic functions $N^{\pm}(p)$:

$$N^{\pm}(p)=\exp\left\{\frac{1}{\pi}\int_0^a \arctan\left[\frac{(a^2-\zeta^2)^{3/2}}{\zeta^3}\right]\frac{d\zeta}{\zeta\pm p}\right\}, \tag{45}$$

with the property $N^{+}(-p)=N^{-}(p)$.

With the product factorisation in hand, Eq. (42) takes the following form that defines a function $E(p)$:

$$\frac{T^{+}(p)}{N^{+}(p)\cdot(a+p)^{1/2}}=\frac{3\mu c}{2}p^2(a-p)^{1/2}N^{-}(p)\cdot W^{-}(p)\equiv E(p). \tag{46}$$

The above equation defines $E(p)$ only in the strip $-\varepsilon<\mathrm{Re}(p)<0$. But the first member in the equation is a nonzero analytic function in $\mathrm{Re}(p)>-\varepsilon$, and the second member is a nonzero analytic function in $\mathrm{Re}(p)<0$. Then, in view of the theorem of analytic continuation (or identity theorem for single-valued analytic functions), the two members define one and the same function that is analytic over the whole p-plane, [38,39]. In other words, $E(p)$ is an *entire* function. Polynomial and exponential functions are the types of entire functions. The case of an exponential function (i.e., a function of the form $\exp[g(p)]$, where $g(p)$ is a polynomial) should be excluded because such a function has an *essential singularity* at infinity. Indeed, an exponential growth of the functions involved in (46) would result in violating the so-called *edge condition*, i.e., the condition of bounded energy density around the geometrical singularity (crack edge) in the physical domain. Therefore, $E(p)$ should be a polynomial since only algebraic growth of the fields in the neighborhood of the crack tip is allowed. Further, determining the coefficients of this polynomial will lead to the desired decoupling of $T^{+}(p)$ and $W^{-}(p)$. Below, we determine the form of $E(p)$ by the use of asymptotic analysis.

In particular, we will use theorems of the Abel and Tauber type having the form

$$\lim_{x\to 0}f(x)\overset{\mathrm{LT}}{\leftrightarrow}\lim_{|p|\to\infty}f^{*}(p), \tag{47}$$

$$\lim_{x\to \infty}f(x)\overset{\mathrm{LT}}{\leftrightarrow}\lim_{|p|\to 0}f^{*}(p), \tag{48}$$

where the symbol $\overset{\mathrm{LT}}{\leftrightarrow}$ means that the image function $f^{*}(p)$ and the original function $f(x)$ are connected through the *one-sided* Laplace-transform relations $f^{*}(p)=\int_0^{\infty}f(x)e^{-px}dx$ and $f(x)=(1/2\pi i)\int_{Br}f^{*}(p)e^{px}dp$, and p is a complex variable which in (47) and (48) tends to infinity or zero along paths in the pertinent half-plane of convergence (analyticity). Relations (47) and (48) hold under certain conditions given, e.g., in [36]. Also, the extended Liouville's theorem, [39], will be utilized. Referring to (46), this states that if $T^{+}(p)\cdot[N^{+}(p)\cdot(a+p)^{1/2}]^{-1}=O(p^{\nu})$ and $(3\mu c/2)p^2(a-p)^{1/2}N^{-}(p)\cdot W^{-}(p)=O(p^{\xi})$ in the respective half-planes of analyticity, then $E(p)$ is a polynomial of degree not exceeding the minimum of $([\nu],[\xi])$, where the symbol $[\ ]$ denotes the integral part of a number.

Now as a first possibility of the near-tip behavior, one may adopt a behavior of the total monopolar stress and the crack-face displacement that is analogous to the classical fracture mechanics behavior, viz.

$$t_{yz}(x,y=0)=O(x^{-1/2})\quad \text{as}\quad x\to+0, \tag{49a}$$

$$w(x,y=0)=O(x^{1/2})\quad \text{as}\quad x\to-0. \tag{49b}$$

This field gives by (47) and the transformation formula $x^{\kappa}\overset{\mathrm{LT}}{\leftrightarrow}\Gamma(\kappa+1)\cdot p^{-\kappa-1}$ (with $\Gamma(\ )$ being the Gamma function and $\kappa>-1$), [36,38], the following asymptotic behavior in the transform domain

$$T^{+}(p)=O(p^{-1/2})\quad \text{as}\quad |p|\to\infty, \tag{50a}$$

$$W^{-}(p)=O(p^{-3/2})\quad \text{as}\quad |p|\to\infty. \tag{50b}$$

Then, Liouville's theorem leads to the conclusion that $E(p)=0$, which, however, is an inadmissible result since it shows that the stress field is zero everywhere (although the cracked body is under loading). Therefore, the possibility of a near-tip behavior given by (49) should be discarded.

Next, prompted by the results of the Williams asymptotic method obtained before, i.e., the results in (28) and (29), we consider the following possibility of near-tip behavior

$$t_{yz}(x,y=0)=O(x^{-3/2}) \quad \text{as} \quad x\to+0, \tag{51a}$$

$$w(x,y=0)=O(x^{3/2}) \quad \text{as} \quad x\to-0. \tag{51b}$$

Here, certain results of the theory of generalized functions will be employed concerning transforms of *singular* distributions, [40,41]. In this connection, we note that the distribution x_+^λ for $\mathrm{Re}(\lambda)>-1$ is identified with the function $x_+^\lambda=x^\lambda$ for $x>0$ and $x_+^\lambda=0$ for $x<0$. For other values of the complex parameter λ (of course, λ here is not to be confused with the Lamé constant) it is defined by analytic continuation of the functional $\langle x_+^\lambda,h\rangle \equiv \int_0^\infty x^\lambda h(x)\,dx$, where $h(x)$ is a test function. In this way, a distribution is obtained for all complex values of λ with the exception of $\lambda=-1,-2,-3,\ldots$. In a similar manner, x_-^λ is defined by starting from $x_-^\lambda=0$ for $x>0$ and $x_-^\lambda=|x|^\lambda$ for $x<0$. Then, (51) and the transformation formula $x^\lambda \overset{\mathrm{LT}}{\leftrightarrow} \Gamma(\lambda+1)\cdot p^{-\lambda-1}$ (with $\lambda\neq-1,-2,-3,\ldots$), [40,41], provide the following asymptotic behavior in the transform domain:

$$T^+(p)=O(p^{1/2}) \quad \text{as} \quad |p|\to\infty, \tag{52a}$$

$$W^-(p)=O(p^{-5/2}) \quad \text{as} \quad |p|\to\infty. \tag{52b}$$

Further the extended Liouville's theorem leads to the conclusion that $E(p)=E_0$, where E_0 is a constant. As shown below this constant will be determined from conditions at remote regions in the physical plane. The previous result is mathematically admissible, while any other case like, e.g., $t_{yz}(x,y=0)=O(x^{-1})$ or $O(x^{-2})$ as $x\to+0$ is precluded since even analytic continuation fails to define one-sided Laplace (or Fourier) transforms of the associated singular distributions (cf. Gel'fand and Shilov [40], p. 171). Of course, it remains to prove that the field in (51) gives a *bounded* value for the energy quantities of J-integral and ERR, despite the hypersingular character of stress. This will be shown in Section 7. Finally, the requirement of boundedness of energy expressions is not only to be imposed on physical grounds but it is generally (Ignaczak [28] and Knowles and Pucik [42]) a necessary condition for uniqueness.

Our task now is to determine E_0. As in the work of Zhang et al. [18], a matching procedure is followed that equates the *inner* solution $\lim_{x\to\infty} t_{yz}(x,y=0)$, as obtained by the present gradient analysis, with the *outer* solution $K_{\mathrm{III}}/(2\pi x)^{1/2}$ provided by the conventional fracture mechanics. K_{III} is the stress intensity factor for each specific problem treated by the conventional fracture mechanics. The latter field (singular solution) dominates over an area that is relatively close to the crack tip but lies outside the domain where gradient effects are pronounced. We notice the following in support of the assertion that this procedure is indeed reasonable: (i) as shown below the stress behaves as $t_{yz}=O(x^{-1/2})$ for $x\to\infty$, (ii) the very form of the field Eq. (16) exhibits the singular-perturbation character of the gradient formulation and therefore suggests a *boundary layer* approach (Van Dyke [43]) to the crack problem (one may observe that an extremely small quantity—the coefficient c—multiplies the higher-order term, which is the one introduced by the nonconventional formulation). Finally, one may observe that this concept is in some respects similar to the one introduced by Rice [44] in analyzing small scale yielding around a crack tip.

The transformed total monopolar stress $T^+(p)$ is given by (46) as

$$T^+(p)=E_0\cdot N^+(p)\cdot(a+p)^{1/2}, \tag{53}$$

an expression that holds for *all* values of the Laplace transform variable p in the right half-plane. For the moment, we need to evaluate only $\lim_{|p|\to0}T^+(p)$ in order to obtain then $\lim_{x\to\infty}t_{yz}(x,y=0)$ by (48). One way to obtain the expression of $\lim_{|p|\to0}N^+(p)$ is to use $\lim_{|p|\to0}N(p)$ and perform a product factorization of the latter limit by *inspection*. This way is easier than finding $\lim_{|p|\to0}N^+(p)$ from (45). Indeed, one may obtain first from (41) and the definition of $N(p)$ the limit $\lim_{|p|\to0}N(p)=2(3c^{1/2})^{-1}(\varepsilon^2-p^2)^{-1/2}$ and then

$$\lim_{|p|\to0} N^+(p)=\left(\frac{2}{3c^{1/2}}\right)^{1/2}\frac{1}{(\varepsilon+p)^{1/2}}. \tag{54}$$

Further, a combination of (53) and (54) provides the limit

$$\lim_{|p|\to0} T^+(p)=E_0\cdot\left(\frac{2}{3c}\right)^{1/2}\frac{1}{p^{1/2}}, \tag{55}$$

which by (48) and the transformation formula $x^\kappa \overset{\mathrm{LT}}{\leftrightarrow} \Gamma(\kappa+1)\cdot p^{-\kappa-1}$ (with $\kappa>-1$) allows writing

$$\lim_{x\to+\infty} t_{yz}(x,y=0)=E_0\cdot\left(\frac{2}{3c}\right)^{1/2}\frac{1}{(\pi x)^{1/2}}. \tag{56}$$

Finally, matching the above expression with $K_{\mathrm{III}}/(2\pi x)^{1/2}$ provides the value of the constant as $E_0=K_{\mathrm{III}}(3c)^{1/2}/2$.

In view of the above, we record the final transformed expressions (valid for all p in the pertinent half-plane of convergence) for the total monopolar stress ahead of the tip and the crack-face displacement

$$T^+(p)=\frac{K_{\mathrm{III}}(3c)^{1/2}}{2}N^+(p)\cdot(a+p)^{1/2}, \tag{57}$$

$$W^-(p)=\frac{K_{\mathrm{III}}}{(3c)^{1/2}\mu p^2(a-p)^{1/2}\cdot N^-(p)}, \tag{58}$$

where it is reminded that $a=(1/c)^{1/2}$, and $N^+(p)$ and $N^-(p)$ are given by (45). Exact expressions for the original functions $t_{yz}(x>0,y=0)$ and $w(x<0,y=0)$ can be derived from (57) and (58) through one-sided Laplace-transform inversions. Such an inversion will be performed in Section 8, where we elaborate more on the stress ahead of the crack tip providing the exact expression and several comparisons. In closing now this section, we give the near-tip asymptotic expressions of $t_{yz}(x>0,y=0)$ and $w(x<0,y=0)$. These expressions, however, suffice for the evaluation of the J-integral and the ERR and possess also much practical importance as explained below.

The limits of the expressions in (57) and (58) for $|p|\to\infty$ are found to be

$$\lim_{|p|\to\infty} T^+(p)=\frac{K_{\mathrm{III}}(3c)^{1/2}}{2}p^{1/2}, \tag{59}$$

$$\lim_{|p|\to\infty} W^-(p)=\frac{K_{\mathrm{III}}}{(3c)^{1/2}\mu}\frac{1}{p^{5/2}}, \tag{60}$$

which by the inversions $p^{1/2} \overset{\mathrm{LT}}{\leftrightarrow} [\Gamma(-1/2)]^{-1}x^{-3/2}=-(2\pi^{1/2})^{-1}x^{-3/2}$ and $p^{-5/2} \overset{\mathrm{LT}}{\leftrightarrow} [\Gamma(5/2)]^{-1}(-x)^{3/2}=4(3\pi^{1/2})^{-1}(-x)^{3/2}$ give the following near-tip field

$$\lim_{x\to+0} t_{yz}(x,y=0)=-\frac{K_{\mathrm{III}}(3c)^{1/2}}{4\pi^{1/2}}\frac{1}{x^{3/2}}, \tag{61}$$

$$\lim_{x\to-0} w(x,y=0)=\frac{4K_{\mathrm{III}}}{3(3\pi c)^{1/2}\mu}(-x)^{3/2}. \tag{62}$$

In view of the fact that K_{III} is the stress intensity factor obtained by a classical elasticity analysis for the same crack problem (same geometry and loading) as that considered through the dipolar gradient approach, Eqs. (61) and (62) provide a kind of *correspondence principle*. This correspondence principle connects any classical fracture mechanics solution (through the pertinent K_{III} value

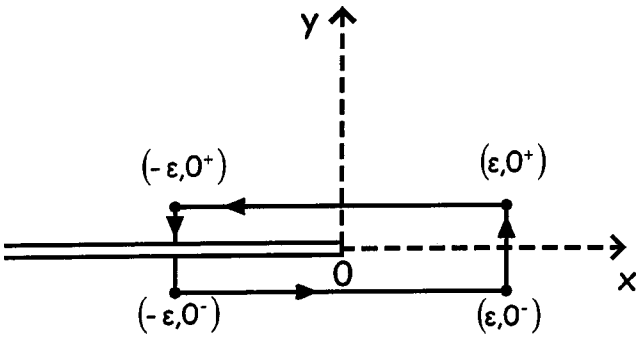


Fig. 6 Rectangular-shaped contour surrounding the crack tip for the evaluations of the *J*-integral and the energy release rate

obtained for each specific problem) with the near-tip field resulting by the nonclassical gradient formulation of the problem in question. Thus, a host of classical fracture mechanics solutions to crack problems may serve within a nonclassical gradient framework as well.

Three final notices pertain to the form of the above asymptotic field. First, the cusp-like closure of the crack faces (a closure smoother than the one predicted by the classical theory) implied by (62) is not unusual in experiments (see, e.g., Mills [45] and Elssner et al. [46]). Secondly, an aggravation of the stress field as compared to the respective result of the conventional theory (this aggravation appears here through the stronger $x^{-3/2}$ singularity) is not unusual in analyses with nonclassical effects (see, e.g., the couple-stress results of Bogy and Sternberg [47] and Zhang et al. [18]). In addition, Prakash et al. [48] have provided an analysis and experimental evidence supporting the possibility of an $x^{-3/2}$ stress singularity in dynamic crack initiation. All this evidence shows that deviations from predictions of classical fracture mechanics are possible in some situations and are, at least, worthy of investigation. Of course, by no means we claim that the results in (61) and (62) carry over to other situations like, e.g., the plane strain/stress case. An appropriate dipolar gradient analysis for the latter case is needed to give the answer. Thirdly, the minus sign in the RHS of (61) shows that the asymptotic gradient crack-tip stress field has a cohesive-traction nature. This point will be further elaborated in Section 8 below. It will be shown also in Section 8 that (61) dominates only within an extremely small region adjacent to the crack tip.

6 Evaluation of the *J*-Integral and the Energy Release Rate (ERR)

The evaluation of the energy quantities is accomplished here by using Fisher's theorem, [49], concerning the product of distributions. For the J-integral, we also consider the new rectangular-shaped contour Γ (see Fig. 6) with vanishing ''height'' along the y-direction and with $\varepsilon \to +0$. This change of contour permits using solely the asymptotic near-tip field in (61) and (62). Notice that Zhang et al. [18] in evaluating the ERR for a mode III crack problem with couple stresses followed a rather involved method based on earlier work by Bueckner [50]. It seems that the procedure followed here is simpler and more direct. Indeed, taking into account (14*d*), (18), (19), and (21), the definitions in (22) and (24) provide the following integral for both energy quantities:

$$J=G=\lim_{\varepsilon\to+0}\left\{2\int_{-\varepsilon}^{\varepsilon} t_{yz}(x,y=0)\cdot\frac{\partial w(x,y=0)}{\partial x}dx\right\}. \quad (63)$$

Now, by using the solution (61) and (62), we obtain

$$J=G=\lim_{\varepsilon\to+0}\left\{2(-1)\frac{K_{\mathrm{III}}(3c)^{1/2}}{4\pi^{1/2}}\times\frac{4K_{\mathrm{III}}}{3(3\pi c)^{1/2}\mu}\frac{3}{2}\int_{-\varepsilon}^{\varepsilon}(x_+)^{-3/2}(x_-)^{1/2}dx\right\}. \quad (64)$$

Further, the product of distributions inside the integral is obtained through the use of Fisher's theorem, [49], i.e., of the operational relation $(x_-)^{\lambda}(x_+)^{-1-\lambda}=-\pi\delta(x)[2\sin(\pi\lambda)]^{-1}$ with $\lambda\neq -1, -2, -3, \ldots$ and $\delta(x)$ being the Dirac delta distribution. Then, in view of the fundamental property of the Dirac delta distribution that $\int_{-\varepsilon}^{\varepsilon}\delta(x)dx=1$, Eq. (64) provides the result

$$J=G=\frac{K_{\mathrm{III}}^2}{2\mu}, \quad (65)$$

which shows that the J-integral and the ERR are *bounded* (despite the hypersingular nature of the near-tip stress) and identical with the respective classical elasticity result. Our findings suggest therefore that, at least for the one-parameter theory of microstructure employed here, the *overall* energy situation (rate of total potential energy) of the cracked body is not affected by the material microstructure and only the *local* crack-tip field is influenced.

7 Exact Expression for the Stress Ahead of the Crack Tip

In this section we elaborate more on the stress ahead of the crack tip $t_{yz}(x>0,y=0)$ and its nature, and also provide comparisons of the exact expression with both the asymptotic form in (61) and the classical $x^{-1/2}$ field. First, an exact one-sided Laplace transform inversion of $T^+(p)$ in (57) will be obtained.

One may write formally

$$t_{yz}(x>0,y=0)=\frac{K_{\mathrm{III}}(3c)^{1/2}}{2}\frac{1}{2\pi i}\int_{\mathrm{Br}}[N^+(p)\cdot(a+p)^{1/2}]e^{px}dp \equiv \frac{K_{\mathrm{III}}(3c)^{1/2}}{2}\frac{1}{2\pi i}I, \quad (66)$$

where the integration variable takes values only in the half-plane $\mathrm{Re}(p)>-\varepsilon$ $(\varepsilon\to+0)$ and any line, in this half-plane, parallel to the $\mathrm{Im}(p)$-axis may serve as the Bromwich path. The I-integral defined above depends upon x and c. I is evaluated by deforming the integration path in the left half-plane (see Fig. 7) where the integrand is nonanalytic, exploiting in this way the existence of branch cuts for the functions $N^+(p)$ and $(a+p)^{1/2}$. Noting the property $\lim_{|p|\to\infty}N^+(p)=1$ and also that $N^+(p)=N(p)/N^-(p)$ (cf. Eq. (43)), the I-integral is written by Cauchy's theorem as

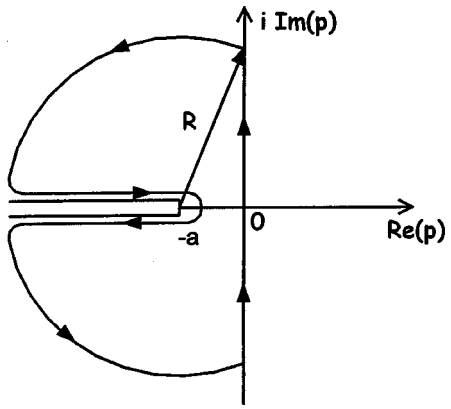


Fig. 7 Contour integration for the evaluation of the complex integral in Eq. (66)

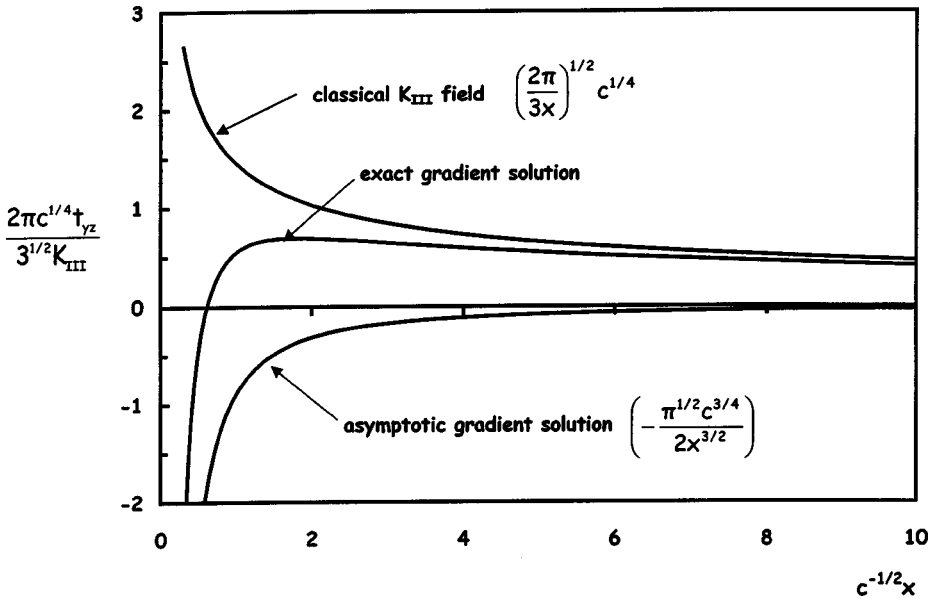


Fig. 8 Graphs of the exact gradient (total monopolar stress), asymptotic gradient (total monopolar stress), and classical K_{III} field solutions in normalized forms

$$\frac{1}{2\pi i} I = -\frac{1}{2\pi i}\left\{ i\int_{\pi/2}^{\pi} R^{3/2}\exp\left(i\frac{3\varphi}{2} - ax + Rxe^{i\varphi}\right) d\varphi \right.$$
$$+ i\int_{-\infty}^{-a} N^{+}(p)\cdot(|a+p|)^{1/2} e^{px} dp$$
$$+ \int_{-a}^{0} \frac{[\mathrm{Re}\, N(p) - i\,\mathrm{Im}\, N(p)](|a+p|)^{1/2} e^{px}}{N^{-}(p)} dp$$
$$+ \int_{0}^{-a} \frac{[\mathrm{Re}\, N(p) + i\,\mathrm{Im}\, N(p)](|a+p|)^{1/2} e^{px}}{N^{-}(p)} dp$$
$$- i\int_{-a}^{-\infty} N^{+}(p)\cdot(|a+p|)^{1/2} e^{px} dp + i\int_{-\pi}^{-\pi/2} R^{3/2}$$
$$\left.\times\exp\left(i\frac{3\varphi}{2} - ax + Rxe^{i\varphi}\right) d\varphi \right\}, \tag{67}$$

where R is the radius of the two quarter-circular paths having a center at the point $p=-a$ (see Fig. 7) and the angle φ is defined by the relation $p+a=R\cdot\exp(i\varphi)$. Also, $R\to\infty$ in the left half-plane, and $\mathrm{Re}\, N(p)=2cp^2/3$ and $\mathrm{Im}\, N(p)=2(1-cp^2)(a^2-p^2)^{1/2}[3|p|]^{-1}$ for p real and $|p|\leqslant a$. Further, it can be shown

$$\frac{1}{2\pi i} I = \frac{1}{\pi}\left\{ \int_{0}^{a} \frac{[\mathrm{Im}\, N(p)](a-p)^{1/2} e^{-px}}{N^{+}(p)} dx \right.$$
$$- \int_{a}^{\infty} N^{-}(p)\cdot(p-a)^{1/2} e^{-px} dp + R^{3/2} e^{-ax}$$
$$\left.\times\int_{\pi/2}^{\pi} \exp(Rx\cdot\cos\varphi)\cdot\cos\left(\frac{3\varphi}{2} + Rx\cdot\sin\varphi\right) d\varphi \right\}. \tag{68}$$

The third integral inside the braces vanishes as $R\to\infty$ and it is interesting to note that although the conditions for Jordan's lemma are not met by the integrand in (66), the contribution of the quarter-circular paths is zero because of the existence of the branch cut for the function $(a+p)^{1/2}$. Therefore, the total monopolar stress ahead of the crack tip is found from the following expression involving two real integrals:

$$t_{yz}(x, y=0) = \frac{K_{III}(3c)^{1/2}}{2\pi}\left\{ \int_{0}^{a} \frac{[\mathrm{Im}\, N(p)](a-p)^{1/2} e^{-px}}{N^{+}(p)} dx \right.$$
$$\left. - \int_{a}^{\infty} N^{-}(p)\cdot(p-a)^{1/2} e^{-px} dp \right\}. \tag{69}$$

It can be checked that both integrals are convergent. Also, a numerical evaluation of these integrals can easily be accomplished. Finally, the above expression can be written in a more convenient dimensionless form as

$$t_{yz}(x, y=0)$$
$$= \frac{K_{III} 3^{1/2}}{2\pi c^{1/4}}\left\{ \int_{0}^{1} \frac{[\mathrm{Im}\, N(p)](1-p)^{1/2}\exp(-c^{1/2}xp)}{N^{+}(p)} dp \right.$$
$$\left. - \int_{1}^{\infty} N^{-}(p)\cdot(p-1)^{1/2}\exp(-c^{1/2}xp) dp \right\}, \tag{70}$$

where

$$N^{\pm}(p) = \exp\left\{ \frac{1}{\pi}\int_{0}^{1} \arctan\left[\frac{(1-\zeta^2)^{3/2}}{\zeta^3}\right]\frac{1}{\zeta \pm p} d\zeta \right\}, \tag{71}$$

$$\mathrm{Im}\, N(p) = \frac{2(1-p^2)^{3/2}}{3p} \quad \text{for} \quad 0\leqslant p\leqslant 1. \tag{72}$$

The graph of the exact gradient expression for the total monopolar stress ahead of the crack tip in the normalized form $(2\pi c^{1/4} t_{yz}/3^{1/2} K_{III})$ versus $c^{-1/2}x$ is given in Fig. 8. In the same figure the normalized graphs of the asymptotic gradient solution $(-\pi^{1/2}c^{3/4}/2x^{3/2})$ and the classical K_{III} field solution $(2\pi/3x)^{1/2}c^{1/4}$ versus $c^{-1/2}x$ are also shown. The latter two graphs are provided for the purpose of comparison with the exact gradient stress distribution. Also, Fig. 9 presents the variation of the exact stress, in the normalized form $(2\pi t_{yz}/(3c)^{1/2} K_{III})$ with (x/h), where $2h$ is the size of the unit cell of the structured material (intrinsic material length—see Section 2). The two

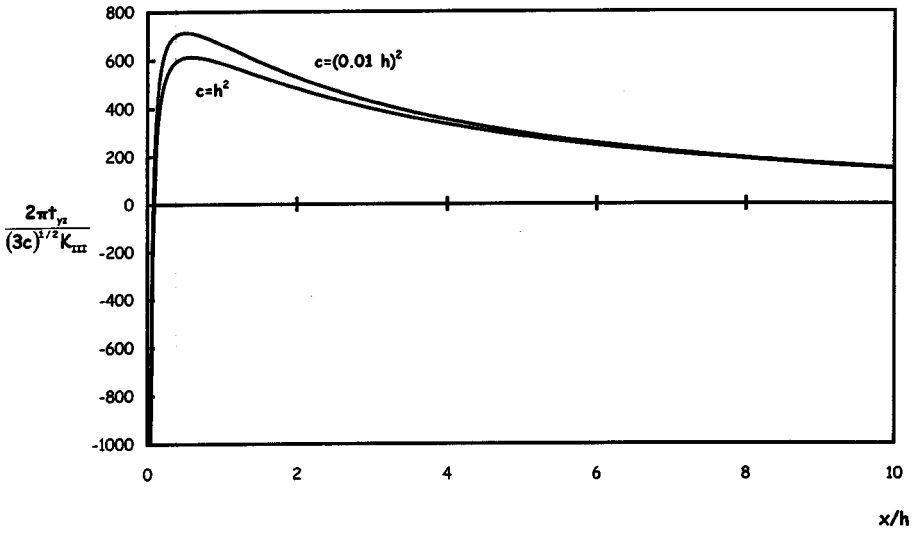


Fig. 9 Variation of the exact total monopolar stress (according to the gradient theory) with (x/h) for the cases $c=h^2$ and $c=(0.01h)^2$. The graphs depict that the cohesive zone is small as compared to the intrinsic material length h and that the stress ahead of the cohesive zone exhibits a bounded maximum.

graphs of Fig. 9 were obtained for the relations $c=(0.01h)^2$ and $c=h^2$. As mentioned in the Introduction, the study by Georgiadis et al. [22] gives the estimate $c=(0.1h)^2$. Thus, in the latter case the stress graph will be in between the two graphs of Fig. 9. The purpose of presenting these two graphs is to make apparent the *bounds* of the region ahead of the tip at which the stress takes on negative values for possible relations between the gradient coefficient c and the intrinsic length h.

On Fig. 8 now, an immediate observation is that the asymptotic gradient solution is inaccurate except for the region very near to the crack tip. Another observation is that the exact gradient stress-field tends to the classical K_{III} stress field at points lying outside the domain where the effects of microstructure are pronounced, i.e., for $x \gg c^{1/2}$. However, in the near-tip region where the distance from the crack-tip is comparable to the length $c^{1/2}$, the two fields differ radically indicating therefore that material microstructure is a significant factor in the fracture behavior of solids. The behavior of the exact solution depicted in Fig. 8 reminds somehow typical *boundary layer* behavior as, e.g., that found for the surface pressure near the leading edge of a Joukowski airfoil (Van Dyke [43]). In particular, the following remarks deserve more attention. For $x<0.5c^{1/2}$, the stress $t_{yz}(x>0,y=0)$ takes on negative values exhibiting therefore a *cohesive-traction* character along the prospective fracture zone (see, e.g., [51,52] for the nature of fracture cohesive zones). However, in view of the relation between c and h, the length L_c (cohesive-zone length of the order of $0.5c^{1/2}$) along which $t_{yz}<0$ is *extremely* small. For instance, even if h is rather large, say $h=2\times10^{-4}$ m (case of a geomaterial—see [13]), for $c=(0.1h)^2$ we have $L_c=0.05h=10^{-5}$ m. The same conclusion can also be reached by observing the graphs of Fig. 9 which show that L_c is a very small fraction of h. It is also interesting to note that L_c does not vary appreciably although c varies over a wide range, i.e., from $c=(0.01h)^2$ to $c=h^2$. Therefore, the length L_c can be considered practically equal to zero and be ignored. Accordingly, the domain of dominance of the $x^{-3/2}$-singularity being of extremely small size can be considered of no physical importance. Instead, one may attribute physical importance to the solution outside the cohesive zone, where the stress exhibits a maximum that is *bounded*. This maximum may serve as a measure of the critical stress level at which further advancement of the crack may occur. In other words, this result of the present gradient formulation of the crack problem permits a simple statement of the fracture criterion. Of course, the classical fracture mechanics analysis does not possess this feature since the stress maximum is unbounded at the crack-tip position $x=0$ and the stress drops monotonically for $x>0$ with no any *local* maximum. Finally, outside the cohesive zone, the stress $t_{yz}(x>L_c,y=0)$ predicted by the gradient theory is lower than that predicted by the classical elasticity theory.

8 Dynamical Time-Harmonic Mode III Crack Problem

We consider again the semi-infinite crack configuration of Section 4 but now assume a dynamical antiplane shear state. The transient problem leads to an extremely difficult mathematical initial/boundary value problem. Here, as a first step we deal with the *time-harmonic* inertial crack problem which, to our knowledge, consists the first attempt to analyze a dynamical crack problem within gradient elasticity. The more general transient solution may follow from the present one through Fourier synthesis. It is also expected that the basic *spatial* behavior of the solution (e.g., the order of singularities and the near-tip behavior) will be retained in the transient case as well. Within classical elasticity, problems involving cracks under remotely applied time-harmonic loading have been considered by, among others, Cherepanov [53] and Freund [54].

The cracked body is subjected to a remotely applied time-harmonic loading and the crack faces are traction-free. In view of the general expressions given in Section 2, Eqs. (12)–(14) remain the same but (11) and (15)–(17) are replaced by

$$u_x=u_y=0, \tag{73a}$$

$$u_z\equiv w\neq 0, \tag{73b}$$

$$w\equiv w(x,y,t)=w(x,y)\cdot\exp(i\Omega t), \tag{73c}$$

$$\frac{\partial}{\partial x}\left(\tau_{xz}-\frac{\partial m_{xxz}}{\partial x}-\frac{\partial m_{yxz}}{\partial y}\right)+\frac{\partial}{\partial y}\left(\tau_{yz}-\frac{\partial m_{xyz}}{\partial x}-\frac{\partial m_{yyz}}{\partial y}\right)$$
$$=\rho\frac{\partial^2 w}{\partial t^2}-\frac{\rho h^2}{3}\nabla^2\left(\frac{\partial^2 w}{\partial t^2}\right), \tag{74}$$

$$c\nabla^4 w-g\nabla^2 w-k^2 w=0, \tag{75}$$

$$\sigma_{xz}=\mu g\frac{\partial w}{\partial x}-\mu c\nabla^2\left(\frac{\partial w}{\partial x}\right), \tag{76a}$$

$$\sigma_{yz}=\mu g\frac{\partial w}{\partial y}-\mu c\nabla^2\left(\frac{\partial w}{\partial y}\right), \tag{76b}$$

where Ω is the frequency of the time-harmonic state, $g=(1-\Omega^2(\rho h^2/3\mu))$, and $k=(\Omega/V)$ with $V=(\mu/\rho)^{1/2}$ being the shear-wave velocity in the absence of gradient effects (i.e., in classical elasticity). Equation (75) is the field equation of the problem. It is called metaharmonic and appears also in the problem of bending vibrations of thin plates (Vekua [55]). More details about it can be found in [13,20]. In what follows, as is standard in time-harmonic problems, it is understood that all field quantities are to be multiplied by the factor $\exp(i\Omega t)$ and that the real part of the resulting expression is to be taken.

The above equations are also supplied by the boundary conditions (18)–(21). The resulting boundary value problem is attacked again by the Wiener-Hopf method. First, transforming (75) with (30*a*) gives the ordinary differential equation

$$c\frac{d^4w^*}{dy^4}+(2cp^2-g)\frac{d^2w^*}{dy^2}+(cp^4-gp^2-k^2)w^*=0, \tag{77}$$

with the following general solution (bounded as $y\to+\infty$)

$$w^*(p,y)=B(p)\cdot\exp(-\overline{\beta}y)+C(p)\cdot\exp(-\overline{\gamma}y)\quad\text{for}\quad y\geqslant 0, \tag{78}$$

where

$$\overline{\beta}\equiv\overline{\beta}(p)=i(p^2+\sigma^2)^{1/2} \tag{79a}$$

with

$$\sigma=\frac{[(g^2+4ck^2)^{1/2}-g]^{1/2}}{(2c)^{1/2}}>0, \tag{79b}$$

$$\overline{\gamma}\equiv\overline{\gamma}(p)=(\tau^2-p^2)^{1/2} \tag{80a}$$

with

$$\tau=\frac{[(g^2+4ck^2)^{1/2}+g]^{1/2}}{(2c)^{1/2}}>0. \tag{80b}$$

In the above equations, $B(p)$ and $C(p)$ are unknown functions, and the complex p-plane should be cut in the way shown in Fig. 10. Finally, the Laplace-transformed stresses that enter the boundary conditions are found to be

$$\sigma^*_{yz}(p,y)=-\mu c(\tau^2B\overline{\beta}e^{-\overline{\beta}y}-\sigma^2C\overline{\gamma}e^{-\overline{\gamma}y}), \tag{81}$$

$$m^*_{yyz}(p,y)=\mu c(B\overline{\beta}^2e^{-\overline{\beta}y}+C\overline{\gamma}^2e^{-\overline{\gamma}y}), \tag{82}$$

$$m^*_{yxz}(p,y)=-\mu cp(B\overline{\beta}e^{-\overline{\beta}y}+C\overline{\gamma}e^{-\overline{\gamma}y}). \tag{83}$$

Next, to formulate the Wiener-Hopf equation, the same "half-line" transforms are defined as in (36) and (37). Also, (38) and (39) apply in the present case too. The usual procedure of eliminating the functions (B,C) in the system of equations resulting from the transformed boundary conditions leads then to the following Wiener-Hopf equation

$$T^+(p)=\frac{\mu c}{\chi^2}\overline{\beta}\overline{\gamma}(\overline{\beta}^3-\overline{\gamma}^3)\cdot W^-(p), \tag{84}$$

where $\chi^2=(g^2+4ck^2)^{1/2}/c$ is a positive real constant dependent upon the material properties and the frequency. Notice also that $\chi^2=(\sigma^2+\tau^2)=\overline{\gamma}^2-\overline{\beta}^2$.

Further, since a product factorization of the function $\overline{\gamma}$ is immediately accomplished by inspection as $\overline{\gamma}(p)=(\tau+p)^{1/2}(\tau-p)^{1/2}$, Eq. (84) takes the form

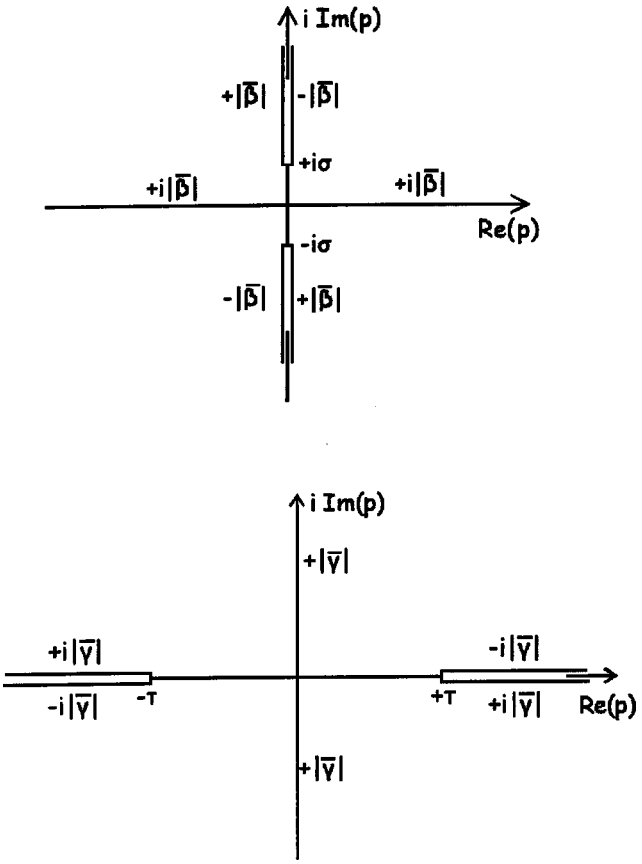


Fig. 10 Branch cuts for the functions $(\overline{\beta},\overline{\gamma})$

$$\frac{T^+(p)}{(\tau+p)^{1/2}}=-\frac{\mu c}{\chi^2}(\tau-p)^{1/2}(\sigma^2+p^2)\cdot\overline{L}(p)\cdot W^-(p), \tag{85}$$

where the kernel function $\overline{L}(p)$ is given as

$$\overline{L}(p)=(\sigma^2+p^2)+\frac{(\tau^2-p^2)^{3/2}}{i(\sigma^2+p^2)^{1/2}}. \tag{86}$$

Now, contrary to the static case analyzed in Section 6, the kernel function in the present dynamic case exhibits two zeros in the complex plane. This was found through a rather involved procedure using the principle of the argument, [37], and taking care of the behavior and the branch cuts of the functions $(\overline{\beta},\overline{\gamma})$. In addition, a check was made by the symbolic program MATHEMATICA™. Thus, the function $\overline{L}(p)$ exhibits the (non-extraneous) zeros

$$\pm Z=\pm\left\{\frac{g}{2c}+i\left[\frac{(g^2/4c^2)+(k^2/c)}{3}\right]^{1/2}\right\}^{1/2}, \tag{87}$$

and, in addition, has the asymptotic behavior $\lim_{|p|\to\infty}\overline{L}(p)\to 3\chi^2/2$. Next, the function $M(p)$ is introduced as

$$M(p)=\frac{2}{3\chi^2}\frac{(\tau^2-p^2)\cdot\overline{L}(p)}{p^2-Z^2}, \tag{88}$$

which no longer exhibits zeros and also has the desired asymptotic property $\lim_{|p|\to\infty}M(p)\to 1$. This new form of the kernel permits its product factorization through Cauchy's integral theorem.

In view of the above, the Wiener-Hopf equation of the problem becomes

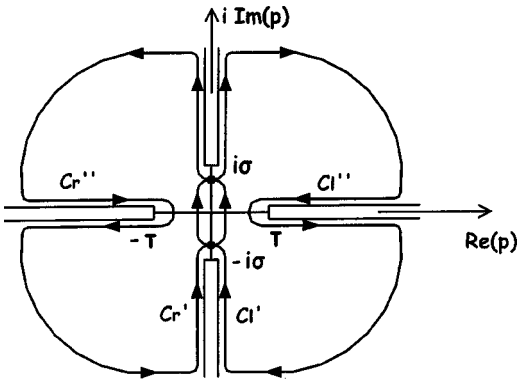


Fig. 11 Contour integrations for the factorization of the kernel function defined in Eq. (88)

$$\frac{T^+(p)\cdot(\tau+p)^{1/2}}{(p+Z)} = -\frac{3\mu c}{2}\frac{(\sigma^2+p^2)(p-Z)}{(\tau-p)^{1/2}}M(p)\cdot W^-(p), \tag{89}$$

and the kernel is written as the following product of two analytic and nonzero functions defined in pertinent half-plane domains of the complex plane

$$M(p)=M^+(p)\cdot M^-(p), \tag{90}$$

where

$$M^+(p)=\exp\left\{-\frac{1}{2\pi i}\int_{C_l'+C_l''}\frac{\ln[M(\zeta)]}{\zeta-p}d\zeta\right\}, \tag{91a}$$

$$M^-(p)=\exp\left\{\frac{1}{2\pi i}\int_{C_r'+C_r''}\frac{\ln[M(\zeta)]}{\zeta-p}d\zeta\right\}. \tag{91b}$$

The use of Cauchy's integral theorem to accomplish (90) is depicted in Fig. 11. Notice that Cauchy's theorem still applies in this case of a *nonsimple* contour (a contour with self-intersections) because the number of intersections is finite (see for the general result in, e.g., Ablowitz and Fokas [56]). $M^+(p)$ is analytic and nonzero in $\mathrm{Re}(p)>0$ and $M^-(p)$ is analytic and nonzero in $\mathrm{Re}(p)<0$. The integration path $(C_l'+C_l'')$ begins from the point Σ at $(-i\sigma+i\varepsilon)$, with ε real such $\varepsilon\to+0$, and runs along the entire imaginary axis (along the two cuts, it runs parallel to them on the right) and around the cut along the positive real axis. The integration path $(C_r'+C_r'')$ begins from the point Σ, it runs along the entire imaginary axis (along the two cuts, it runs parallel to them on the left) and around the cut along the negative real axis. Both integration paths end at the point Σ, and the second path is considered a continuation of the first so that Cauchy's theorem is applied and (90) is obtained. In both cases, the quarter-circular paths at infinity have a zero contribution according to Jordan's lemma. Finally, the small semi-circular paths around the branch points have a zero contribution.

Then, with the formal product factorization in hand, Eq. (89) is written under the following form that defines a function $\bar{E}(p)$:

$$\frac{T^+(p)\cdot(\tau+p)^{1/2}}{(p+Z)\cdot M^+(p)} = -\frac{3\mu c}{2}\frac{(\sigma^2+p^2)(p-Z)}{(\tau-p)^{1/2}}M^-(p)\cdot W^-(p) \equiv \bar{E}(p). \tag{92}$$

The above equation strictly holds along the segment of the imaginary axis $(\mathrm{Re}(p)=0, -\sigma<\mathrm{Im}(p)<\sigma)$ and $\bar{E}(p)$ is therefore defined only along this segment. This restriction of the validity of the Wiener-Hopf equation to a *finite* segment only (notice that in the static case treated before and invariably in crack problems within classical elasticity—both static and dynamic—the Wiener-Hopf equation holds along an infinite line or strip) is another novel feature of the present mathematical problem. Still, the theorem of analytic continuation applies and leads us to conclude that $\bar{E}(p)$ is an entire function. Working also along the same lines as those in the respective analysis of the previous static case, we find that the near-tip stress and displacement fields behave as in (51). Results analogous to the ones in the static case can be further obtained from the basic analysis of this section.

9 Conclusions

The present work was concerned with the exact determination of mode III crack-tip fields in a microstructured body under a remotely applied loading. The material microstructure was modeled according to the Mindlin-Green-Rivlin theory of generalized elastic continua (dipolar gradient or strain-gradient theory of grade two). A simple but yet rigorous version of this theory was employed by considering an isotropic linear expression of the elastic strain-energy density in antiplane shearing that involves only two material constants (the shear modulus and the gradient coefficient). The formulation of the problem and the solution methods were exact. The boundary value problem was attacked by the Wiener-Hopf technique but the asymptotic Williams technique was also employed in a preliminary analysis. Both static and time-harmonic dynamic analyses were provided. A singular-perturbation character was exhibited within the gradient formulation and the concept of a boundary layer was employed.

The results for the near-tip field showed significant departure from the predictions of the classical fracture mechanics. In particular, it was found that cohesive stresses develop in the immediate vicinity of the crack tip and that, ahead of the small cohesive zone, the stress distribution exhibits a local maximum that is bounded. This maximum value may serve, therefore, as a measure of the critical stress level at which further advancement of the crack may occur. In addition, the crack-face displacement closes more smoothly, in the vicinity of the crack tip, as compared to the classical result. The new formulation of the crack problem required also new extended definitions for the J-integral and the energy release rate. The determination of these quantities was made possible through the use of the theory of generalized functions.

A final notice pertains to the possibility of generalizing the present analysis by considering a continuum theory of even higher order than that of dipolar gradient theory. The next step could be a tripolar theory. The dipolar theory involves doublets of forces (double forces) as ''internal'' forces. The tripolar theory will involve rather doublets of moments (triple forces). Besides the fact that the latter generalized forces possess a not so clear physical meaning, the increased complexity of such a theory does not hold much hope for treating practical problems.

Acknowledgments

The author is thankful to Prof. L.M. Brock (University of Kentucky) for discussions on aspects of the mathematical analysis contained in this work. Also, the author is thankful to Prof. I. Vardoulakis (NTU Athens) and N. Aravas (University of Thessaly) for discussions on generalized continuum theories. Financial support of this work under the ''Thales'' program of the NTUA is gratefully acknowledged.

References

[1] Mindlin, R. D., 1964, ''Micro-Structure in Linear Elasticity,'' Arch. Ration. Mech. Anal., **16**, pp. 51–78.
[2] Green, A. E., and Rivlin, R. S., 1964, ''Multipolar Continuum Mechanics,'' Arch. Ration. Mech. Anal., **17**, pp. 113–147.
[3] Green, A. E., 1965, ''Micro-Materials and Multipolar Continuum Mechanics,'' Int. J. Eng. Sci., **3**, pp. 533–537.
[4] Mindlin, R. D., and Eshel, N. N., 1968, ''On First Strain-Gradient Theories in Linear Elasticity,'' Int. J. Solids Struct., **4**, pp. 109–124.
[5] Weitsman, Y., 1966, ''Strain-Gradient Effects Around Cylindrical Inclusions and Cavities in a Field of Cylindrically Symmetric Tension,'' ASME J. Appl. Mech., **33**, pp. 57–67.

[6] Day, F. D., and Weitsman, Y., 1966, ''Strain-Gradient Effects in Microlayers,'' ASCE J. Eng. Mech., **92**, pp. 67–86.
[7] Cook, T. S., and Weitsman, Y., 1966, ''Strain-Gradient Effects around Spherical Inclusions and Cavities,'' Int. J. Solids Struct., **2**, pp. 393–406.
[8] Herrmann, G., and Achenbach, J. D., 1968, ''Applications of Theories of Generalized Cosserat Continua to the Dynamics of Composite Materials,'' *Mechanics of Generalized Continua*, E. Kroener, ed., Springer, Berlin, pp. 69–79.
[9] Achenbach, J. D., Sun, C. T., and Herrmann, G., 1968, ''On the Vibrations of a Laminated Body,'' ASME J. Appl. Mech., **35**, pp. 689–696.
[10] Vardoulakis, I., and Sulem, J., 1995, *Bifurcation Analysis in Geomechanics*, Blackie Academic and Professional (Chapman and Hall), London.
[11] Fleck, N. A., Muller, G. M., Ashby, M. F., and Hutchinson, J. W., 1994, ''Strain Gradient Plasticity: Theory and Experiment,'' Acta Metall. Mater., **42**, pp. 475–487.
[12] Lakes, L., 1995, ''Experimental Methods for Study of Cosserat Elastic Solids and Other Generalized Elastic Continua,'' *Continuum Models for Materials with Microstructure*, H.-B. Muhlhaus, ed., John Wiley and Sons, Chichester, pp. 1–25.
[13] Vardoulakis, I., and Georgiadis, H. G., 1997, ''SH Surface Waves in a Homogeneous Gradient Elastic Half-Space with Surface Energy,'' J. Elast., **47**, pp. 147–165.
[14] Wei, Y., and Hutchinson, J. W., 1997, ''Steady-State Crack Growth and Work of Fracture for Solids Characterized by Strain Gradient Plasticity,'' J. Mech. Phys. Solids, **45**, pp. 1253–1273.
[15] Begley, M. R., and Hutchinson, J. W., 1998, ''The Mechanics of Size-Dependent Indentation,'' J. Mech. Phys. Solids, **46**, pp. 2049–2068.
[16] Exadaktylos, G., and Vardoulakis, I., 1998, ''Surface Instability in Gradient Elasticity With Surface Energy,'' Int. J. Solids Struct., **35**, pp. 2251–2281.
[17] Huang, Y., Zhang, L., Guo, T. F., and Hwang, K. C., 1997, ''Near-Tip Fields for Cracks in Materials With Gradient Effects,'' *Proceedings of the IUTAM Symposium on Nonlinear Analysis of Fracture*, J. R. Willis, ed., Kluwer Academic Publishers, Dordrecht, pp. 231–243.
[18] Zhang, L., Huang, Y., Chen, J. Y., and Hwang, K. C., 1998, ''The Mode-III Full-Field Solution in Elastic Materials With Strain Gradient Effects,'' Int. J. Fract., **92**, pp. 325–348.
[19] Chen, J. Y., Huang, Y., and Ortiz, M., 1998, ''Fracture Analysis of Cellular Materials: A Strain Gradient Model,'' J. Mech. Phys. Solids, **46**, pp. 789–828.
[20] Georgiadis, H. G., and Vardoulakis, I., 1998, ''Anti-Plane Shear Lamb's Problem Treated by Gradient Elasticity With Surface Energy,'' Wave Motion, **28**, pp. 353–366.
[21] Georgiadis, H. G., Vardoulakis, I., and Lykotrafitis, G., 2000, ''Torsional Surface Waves in a Gradient-Elastic Half-Space,'' Wave Motion, **31**, pp. 333–348.
[22] Georgiadis, H. G., Vardoulakis, I., and Velgaki, E. G., 2002, ''Dispersive Rayleigh-Wave Propagation in Microstructured Solids Characterized by Dipolar Gradient Elasticity,'' J. Elast., submitted for publication.
[23] Georgiadis, H. G., and Velgaki, E. G., 2002, ''High-Frequency Rayleigh Waves in Materials With Microstructure and Couple-Stress Effects,'' Int. J. Solids Struct., **40**, pp. 2501–2520.
[24] Amanatidou, E., and Aravas, N., 2001, ''Finite Element Techniques for Gradient Elasticity Problems,'' *Proceedings of the 6th Greek National Congress of Mechanics*, Hellenic Society of Theoretical and Applied Mechanics, **2**, pp. 149–154.
[25] Gazis, D. C., Herman, R., and Wallis, R. F., 1960, ''Surface Elastic Waves in Cubic Crystals,'' Phys. Rev., **119**, pp. 533–544.
[26] Jaunzemis, W., 1967, *Continuum Mechanics*, MacMillan, New York.
[27] Fung, Y. C., 1965, *Foundations of Solid Mechanics*, Prentice-Hall, Englewood Cliffs, NJ.
[28] Ignaczak, J., 1971, ''Tensorial Equations of Motion for Elastic Materials With Microstructure,'' *Trends in Elasticity and Thermoelasticity* (W. Nowacki Anniversary Volume), Wolters-Noordhoff, Groningen, pp. 89–111.
[29] Rice, J. R., 1968, ''A Path Independent Integral and the Approximate Analysis of Strain Concentration by Notches and Cracks,'' ASME J. Appl. Mech., **35**, pp. 379–386.
[30] Rice, J. R., 1968, ''Mathematical Analysis in the Mechanics of Fracture,'' *Fracture*, H. Liebowitz, ed., **2**, Academic Press, New York, pp. 191–311.
[31] Atkinson, C., and Leppington, F. G., 1974, ''Some Calculations of the Energy-Release Rate G for Cracks in Micropolar and Couple-Stress Elastic Media,'' Int. J. Fract., **10**, pp. 599–602.
[32] Lubarda, V. A., and Markenscoff, X., 2000, ''Conservation Integrals in Couple Stress Elasticity,'' J. Mech. Phys. Solids, **48**, pp. 553–564.
[33] Williams, M. L., 1952, ''Stress Singularities Resulting from Various Boundary Conditions in Angular Corners of Plates in Extension,'' ASME J. Appl. Mech., **74**, pp. 526–528.
[34] Williams, M. L., 1957, ''On the Stress Distribution at the Base of a Stationary Crack,'' ASME J. Appl. Mech., **79**, pp. 109–114.
[35] Barber, J. R., 1992, *Elasticity*, Kluwer Academic Publishers, Dordrecht.
[36] van der Pol, B., and Bremmer, H., 1950, *Operational Calculus Based on the Two-Sided Laplace Integral*, Cambridge University Press, Cambridge, UK.
[37] Carrier, G. A., Krook, M., and Pearson, C. E., 1966, *Functions of a Complex Variable*, McGraw-Hill, New York.
[38] Roos, B. W., 1969, *Analytic Functions and Distributions in Physics and Engineering*, John Wiley and Sons, New York.
[39] Mittra, R., and Lee, S. W., 1971, *Analytical Techniques in the Theory of Guided Waves*, MacMillan, New York.
[40] Gel'fand, I. M., and Shilov, G. E., 1964, *Generalized Functions*, **1**, Academic Press, New York.
[41] Lauwerier, H. A., 1963, ''The Hilbert Problem for Generalized Functions,'' Arch. Ration. Mech. Anal., **13**, pp. 157–166.
[42] Knowles, J. K., and Pucik, T. A., 1973, ''Uniqueness for Plane Crack Problems in Linear Elastostatics,'' J. Elast., **3**, pp. 155–160.
[43] Van Dyke, M., 1964, *Perturbation Methods in Fluid Mechanics*, Academic Press, New York.
[44] Rice, J. R., 1974, ''Limitations to the Small Scale Yielding Approximation for Crack Tip Plasticity,'' J. Mech. Phys. Solids, **22**, pp. 17–26.
[45] Mills, N. J., 1974, ''Dugdale Yielded Zones in Cracked Sheets of Glassy Polymers,'' Eng. Fract. Mech., **6**, pp. 537–549.
[46] Elssner, G., Korn, D., and Ruhle, M., 1994, ''The Influence of Interface Impurities on Fracture Energy of UHV Diffusion Bonded Metal-Ceramic Bicrystals,'' Scr. Metall. Mater., **31**, pp. 1037–1042.
[47] Bogy, D. B., and Sternberg, E., 1967, ''The Effect of Couple-Stresses on Singularities due to Discontinuous Loadings,'' Int. J. Solids Struct., **3**, pp. 757–770.
[48] Prakash, V., Freund, L. B., and Clifton, R. J., 1992, ''Stress Wave Radiation From a Crack Tip During Dynamic Initiation,'' ASME J. Appl. Mech., **59**, pp. 356–365.
[49] Fisher, B., 1971, ''The Product of Distributions,'' Quart. J. Math. Oxford, **22**, pp. 291–298.
[50] Bueckner, H. F., 1958, ''The Propagation of Cracks and the Energy of Elastic Deformation,'' Trans. ASME, **24**, pp. 1225–1230.
[51] Barenblatt, G. I., 1962, ''The Mathematical Theory of Equilibrium Cracks in Brittle Fracture,'' Adv. Appl. Mech., **7**, pp. 55–129.
[52] Nilsson, F., and Stahle, P., 1988, ''Crack Growth Criteria and Crack Tip Models,'' SM Arch., **13**(4), pp. 193–238.
[53] Cherepanov, G. P., 1979, *Mechanics of Brittle Fracture*, McGraw-Hill, New York.
[54] Freund, L. B., 1990, *Dynamic Fracture Mechanics*, Cambridge University Press, Cambridge, UK.
[55] Vekua, I. N., 1968, *New Methods for Solving Elliptic Equations*, North-Holland, Amsterdam.
[56] Ablowitz, M. J., and Fokas, A. S., 1997, *Complex Variables: Introduction and Applications*, Cambridge University Press, Cambridge, UK.

Gradient Elasticity Theory for Mode III Fracture in Functionally Graded Materials—Part I: Crack Perpendicular to the Material Gradation

G. H. Paulino
Department of Civil and Environmental Engineering, University of Illinois, Newmark Laboratory, 205 North Mathews Avenue, Urbana, IL 61801
Mem. ASME

A. C. Fannjiang
Department of Mathematics, University of California, Davis, CA 95616

Y.-S. Chan
Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

Anisotropic strain gradient elasticity theory is applied to the solution of a mode III crack in a functionally graded material. The theory possesses two material characteristic lengths, ℓ and ℓ', which describe the size scale effect resulting from the underlining microstructure, and are associated to volumetric and surface strain energy, respectively. The governing differential equation of the problem is derived assuming that the shear modulus is a function of the Cartesian coordinate y, i.e., $G=G(y)=G_0e^{\gamma y}$, where G_0 and γ are material constants. The crack boundary value problem is solved by means of Fourier transforms and the hypersingular integrodifferential equation method. The integral equation is discretized using the collocation method and a Chebyshev polynomial expansion. Formulas for stress intensity factors, K_{III}, are derived, and numerical results of K_{III} for various combinations of ℓ, ℓ', and γ are provided. Finally, conclusions are inferred and potential extensions of this work are discussed. [DOI: 10.1115/1.1532321]

1 Introduction

Classical (local) continuum theories possess no intrinsic length scale. Typical dimensions of length are generally associated with the overall geometry of the domain under consideration. Thus classical elasticity and plasticity are scale-free continuum theories in which there is no microstructure associated with material points, [1]. In contrast, strain gradient theories enrich the classical continuum with additional material characteristic lengths in order to describe the size (or scale) effects resulting from the underlining microstructures. Recent work on strain gradient theories to account for size (or scale) effects in materials can be found in the articles by Wu [2], Fleck and Hutchinson [3], Lakes [4,5], Smyshlyaev and Fleck [6], and Van Vliet and Van Mier [7]. Recent applications of gradient elasticity to fracture mechanics include the work by Fannjiang et al. [8], Paulino et al. [9], Exadaktylos et al. [10], Vardoulakis et al. [11], Aifantis [12], Zhang et al. [13], Hwang et al. [14], and the review paper by Hutchinson and Evans [15]. The present work focuses on anisotropic strain gradient elasticity theory for fracture problems in functionally graded materials (FGMs). To the best of the authors' knowledge, this is the first (or one of the first) solutions for FGMs with gradient terms.

The emergence of FGMs is the outcome of the need to accommodate material exposure to nonuniform service requirements. These multiphased materials feature gradual transition in composition and/or microstructure for the specific purpose of controlling variations in thermal, structural, or functional properties. The spatial variation of microstructure is accomplished through nonuniform distribution of the reinforcement phase with different properties, sizes, and shapes, as well as by interchanging the roles of reinforcement and matrix (base) materials in a continuous manner. This concept is illustrated by Fig. 1, which shows an FGM with a continuously graded microstructure. Typical examples of FGMs include ceramic/ceramic (e.g., $MoSi_2$/SiC [16] and TiC/SiC [17]), and metal/ceramic (e.g., Nb/Nb_5Si_3 [18] and Ti/TiB [19]), systems. Comprehensive reviews on several aspects of FGMs can be found in the articles by Markworth et al. [20], Erdogan [21], and Hirai [22], and in the book by Suresh and Mortensen [23].

This paper presents a linkage between gradient elasticity and graded materials within the framework of fracture mechanics. The remainder of the paper is organized as follows. First, the constitutive equations of anisotropic gradient elasticity for nonhomogeneous materials subjected to antiplane shear deformation are given. Then, the governing partial differential equations (PDEs)

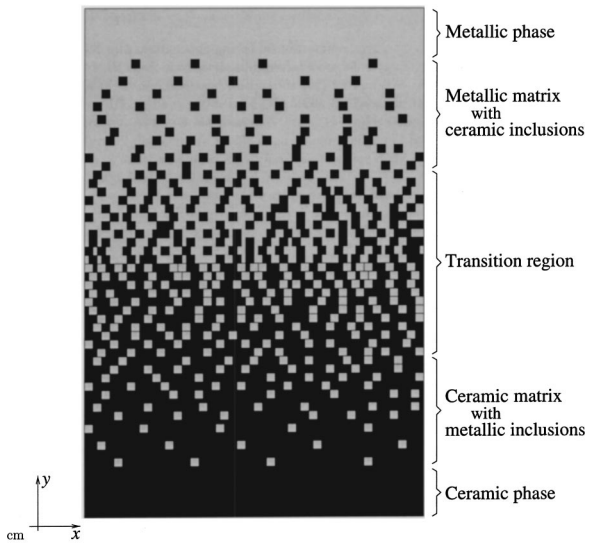


Fig. 1 Functionally graded material (FGM) with continuously graded microstructure

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Oct. 18, 2000; final revision, Sept. 6, 2001. Associate Editor: B. M. Moran. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

are derived and the Fourier transform method is introduced and applied to convert the governing PDE into an ordinary differential equation (ODE). Afterwards, the crack boundary value problem is described and a specific complete set of boundary conditions is given. The governing hypersingular integrodifferential equation is derived and discretized using the collocation method. Next, various relevant aspects of the numerical discretization are described in detail. Subsequently, numerical results are given, conclusions are inferred, and potential extensions of this work are discussed. Two appendices supplement the paper. One contains the lengthy expression of the regular kernel in the final (governing) hypersingular integrodifferential equation, and the other provides some useful formulas for evaluating hypersingular integrals and computing stress intensity factors (SIFs).

2 Constitutive Equations of Gradient Elasticity

This section introduces the notation and constitutive equations of gradient elasticity, which will be used to investigate antiplane shear cracks in functionally graded materials (FGMs). In three-dimensional space, the displacement components are defined as

$$u_x \equiv u, \quad u_y \equiv v, \quad u_z \equiv w, \tag{1}$$

and for antiplane shear problems, the following relations hold:

$$u = v = 0, \quad w = w(x,y). \tag{2}$$

Strains are defined as

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \tag{3}$$

where both the indices i and j run through $(x_1, x_2, x_3) = (x,y,z)$. For antiplane shear problems, the nontrivial strains are

$$\epsilon_{xz} = \frac{1}{2}\frac{\partial w}{\partial x}, \quad \epsilon_{yz} = \frac{1}{2}\frac{\partial w}{\partial y}. \tag{4}$$

Casal [24–26] has established the connection between surface tension effects and anisotropic gradient elasticity theory. For a material graded in the y-direction, the Casal's continuum can be extended so that the strain-energy density has the following form

$$\mathcal{W} = \tfrac{1}{2}\lambda(y)\epsilon_{ii}\epsilon_{jj} + G(y)\epsilon_{ij}\epsilon_{ji} + G(y)\ell^2(\partial_k \epsilon_{ij})(\partial_k \epsilon_{ji}) + \ell' \nu_k \partial_k [G(y)\epsilon_{ij}\epsilon_{ji}], \quad \ell > 0, \tag{5}$$

which has been generalized for an FGM with Lamé moduli $\lambda \equiv \lambda(y)$ and $G \equiv G(y)$. Moreover, $\partial_k = \partial/\partial x_k$. When the formulation is derived by means of a variational principle (or principle of virtual work), terms associated with ℓ undertake a volume integral, and terms associated with ℓ' can be reduced to a surface integral using the divergence theorem. In this sense, the characteristic length ℓ is responsible for volumetric strain-gradient terms, and the characteristic ℓ' is responsible for surface strain-gradient terms. Moreover, ν_k, $\partial_k \nu_k = 0$, is a director field equal to the unit outer normal n_k on the boundaries.

The Cauchy stresses τ_{ij}, the couple stresses μ_{kij} and the total stresses σ_{ij} are defined as

$$\tau_{ij} = \partial \mathcal{W}/\partial \epsilon_{ij} \tag{6}$$

$$\mu_{kij} = \partial \mathcal{W}/\partial \epsilon_{ij,k} \tag{7}$$

$$\sigma_{ij} = \tau_{ij} - \partial_k \mu_{kij}. \tag{8}$$

For homogeneous materials (i.e., λ and G constants), the stress fields are expressed in terms of strains and strain derivatives as

$$\sigma_{ij} = \lambda \epsilon_{kk}\delta_{ij} + 2G(\epsilon_{ij} - \ell^2 \nabla^2 \epsilon_{ij}) \tag{9}$$

$$\tau_{ij} = \lambda \epsilon_{kk}\delta_{ij} + 2G\epsilon_{ij} + 2G\ell' \nu_k \partial_k \epsilon_{ij} \tag{10}$$

$$\mu_{kij} = 2G(\ell' \nu_k \epsilon_{ij} + \ell^2 \partial_k \epsilon_{ij}). \tag{11}$$

As pointed out by Chan et al. [27], the constitutive equations of gradient elasticity for FGMs have a different form from the ones above. Thus, for FGMs with material gradation along the Cartesian coordinate y, the constitutive equations of gradient elasticity are

$$\sigma_{ij} = \lambda(y)\epsilon_{kk}\delta_{ij} + 2G(y)(\epsilon_{ij} - \ell^2 \nabla^2 \epsilon_{ij}) - 2\ell^2 [\partial_k G(y)](\partial_k \epsilon_{ij}) \tag{12}$$

$$\tau_{ij} = \lambda(y)\epsilon_{kk}\delta_{ij} + 2G(y)\epsilon_{ij} + 2\ell' \nu_k [\epsilon_{ij}\partial_k G(y) + G(y)\partial_k \epsilon_{ij}] \tag{13}$$

$$\mu_{kij} = 2\ell' \nu_k G(y)\epsilon_{ij} + 2\ell^2 G(y)\partial_k \epsilon_{ij}. \tag{14}$$

Note that the Cauchy stresses τ_{ij} are influenced by a term containing the spatial derivative of the shear modulus, and so are the total stresses σ_{ij}. The term "$-2\ell^2[\partial_k G(y)](\partial_k \epsilon_{ij})$" that appear in (12), but not in (9), can be interpreted as the interaction between the material gradation and the nonlocal strain gradient effect, which will play a role in the governing partial differential equation (PDE) (17) discussed in the next section. Moreover, if λ and G are constants, the constitutive equations for homogeneous materials (see Vardoulakis et al. [11], Exadaktylos et al. [10], and Fannjiang et al. [8]) are recovered as a particular case of Eqs. (12)–(14). If the shear modulus G is a function of y (see Fig. 2) and a mode III problem is under consideration, then each component of the stress field can be written as, [27]:

$$\begin{aligned} &\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = 0, \quad \sigma_{xy} = 0 \\ &\sigma_{xz} = 2G(y)(\epsilon_{xz} - \ell^2 \nabla^2 \epsilon_{xz}) - 2\ell^2 [\partial_y G(y)](\partial_y \epsilon_{xz}) \neq 0 \\ &\sigma_{yz} = 2G(y)(\epsilon_{yz} - \ell^2 \nabla^2 \epsilon_{yz}) - 2\ell^2 [\partial_y G(y)](\partial_y \epsilon_{yz}) \neq 0 \\ &\mu_{xxz} = 2G(y)\ell^2 \partial_x \epsilon_{xz} \\ &\mu_{xyz} = 2G(y)\ell^2 \partial_x \epsilon_{yz} \end{aligned} \tag{15}$$

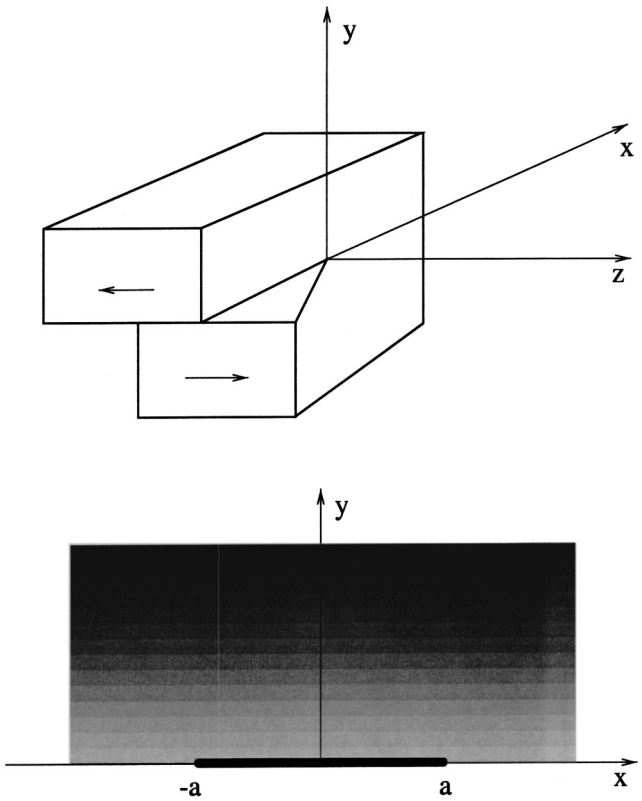


Fig. 2 Mode III crack in a functionally graded material

$$\mu_{yxz}=2G(y)(\ell^2\partial_y\epsilon_{xz}-\ell'\epsilon_{xz})$$
$$\mu_{yyz}=2G(y)(\ell^2\partial_y\epsilon_{yz}-\ell'\epsilon_{yz}).$$

Again, it is worth pointing out that there is an extra term in σ_{xz} and σ_{yz} as compared to the homogeneous material case (see Vardoulakis et al. [11] p. 4534).

3 Governing Partial Differential Equation

By imposing the only nontrivial equilibrium equation

$$\frac{\partial\sigma_{xz}}{\partial x}+\frac{\partial\sigma_{yz}}{\partial y}=0, \tag{16}$$

the following partial differential equation (PDE) for general form of $G(y)$ is obtained:

$$\frac{\partial}{\partial x}\left[G(y)\left(\frac{\partial w}{\partial x}-\ell^2\nabla^2\frac{\partial w}{\partial x}\right)\right]+\frac{\partial}{\partial y}\left[G(y)\left(\frac{\partial w}{\partial y}-\ell^2\nabla^2\frac{\partial w}{\partial y}\right)\right]$$
$$-\ell^2\left[\frac{\partial^2G(y)}{\partial y^2}\frac{\partial^2w}{\partial y^2}+\frac{\partial G(y)}{\partial y}\frac{\partial^3w}{\partial y^3}+\frac{\partial G(y)}{\partial y}\frac{\partial^3w}{\partial x^2\partial y}\right]=0. \tag{17}$$

If the shear modulus G is an exponential function of y, i.e.,

$$G\equiv G(y)=G_0e^{\gamma y}, \tag{18}$$

then (17) can be simplified as

$$-\ell^2\nabla^4w-2\gamma\ell^2\nabla^2\frac{\partial w}{\partial y}+\nabla^2w-\gamma^2\ell^2\frac{\partial^2w}{\partial y^2}+\gamma\frac{\partial w}{\partial y}=0, \tag{19}$$

or in a factored form

$$\left(1-\gamma\ell^2\frac{\partial}{\partial y}-\ell^2\nabla^2\right)\left(\nabla^2+\gamma\frac{\partial}{\partial y}\right)w=0. \tag{20}$$

In terms of the differential operator notation, (20) can be written in the form as

$$\mathsf{H}_\gamma\mathsf{L}_\gamma w=0;\quad \mathsf{H}_\gamma=1-\gamma\ell^2\frac{\partial}{\partial y}-\ell^2\nabla^2,\quad \mathsf{L}_\gamma=\nabla^2+\gamma\frac{\partial}{\partial y}, \tag{21}$$

where H_γ is the perturbed Helmholtz operator, L_γ is the perturbed Laplacian operator, and the two operators commute, i.e., $\mathsf{H}_\gamma\mathsf{L}_\gamma=\mathsf{L}_\gamma\mathsf{H}_\gamma$. Thus, the PDE (20) can be considered as a double perturbation of the composition of the Helmholtz and harmonic equations,

$$(1-\ell^2\nabla^2)\nabla^2w=0, \tag{22}$$

that is, one perturbation is to the Helmholtz operator $(1-\ell^2\nabla^2)$, and the other perturbation is to the Laplacian operator ∇^2. Both the Helmholtz and the Laplacian operators are invariant under "rigid-body motions." However, FGMs bring in the perturbation and destroy such invariance. By setting $\gamma\to0$ in (20), one gets (22), which is the PDE for gradient elasticity.

Another viewpoint of the perturbation is focused on the role of the characteristic length ℓ. By taking $\ell\to0$ (at the level of the differential equation), we obtain a lower order of PDE,

$$\left(\nabla^2+\gamma\frac{\partial}{\partial y}\right)w=0,$$

i.e., the perturbed harmonic equation, which has been investigated by Erdogan and Ozturk [28]. However, because the corresponding term to the coefficient ℓ^2 affects the highest differential in the governing PDE (19), a singular perturbation is expected as the limit $\ell\to0$ is considered. By taking both $\gamma\to0$ and $\ell\to0$, we obtain the harmonic equation for classical elasticity. Various combination of parameters ℓ and γ with the corresponding governing PDE are listed in Table 1.

Table 1 Governing partial differential equations (PDEs) in antiplane shear problems

Cases	Governing PDE	References
$\ell=0,\ \gamma=0$	Laplace equation: $\nabla^2w=0$	Standard textbooks.
$\ell=0,\ \gamma\neq0$	Perturbed Laplace equation: $\left(\nabla^2+\gamma\frac{\partial}{\partial y}\right)w=0$	Erdogan and Ozturk [28].
$\ell\neq0,\ \gamma=0$	Helmholtz-Laplace equation: $(1-\ell^2\nabla^2)\nabla^2w=0$	Vardoulakis *et al.* [11]. Fannjiang *et al.* [8].
$\ell\neq0,\ \gamma\neq0$	Equation (20): $\left(1-\gamma\ell^2\frac{\partial}{\partial y}-\ell^2\nabla^2\right)\left(\nabla^2+\gamma\frac{\partial}{\partial y}\right)w=0$	Studied in this paper.

4 Fourier Transform

Let the Fourier transform be defined by

$$\mathcal{F}(w)(\xi)=W(\xi)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}w(x)e^{ix\xi}dx. \tag{23}$$

The inverse Fourier transform theorem gives

$$\mathcal{F}^{-1}(W)(x)=w(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}W(\xi)e^{-ix\xi}d\xi, \tag{24}$$

where $i=\sqrt{-1}$. Now let us assume that

$$w(x,y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}W(\xi,y)e^{-ix\xi}d\xi, \tag{25}$$

i.e., $w(x,y)$ is the inverse Fourier transform of the function $W(\xi,y)$.

Considering each term in Eq. (17) term by term, and using Eq. (25), one obtains

$$-\ell^2\nabla^4w=-\ell^2\left(\frac{\partial^4w(x,y)}{\partial x^4}+2\frac{\partial^4w(x,y)}{\partial x^2\partial y^2}+\frac{\partial^4w(x,y)}{\partial y^4}\right)$$
$$=\frac{-\ell^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(\xi^4W(\xi,y)-2\xi^2\frac{\partial^2W}{\partial y^2}+\frac{\partial^4W}{\partial y^4}\right)e^{-ix\xi}\ d\xi \tag{26}$$

$$-2\gamma\ell^2\nabla^2\frac{\partial w}{\partial y}=-2\gamma\ell^2\left(\frac{\partial^3w(x,y)}{\partial x^2\partial y}+\frac{\partial^3w(x,y)}{\partial y^3}\right)$$
$$=-2\frac{\gamma\ell^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(-\xi^2\frac{\partial W(\xi,y)}{\partial y}+\frac{\partial^3W}{\partial y^3}\right)e^{-ix\xi}\ d\xi \tag{27}$$

$$\nabla^2w=\frac{\partial^2w(x,y)}{\partial x^2}+\frac{\partial^2w(x,y)}{\partial y^2}$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(-\xi^2W(\xi,y)+\frac{\partial^2W}{\partial y^2}\right)e^{-ix\xi}\ d\xi \tag{28}$$

$$-\gamma^2\ell^2\frac{\partial^2w(x,y)}{\partial y^2}=-\frac{\gamma^2\ell^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\partial^2W(\xi,y)}{\partial y^2}e^{-ix\xi}\ d\xi \tag{29}$$

$$\gamma\frac{\partial w(x,y)}{\partial y}=\frac{\gamma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\partial W(\xi,y)}{\partial y}e^{-ix\xi}\ d\xi. \tag{30}$$

Equations (26) to (30) are added (according to Eq. (19)), and after simplification, the governing ordinary differential equation (ODE) is obtained:

$$\left[\ell^2\frac{d^4}{dy^4}+2\gamma\ell^2\frac{d^3}{dy^3}-(2\ell^2\xi^2+\gamma^2\ell^2+1)\frac{d^2}{dy^2}-\gamma(1+2\ell^2\xi^2)\frac{d}{dy}\right.$$
$$\left.+(\ell^2\xi^4+\xi^2)\right]W=0. \tag{31}$$

Table 2 Roots λ_i together with corresponding mechanics theory and type of material

Cases	Number of roots	Roots	Mechanics theory and type of material	References
$\ell=0,\ \gamma=0$	2	$\pm\lvert\xi\rvert$	Classical LEFM, homogeneous materials	Standard textbooks.
$\ell=0,\ \gamma\neq 0$	2	$-\gamma/2\pm\sqrt{\gamma^2/4+\xi^2}$	Classical LEFM, nonhomogeneous materials	Erdogan and Ozturk [28].
$\ell\neq 0,\ \gamma=0$	4	$\pm\lvert\xi\rvert,\ \pm\sqrt{\xi^2+1/\ell^2}$	Gradient theories, homogeneous materials	Vardoulakis *et al.* [11]. Fannjiang *et al.* [8].
$\ell\neq 0,\ \gamma\neq 0$	4	$-\gamma/2\pm\sqrt{\gamma^2/4+\xi^2}$, $-\gamma/2\pm\sqrt{\xi^2+\gamma^2/4+1/\ell^2}$	Gradient theories, nonhomogeneous materials	Studied in this paper.

5 Solutions of the Ordinary Differential Equation

The corresponding characteristic equation to the ordinary differential equation (ODE) (31) is

$$\ell^2\lambda^4+2\gamma\ell^2\lambda^3-(2\ell^2\xi^2+\gamma^2\ell^2+1)\lambda^2-\gamma(1+2\ell^2\xi^2)\lambda+(\ell^2\xi^4+\xi^2)=0, \tag{32}$$

which can be further factored as

$$[\ell^2\lambda^2+\gamma\ell^2\lambda-(1+\ell^2\xi^2)](\lambda^2+\gamma\lambda-\xi^2)=0. \tag{33}$$

Clearly the four roots λ_i $(i=1,2,3,4)$ of the polynomial (33) above can be obtained as

$$\lambda_1=\frac{-\gamma}{2}-\frac{\sqrt{\gamma^2+4\xi^2}}{2},\quad \lambda_2=\frac{-\gamma}{2}+\frac{\sqrt{\gamma^2+4\xi^2}}{2}, \tag{34}$$

$$\lambda_3=\frac{-\gamma}{2}-\sqrt{\xi^2+\gamma^2/4+1/\ell^2},\quad \lambda_4=\frac{-\gamma}{2}+\sqrt{\xi^2+\gamma^2/4+1/\ell^2}, \tag{35}$$

where we let $\lambda_1<0$ and $\lambda_3<0$. As $\gamma\to 0$, we recover the roots found by Vardoulakis et al. [11] and Fannjiang et al. [8]. The roots λ_1 and λ_2 correspond to the solution of the perturbed harmonic equation, and the roots λ_3 and λ_4 match with the solution of the perturbed Helmholtz's equation. Various choices of parameters ℓ and γ with their corresponding mechanics theories and material types are listed in Table 2.

By taking account of the far-field boundary condition

$$w(x,y)\to 0 \ \text{ as } \ \sqrt{x^2+y^2}\to+\infty, \tag{36}$$

and with $y>0$ (the upper half plane), one obtains

$$W(\xi,y)=A(\xi)e^{\lambda_1 y}+B(\xi)e^{\lambda_3 y}. \tag{37}$$

Accordingly, the displacement $w(x,y)$ takes the form

$$w(x,y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}[A(\xi)e^{\lambda_1 y}+B(\xi)e^{\lambda_3 y}]e^{-ix\xi}d\xi. \tag{38}$$

Both $A(\xi)$ and $B(\xi)$ are determined by the boundary conditions.

6 Boundary Conditions

Figure 2 shows the geometry of the mode III crack problem in which a functionally graded material (FGM), with shear modulus $G(y)=G_0e^{\gamma y}$, bonded to a half-space is considered. Thus the problem reduces to the upper half-plane, and $y=0$ is treated as the boundary. By the principle of virtual work, the following mixed boundary conditions can be derived:

$$\begin{cases}\sigma_{yz}(x,0)=p(x), & |x|<a\\ w(x,0)=0, & |x|>a\\ \mu_{yyz}(x,0)=0, & -\infty<x<+\infty,\end{cases} \tag{39}$$

which are adopted in this paper. One may observe that the first two boundary conditions (BCs) in (39) are from classical elasticity, e.g., linear elastic fracture mechanics (LEFM). The last BC regarding the couple-stress μ_{yyz} is needed as the higher order theory is considered.

7 Hypersingular Integrodifferential Equation Approach

By taking account of the symmetry along the x-axis, we may consider that $w(x,y)$ takes the following general solution form (for the upper half-plane):

$$\begin{aligned}w(x,y)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}[A(\xi)e^{\lambda_1 y}+B(\xi)e^{\lambda_3 y}]e^{-ix\xi}d\xi,\quad y\geqslant 0\\ &=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}[A(\xi)e^{-(\gamma+\sqrt{4\xi^2+\gamma^2})y/2}\\ &\quad+B(\xi)e^{-(\gamma+\sqrt{4\xi^2+\gamma^2+4/\ell^2})y/2}]e^{-ix\xi}d\xi,\quad y\geqslant 0,\end{aligned} \tag{40}$$

where $A(\xi)$ and $B(\xi)$ need to be determined from the boundary conditions (39). As Eq. (40) provides the form of the solution for $w(x,y)$, it can be used in conjunction with Eq. (15) such that

$$\begin{aligned}\sigma_{yz}(x,y)&=2G(y)(\epsilon_{yz}-\ell^2\nabla^2\epsilon_{yz})-2\ell^2[\partial_y G(y)](\partial_y\epsilon_{yz})\\ &=\frac{G(y)}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\lambda_1(\gamma,\xi)A(\xi)e^{-(\gamma+\sqrt{\gamma^2+4\xi^2})y/2-ix\xi}d\xi,\\ &\quad y\geqslant 0.\end{aligned} \tag{41}$$

Notice that the term associated with $B(\xi)$ has been dropped out from $\sigma_{yz}(x,y)$. Moreover,

$$\begin{aligned}\mu_{yyz}(x,y)&=2G(y)\left(\ell^2\frac{\partial\epsilon_{yz}}{\partial y}-\ell'\epsilon_{yz}\right),\quad y\geqslant 0,\\ &=\frac{G(y)}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\{(\ell^2\lambda_1^2-\ell'\lambda_1)A(\xi)e^{\lambda_1 y}+(\ell^2\lambda_3^2\\ &\quad-\ell'\lambda_3)B(\xi)e^{\lambda_3 y}\}e^{-ix\xi}d\xi\\ &=\frac{G(y)}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\{c_A(\gamma,\xi)A(\xi)e^{-(\gamma+\sqrt{\gamma^2+4\xi^2})y/2}\\ &\quad+c_B(\gamma,\xi)B(\xi)e^{-(\gamma+\sqrt{4\xi^2+\gamma^2+4/\ell^2})y/2}\}e^{-ix\xi}d\xi,\end{aligned} \tag{42}$$

where

$$c_A(\gamma,\xi)=\ell^2\lambda_1^2-\ell'\lambda_1$$
$$=\frac{\gamma}{2}(\gamma\ell^2+\ell')+\frac{1}{2}(\gamma\ell^2+\ell')\sqrt{\gamma^2+4\xi^2}+\ell^2\xi^2, \tag{43}$$

and

$$c_B(\gamma,\xi)=\ell^2\lambda_3^2-\ell'\lambda_3$$
$$=\ell^2\xi^2+\frac{\gamma}{2}(\gamma\ell^2+\ell')+1$$
$$+\frac{1}{2}(\gamma\ell^2+\ell')\sqrt{4\xi^2+\gamma^2+4/\ell^2}. \tag{44}$$

In order to derive the Fredholm integral equation, we define the density as the slope function

$$\phi(x)=\partial w(x,0^+)/\partial x. \tag{45}$$

The second boundary condition in (39), and Eq. (45), imply that

$$\phi(x)=0, \quad |x|>a, \tag{46}$$

and

$$\int_{-a}^{a}\phi(x)dx=0, \tag{47}$$

which is the single-valuedness condition. The definition (45), together with Eq. (40), lead to

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(-i\xi)[A(\xi)+B(\xi)]e^{-ix\xi}d\xi=\phi(x), \quad -\infty<x<\infty. \tag{48}$$

By inverting the Fourier transform and using (46), one obtains

$$(i\xi)[A(\xi)+B(\xi)]=\frac{-1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\phi(x)e^{ix\xi}dx, \quad -\infty<x<\infty$$
$$=\frac{-1}{\sqrt{2\pi}}\int_{-a}^{a}\phi(t)e^{i\xi t}dt. \tag{49}$$

The last boundary condition in (39), imposed on $\mu_{yyz}(x,y)$, provides the following pointwise relationship between $A(\xi)$ and $B(\xi)$:

$$B(\xi)=-\frac{\ell^2\xi^2+(\gamma\ell^2+\ell')\sqrt{\gamma^2/4+\xi^2}+\gamma(\gamma\ell^2+\ell')/2}{\ell^2\xi^2+1+[(\gamma\ell^2+\ell')/2](\gamma+\sqrt{4\xi^2+\gamma^2+4/\ell^2})}A(\xi)$$
$$=\rho(\gamma,\xi)A(\xi), \tag{50}$$

where the notation $\rho(\gamma,\xi)$ is introduced here, i.e.,

$$\rho(\gamma,\xi)=-\frac{\ell^2\xi^2+(\gamma\ell^2+\ell')\sqrt{\gamma^2/4+\xi^2}+\gamma(\gamma\ell^2+\ell')/2}{\ell^2\xi^2+1+[(\gamma\ell^2+\ell')/2](\gamma+\sqrt{4\xi^2+\gamma^2+4/\ell^2})}. \tag{51}$$

Substituting (50) into (49), one obtains

$$A(\xi)=\frac{-1}{\sqrt{2\pi}i\xi}\left[\frac{1}{1+\rho(\gamma,\xi)}\right]\int_{-a}^{a}\phi(t)e^{i\xi t}dt, \tag{52}$$

where

$$\frac{1}{1+\rho(\gamma,\xi)}=\frac{\ell^2\xi^2+1+[(\gamma\ell^2+\ell')/2](\gamma+\sqrt{4\xi^2+\gamma^2+4/\ell^2})}{1+[(\gamma\ell^2+\ell')/2](\sqrt{4\xi^2+\gamma^2+4/\ell^2}-\sqrt{4\xi^2+\gamma^2})}. \tag{53}$$

Replacing $A(\xi)$ in Eq. (41) and using the (first) boundary condition for σ_{yz} (that is, $\lim_{y\to0^+}\sigma_{yz}(x,y)=p(x)$, $|x|<a$) in (39), one obtains the following integral equation in limit form:

$$\lim_{y\to0^+}\frac{G(y)}{2\pi}\int_{-\infty}^{\infty}\left[\frac{-\lambda_1(\gamma,\xi)}{i\xi(1+\rho(\gamma,\xi))}\right]$$
$$\times\left[\int_{-a}^{a}\phi(t)e^{i\xi t}dt\right]e^{-(\gamma+\sqrt{\gamma^2+4\xi^2})y/2-ix\xi}d\xi$$
$$=p(x), \quad |x|<a. \tag{54}$$

By rearranging the order of integration, we obtain

$$\lim_{y\to0^+}\frac{G(y)}{2\pi}\int_{-a}^{a}\phi(t)\int_{-\infty}^{\infty}\frac{-\lambda_1(\gamma,\xi)}{(i\xi)[1+\rho(\gamma,\xi)]}$$
$$\times e^{-(\gamma+\sqrt{\gamma^2+4\xi^2})y/2}e^{i\xi(t-x)}d\xi dt$$
$$=p(x), \quad |x|<a, \tag{55}$$

which can be rewritten as

$$\lim_{y\to0^+}\frac{G}{2\pi}\int_{-a}^{a}\phi(t)\int_{-\infty}^{\infty}K(\xi,y)e^{i\xi(t-x)}d\xi dt=p(x), \quad |x|<a, \tag{56}$$

with the kernel

$$K(\xi,y)=\frac{-\lambda_1(\gamma,\xi)}{i\xi[1+\rho(\gamma,\xi)]}e^{-(\gamma+\sqrt{\gamma^2+4\xi^2})y/2}. \tag{57}$$

Asymptotic analysis allows splitting of the kernel $K(\xi,y)$ into the singular $[K_\infty(\xi,y)=\lim_{|\xi|\to\infty}K(\xi,y)]$ and nonsingular parts:

$$K(\xi,y)=\underbrace{K_\infty(\xi,y)}_{\text{singular}}+\underbrace{[K(\xi,y)-K_\infty(\xi,y)]}_{\text{nonsingular}}, \tag{58}$$

where (as y is set to zero)

$$K_\infty(\xi,0)=\frac{|\xi|}{i\xi}\left\{\left[\frac{5\ell^2\gamma^2}{8}+\frac{\ell'\gamma}{4}+1-\left(\frac{\ell'}{2l}\right)^2\right]\right.$$
$$\left.+\frac{2\gamma\ell^2+\ell'}{2}|\xi|+\ell^2\xi^2\right\}, \tag{59}$$

and $K(\xi,0)-K_\infty(\xi,0)$, denoted by $N(\xi,0)=N(\xi)$, can be expressed as a fraction:

$$N(\xi,0)=N(\xi)=\frac{P(\xi)}{Q(\xi)}, \tag{60}$$

with $P(\xi)$ and $Q(\xi)$ described in Appendix A.

Substitution of Eq. (59) into (56), in the sense of distribution theory, [29], leads to

$$\lim_{y\to0^+}\int_{-\infty}^{\infty}K_\infty(\xi,y)e^{i\xi(t-x)}d\xi$$
$$=\frac{-2\ell^2}{(t-x)^3}-\frac{\pi}{2}(2\ell^2\gamma+\ell')\delta'(t-x)$$
$$+\frac{5\ell^2\gamma^2/8+\ell'\gamma/4+1-[\ell'/(2\ell)]^2}{t-x},$$

and to the following hypersingular integral equation:

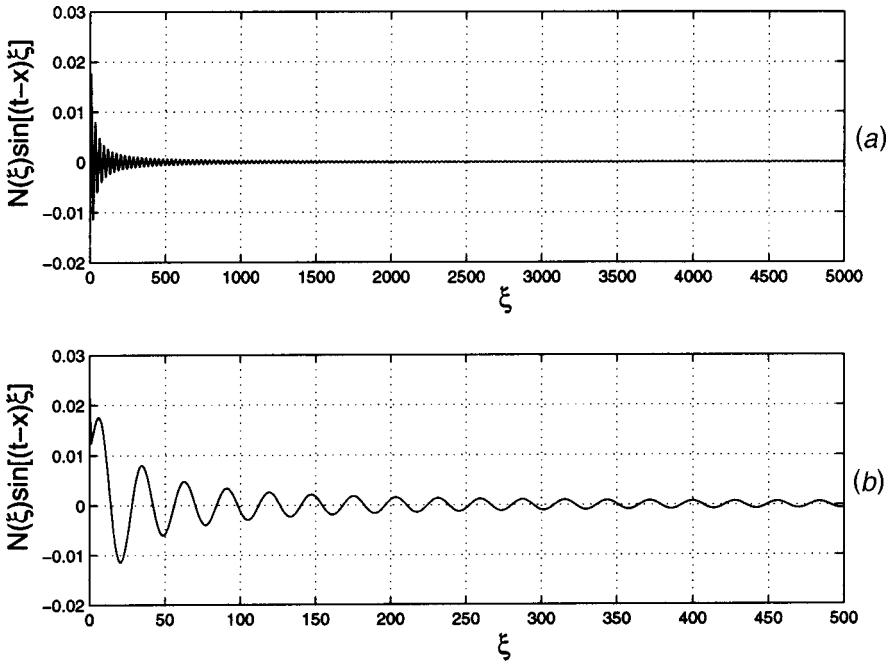


Fig. 3 Plot of the integrand in Eq. (62) for $\ell=0.05$, $\ell'=0.005$, $\gamma=0.1$, $r=\sqrt{3}/7$, and $s=\sqrt{2}/3$. (*a*) $\xi\in[0,5000]$; (*b*) Zoom for the range $\xi\in[0,500]$. Moreover, as $\xi\to 0$, the limit of $N(\xi)\sin[\xi(s-r)]$ is about 22.4×10^{-3}.

$$\frac{G_0}{\pi}\Xint=_{-a}^{a}\left\{\frac{-2\ell^2}{(t-x)^3}-\frac{\pi}{2}(2\ell^2\gamma+\ell')\delta'(t-x)+\frac{5\ell^2\gamma^2/8+\ell'\gamma/4+1-[\ell'/(2\ell)]^2}{t-x}+k(x,t)\right\}\phi(t)dt=p(x),\quad |x|<a, \tag{61}$$

where the regular kernel is

$$k(x,t)=\int_0^\infty N(\xi)\sin[\xi(t-x)]d\xi \tag{62}$$

with $N(\xi)$ described in Eq. (60). Figure 3 permits to graphically evaluate the behavior of the integrand of Eq. (62). Clearly, such kernel is oscillatory, but the magnitude of oscillation decreases and tend to zero as ξ increases, i.e., $\lim_{\xi\to\infty}N(\xi)\sin[\xi(t-x)]=0$. Another point that we need to be cautious about in Eq. (62) is the behavior at $\xi=0$ of $N(\xi)=P(\xi)/Q(\xi)$ as $Q(\xi)$ has the factor ξ in the denominator. However, this would not affect the integrability of the integrand in Eq. (62) because of the term $\sin[\xi(t-x)]$. Thus $\lim_{\xi\to 0}N(\xi)\sin[\xi(t-x)]$ exists and is finite, which depends on the values of t, x, ℓ, ℓ', and γ.

As a result of distribution theory, [29], the differentiation of a delta function, $\delta(t)$, has the following property:

$$\int_{-\infty}^{\infty}\delta'(t-x)\phi(t)dt=-\phi'(x). \tag{63}$$

Thus one may rewrite Eq. (61) as

$$\frac{G_0}{\pi}\Xint=_{-a}^{a}\left\{\frac{-2\ell^2}{(t-x)^3}+\frac{5\ell^2\gamma^2/8+\ell'\gamma/4+1-(\ell'/\ell)^2/4}{t-x}+k(x,t)\right\}\times\phi(t)dt+\frac{G}{2}(\ell'+2\ell^2\gamma)\phi'(x)=p(x),\quad |x|<a, \tag{64}$$

which is an integrodifferential equation with both hypersingular and Cauchy singular kernels. In addition to the single-valuedness condition condition in (47), the integrodifferential Eq. (64) is solved under the physical constraint ("smooth closure condition"):

$$\phi(a)=\phi(-a)=0, \tag{65}$$

so that the solution can be found uniquely (see Refs. [8] and [30]).

8 Numerical Solution

The numerical solution of the mode III fracture boundary value problem is accomplished by means of the collocation method, [31,32]. The process of obtaining the numerical solution of Eq. (64) can be divided into the following steps:

- Normalization,
- representation of the density function,
- Chebyshev polynomial expansion,
- evaluation of the derivative of the density function,
- formation of the linear system of equations,
- evaluation of singular and hypersingular integrals, and
- evaluation of nonsingular integral.

Relevant details for each of the above items are given below.

8.1 Normalization. By the following change of variables,

$$s=[2/(d-c)][t-(c+d)/2],$$

one may convert the integral $\int_c^d g(t)dt$ into the form of $\int_{-1}^{1}f(s)ds$. Because the crack surface is located in the range $(-a,a)$, a convenient change of variables becomes

$$t/a=s\quad\text{and}\quad x/a=r,$$

which is the normalization of the variables t and x, respectively. Thus Eq. (64) can be written in normalized fashion as

$$\frac{1}{\pi}\oint\!\!\!\!\!\!=\int_{-1}^{1}\left\{\frac{-2(\ell/a)^2}{(s-r)^3}+\frac{5(\ell/a)^2(a\gamma)^2/8+(\ell'/a)(a\gamma)/4+1-[(\ell'/a)/(\ell/a)]^2/4}{s-r}+\mathcal{K}(r,s)\right\}\Phi(s)ds+[\ell'/a+2(\ell/a)^2(a\gamma)]\Phi'(r)/2=\mathcal{P}(r)/G_0,\quad |r|<1, \tag{66}$$

where

$$\Phi(r)=\phi(ar),\quad \mathcal{P}(r)=p(ar),\quad \mathcal{K}(r,s)=ak(ar,as).$$

As clearly seen in Eq. (66), the quantities ℓ/a, ℓ'/a, and $a\gamma$ are dimensionless parameters. Thus the following dimensionless parameters are defined:

$$\tilde{\ell}=\ell/a,\quad \tilde{\ell}'=\ell'/a,\quad \tilde{\gamma}=a\gamma, \tag{67}$$

which will be used in the numerical implementation and results.

8.2 Representation of the Density Function. The next step of the numerical approach to the (normalized) hypersingular integral Eq. (66) is to establish the actual behavior of the unknown density function $\Phi(s)$ around the two crack tips $s=\pm 1$. For example, the governing integral equation in classical linear elastic fracture mechanics (LEFM) has Cauchy singularity if the slope function, say $\Phi(s)_{\mathrm{LEFM}}$, is chosen to be the unknown density function. A well-known representation is, [31,32],

$$\Phi(s)_{\mathrm{LEFM}}=f(s)/\sqrt{1-s^2},\quad |s|<1,$$

where $f(\pm 1)\neq 0$. For the cubic hypersingular integral, Eq. (66), the representation of $\Phi(s)$ is found to be, [8],

$$\Phi(s)_{\mathrm{GE}}\equiv\Phi(s)=g(s)\sqrt{1-s^2}, \tag{68}$$

where $g(\pm 1)$ is finite, $g(\pm 1)\neq 0$, and the subscript GE stands for gradient elasticity. Thus by approximating $g(s)$, one can find the numerical solution to $\Phi(s)$.

8.3 Chebyshev Polynomial Expansion. The approximation of $g(s)$ in Eq. (68) is accomplished by means of Chebyshev polynomial expansions. Either Chebyshev polynomials of the first kind $T_n(s)$, or of the second kind $U_n(s)$, may be employed in the approximation, i.e.,

$$g(s)=\sum_{n=0}^{\infty}a_nT_n(s)\quad\text{or}\quad g(s)=\sum_{n=0}^{\infty}A_nU_n(s). \tag{69}$$

The coefficients a_ns or A_ns are determined numerically by the collocation method. As shown by Chan et al. [33], the two expansions should lead to the same numerical results. In this paper, the expansion using $U_n(s)$ is adopted, i.e.,

$$\Phi(s)=\sqrt{1-s^2}\sum_{n=0}^{\infty}A_nU_n(s), \tag{70}$$

where $U_n(s)$ is defined, as usual, by

$$U_n(s)=\frac{\sin[(n+1)\cos^{-1}(s)]}{\sin[\cos^{-1}(s)]},\quad n=0,\,1,\,2,\,\ldots. \tag{71}$$

Satisfaction of the single-valuedness condition (47), or equivalently, $\int_{-1}^{1}\Phi(s)\,ds=0$, requires that the following relation holds:

$$A_0=0. \tag{72}$$

8.4 Evaluation of the Derivative of the Density Function. The term $\Phi'(r)$ in Eq. (66) is evaluated using the expansion (70) and the fact that

$$\frac{d}{dr}[U_n(r)\sqrt{1-r^2}]=-\frac{n+1}{\sqrt{1-r^2}}T_{n+1}(r),\quad n\geq 0. \tag{73}$$

Thus

$$\Phi'(r)=\frac{d}{dr}\left[\sqrt{1-r^2}\sum_{n=0}^{\infty}A_nU_n(r)\right]=\frac{-1}{\sqrt{1-r^2}}\sum_{n=0}^{\infty}(n+1)A_nT_n(r). \tag{74}$$

8.5 Formation of the Linear System of Equations. The strategy to determine the coefficients A_ns consists of forming a set of linear algebraic equations. Replacing $\Phi(s)$ in (66) by the representation (70), and using (74) one obtains the governing integral equation in discretized form:

$$-2\tilde{\ell}^2\sum_{n=1}^{\infty}\frac{A_n}{\pi}\oint\!\!\!\!\!\!=\int_{-1}^{1}\frac{U_n(s)\sqrt{1-s^2}}{(s-r)^3}ds+\left[1+\frac{5\tilde{\ell}^2\tilde{\gamma}^2}{8}+\frac{\tilde{\ell}'\tilde{\gamma}}{4}-\left(\frac{\tilde{\ell}'}{2\tilde{\ell}}\right)^2\right]\sum_{n=1}^{\infty}\frac{A_n}{\pi}\oint\!\!\!\!\!\!=\int_{-1}^{1}\frac{U_n(s)\sqrt{1-s^2}}{s-r}ds+\sum_{n=1}^{\infty}\frac{A_n}{\pi}\int_{-1}^{1}\sqrt{1-s^2}U_n(s)\mathcal{K}(r,s)ds-\frac{\tilde{\ell}'+2\tilde{\ell}^2\tilde{\gamma}}{2\sqrt{1-r^2}}\sum_{n=1}^{\infty}A_n(n+1)T_{n+1}(r)=\frac{\mathcal{P}(r)}{G_0},\quad |r|<1. \tag{75}$$

Notice that the running index n starts from 1 instead of 0 (see (72)).

8.6 Evaluation of Singular and Hypersingular Integrals. The governing integrodifferential Eq. (64), and its discretized version, Eq. (75), contain both Cauchy singular and hypersingular integrals (cubic singularity), which need to be evaluated. Erdogan et al. [31,32] have presented formulas for evaluating Cauchy singular integrals, and Chan et al. [34] have presented formulas for evaluating a broad class of hypersingular integrals, which generalizes previous derivations, [31,32,35], in the literature. Here, such integrals are interpreted in the finite-part sense, and listed in Appendix B (Eq. (93) to (95)).

8.7 Evaluation of Nonsingular Integral. Combining all the results obtained so far in the numerical approximation, one may rewrite Eq. (75) in the following form:

$$\frac{-\tilde{\ell}^2}{2(1-r^2)}\sum_{n=1}^{\infty}A_n[(n^2+n)U_{n+1}(r)-(2n^2+3n+2)U_{n-1}(r)]-\left[1+\frac{5\tilde{\ell}^2\tilde{\gamma}^2}{8}+\frac{\tilde{\ell}'\tilde{\gamma}}{4}-\left(\frac{\tilde{\ell}'}{2\tilde{\ell}}\right)^2\right]\sum_{n=1}^{\infty}A_nT_{n+1}(r)+\sum_{n=1}^{\infty}\frac{A_n}{\pi}\int_{-1}^{1}\sqrt{1-s^2}U_n(s)\mathcal{K}(r,s)ds-\frac{\tilde{\ell}'+2\tilde{\ell}^2\tilde{\gamma}}{2\sqrt{1-r^2}}\sum_{n=1}^{\infty}A_n(n+1)T_{n+1}(r)=\frac{\mathcal{P}(r)}{G},\quad |r|<1. \tag{76}$$

Thus the last step for applying the collocation method consists of evaluating the (regular) integral in (76), which is actually a double integral, i.e.,

$$\int_{-1}^{1}\sqrt{1-s^2}U_n(s)\mathcal{K}(r,s)ds$$

$$=\int_{-1}^{1}\sqrt{1-s^2}U_n(s)ak(ar,as)ds$$

$$=\int_{-1}^{1}\sqrt{1-s^2}U_n(s)\int_0^{\infty}aN(\xi)\sin[a\xi(s-r)]d\xi ds.$$

The integral along $[0, \infty)$ is a Fourier sine transform, and can be efficiently evaluated by applying fast Fourier transform (FFT) [36]. The integral along $[-1,1]$ can be readily obtained by the Gaussian quadrature method, [37].

9 Stress Intensity Factors (SIFs)

Since the (macroscopic) propagation of a crack starts around its tips, it is very important to study and determine the SIFs at both crack tips. In classical linear elastic fracture mechanics (LEFM), the stress $\sigma_{yz}(x,0)$ has $1/\sqrt{x-a}$ singularity as $x\to a^+$ (or $1/\sqrt{x+a}$, as $x\to -a^-$), and thus SIFs are defined and can be calculated by

$$K_{\mathrm{III}}(a)=\lim_{x\to a^+}\sqrt{2\pi(x-a)}\sigma_{yz}(x,0),\quad (x>a), \tag{77}$$

and

$$K_{\mathrm{III}}(-a)=\lim_{x\to -a^-}\sqrt{2\pi(-a-x)}\sigma_{yz}(x,0),\quad (x<-a). \tag{78}$$

However, the same definition may not hold for strain-gradient elasticity because $\sigma_{yz}(x,0)$ may have a stronger singularity, [13]. Thus SIFs will be redefined in the development below.

First, note that the limit in Eqs. (77) and (78) is taken from the region outside the crack surfaces toward both tips, and the integral Eq. (64) is the expression for $\sigma_{yz}(x,0)$ which is valid for $|x|>a$ as well as $|x|<a$, i.e.,

$$\sigma_{yz}(x,0)=\frac{G}{\pi}\int_{-a}^{a}\left\{\frac{-2\ell^2}{(t-x)^3}+\frac{5\ell^2\gamma^2/8+\ell'\gamma/4+1-(\ell'/\ell)^2/4}{t-x}+k(x,t)\right\}\phi(t)dt+\frac{G}{2}(\ell'+2\ell^2\gamma)\phi'(x),\quad |x|>a. \tag{79}$$

Second, after normalization and with the density function $\Phi(t)$ expanded by Chebyshev polynomials of the second kind U_n, some integral formulas, which are useful for deriving SIFs, need to be developed for $|r|>1$ (Chan et al. [34]), and are listed in Appendix B (see Eqs. (96) to (98)). Notice that the highest singularity in the Eqs. (96) to (98) appears in the last term in Eq. (98), and it has singularity $(r^2-1)^{-3/2}$ as $r\to 1^+$ or $r\to -1^-$. Motivated by such asymptotic behavior, we generalize the SIFs for strain gradient elasticity from those of classical LEFM. Thus

$$\ell K_{\mathrm{III}}(a)=\lim_{x\to a^+}2\sqrt{2\pi(x-a)}(x-a)\ \sigma_{yz}(x,0), \tag{80}$$

$$\ell K_{\mathrm{III}}(-a)=\lim_{x\to -a^-}2\sqrt{2\pi(x+a)}(x+a)\sigma_{yz}(x,0). \tag{81}$$

Therefore, the following formulas for the normalized mode III SIFs in the strain-gradient elasticity theory may be derived:

$$\ell K_{\mathrm{III}}(a)=\lim_{x\to a^+}2\sqrt{2\pi(x-a)}(x-a)\sigma_{yz}(x,0),\quad (x>a)$$

$$=\lim_{r\to 1^+}2\sqrt{2\pi(ar-a)}(ar-a)\sigma_{yz}(ar,0),\quad (r>1)$$

$$=2a\sqrt{\pi a}\ G_0\lim_{r\to 1^+}\sqrt{2(r-1)}(r-1)\frac{-2\ell^2}{\pi a^2}$$

$$\times\int_{-1}^{1}\frac{\Phi(s)}{(s-r)^3}ds,\quad (r>1). \tag{82}$$

After cancellation of the common terms, Eq. (82) can be continued by introducing formula (98), and using the representation (70), i.e.,

$$K_{\mathrm{III}}(a)=2\sqrt{2\pi a}\left(\frac{-2\ell}{a}\right)G_0\lim_{r\to 1^+}(r-1)^{3/2}\sum_{n=0}^{N}\frac{-(n+1)}{2}$$

$$\times\left(r-\frac{|r|}{r}\sqrt{r^2-1}\right)^{n-1}\left[n\left(1-\frac{|r|}{\sqrt{r^2-1}}\right)^2+\frac{r-\frac{|r|}{r}\sqrt{r^2-1}}{\sqrt{r^2-1^3}}\right]A_n$$

$$=\sqrt{\pi a}\ (\ell/a)G_0\sum_{n=0}^{\infty}(n+1)A_n. \tag{83}$$

Similarly,

$$K_{\mathrm{III}}(-a)=\sqrt{\pi a}\ (\ell/a)G_0\sum_{n=0}^{\infty}(-1)^n(n+1)A_n. \tag{84}$$

Formulas (83) and (84) will be used to obtain numerical results for SIFs.

10 Results and Discussion

The boundary value problem illustrated in Fig. 2 is considered for all the examples in this paper. To validate the present formulation, consider the case where $\ell,\ \ell'\to 0$ in a certain special limit sense (see Fannjiang et al. [8]), so that the classical elasticity solution is represented. The results for classical stress intensity factors (SIFs) (Eqs. (77) and (78)) are given in Table 3. It is clearly seen from Table 3 that the present results are in agreement with those of Erdogan and Ozturk [28]. Note that the SIFs decrease monotonically as γ increases. Moreover, it is interesting to investigate the asymptotic behavior of the SIFs as $\gamma\to\pm\infty$. As $\gamma\to\infty$ the stiffness of the medium increases indefinitely and, under finite loading (p_0), the crack-opening displacement and the SIFs $K_{\mathrm{III}}(a)$ tend to zero. Similarly, as $\gamma\to-\infty$ the stiffness of the

Table 3 Variation of classical (normalized) stress intensity factors (SIFs) with the material gradation parameter $\tilde{\gamma}=\gamma/a$

	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$	
$\tilde{\gamma}$	Present Study	Erdogan and Ozturk [28]
-2.0	1.476	1.481
-1.6	1.381	1.397
-1.2	1.293	1.308
-0.8	1.204	1.214
-0.4	1.117	1.113
-0.2	1.061	1.059
0.0	1.000	1.000
0.2	0.934	0.934
0.4	0.866	0.869
0.6	0.807	0.810
0.8	0.755	0.758
1.0	0.709	0.712
1.2	0.669	0.671
1.6	0.602	0.604
2.0	0.556	0.550
3.0	0.458	0.457
5.0	0.359	0.356
6.0	0.329	0.324

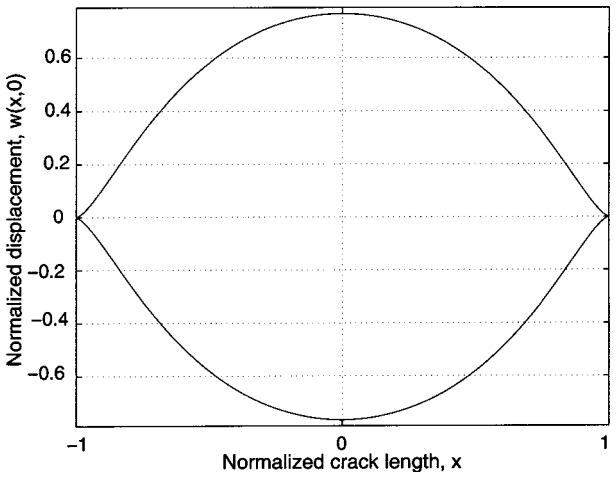


Fig. 4 Full crack displacement profile in an infinite medium of homogeneous material ($\tilde{\gamma}=0$) under uniform crack surface shear loading $\sigma_{yz}(x,0)=-p_0$ with choice of (normalized) $\tilde{\ell}=0.2$ and $\tilde{\ell}'=0$

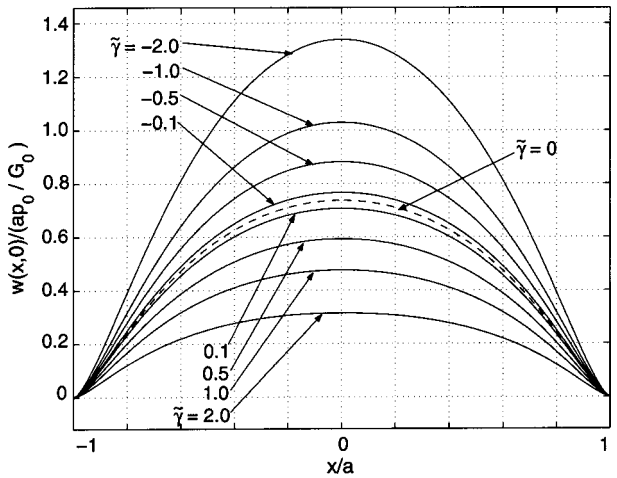


Fig. 6 Crack surface displacement under uniform crack surface shear loading $\sigma_{yz}(x,0)=-p_0$ and shear modulus $G(y)=G_0e^{\gamma y}$ with choice of (normalized) $\tilde{\ell}=0.2$, $\tilde{\ell}'=0.04$, and various $\tilde{\gamma}$. The dashed line stands for the homogeneous material ($\tilde{\gamma}=0$) in a gradient elastic medium.

medium decreases indefinitely, and consequently $K_{\mathrm{III}}(a)$ tend to infinity. These physically expected trends can be observed in Table 3.

Once the slope function is found numerically using the representation (68), the crack displacement profile $w(r,0)$ can be obtained as

$$w(r,0)=\int_{-1}^{r}\Phi(s)ds=\int_{-1}^{r}\sqrt{1-s^2}\sum_{n=0}^{N}A_nU_n(s)ds. \quad (85)$$

Figure 4 shows the normalized crack displacement profile in an infinite medium of homogeneous material ($\gamma=0$) under uniform crack surface loading for $\tilde{\ell}=0.2$ and $\tilde{\ell}'=0$. Notice that the crack tips form a cusp with zero enclosed angle and zero first derivative of the displacement at the crack tips (see (65)). This crack shape is similar to the one obtained by Barenblatt [38] using "cohesive zone theory," but without the assumption regarding existence of interatomic forces.

The solutions obtained in this study for a nonhomogeneous half-plane having shear modulus $G\equiv G(y)$, $y>0$, is also valid for the corresponding infinite medium in which $y=0$ is a plane of symmetry (see Fig. 2), i.e.,

$$G(-y)=G(y).$$

Unless otherwise stated, uniform loading is considered on the crack face, i.e., $\sigma_{yz}(x,0)=-p_0$, and the normalization p_0/G_0 has been employed.

Further normalized crack displacement profiles for various combinations of the gradient parameters $(\tilde{\ell},\tilde{\ell}')$ and material gradation parameter $(\tilde{\gamma})$ are presented in Fig. 5 to Fig. 8. Figures 5 and 6 show crack displacement profiles for selected values of $\tilde{\ell}$, $\tilde{\ell}'$, and various γ. Figure 5 considers $\tilde{\ell}=0.05$, $\tilde{\ell}'=0$ and thus $\rho=\ell'/\ell=0$; while Fig. 6 considers $\tilde{\ell}=0.20$, $\tilde{\ell}'=0.04$ and thus $\rho=\ell'/\ell=0.2$. In both graphs, the broken lines stand for the homogeneous material ($\gamma=0$) in a gradient elastic medium. A com-

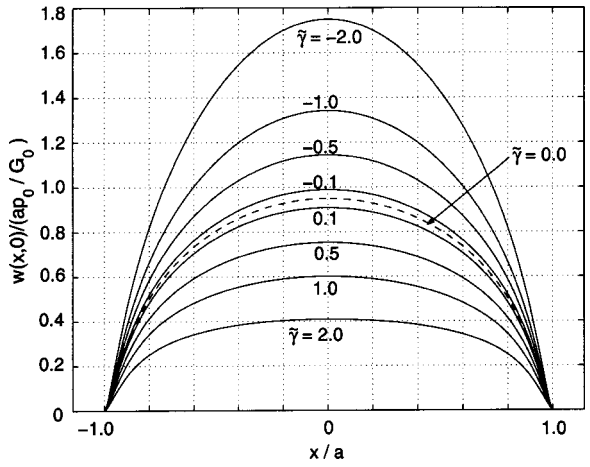


Fig. 5 Crack surface displacement under uniform crack surface shear loading $\sigma_{yz}(x,0)=-p_0$ and shear modulus $G(y)=G_0e^{\gamma y}$ with choice of (normalized) $\tilde{\ell}=0.05$, $\tilde{\ell}'=0$, and various $\tilde{\gamma}$. The dashed line stands for the homogeneous material case ($\tilde{\gamma}=0$).

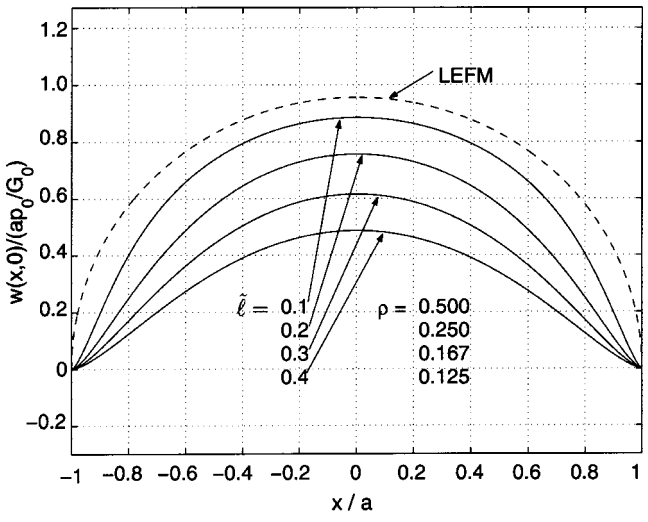


Fig. 7 Crack surface displacement profiles under uniform crack surface shear loading $\sigma_{yz}(x,0)=-p_0$ and shear modulus $G(y)=G_0e^{\gamma y}$ with choice of (normalized) $\tilde{\ell}'=0.05$, $\tilde{\gamma}=0.1$, and various $\tilde{\ell}$. The values of $\tilde{\ell}$ are listed in the same order as the solid-line curves.

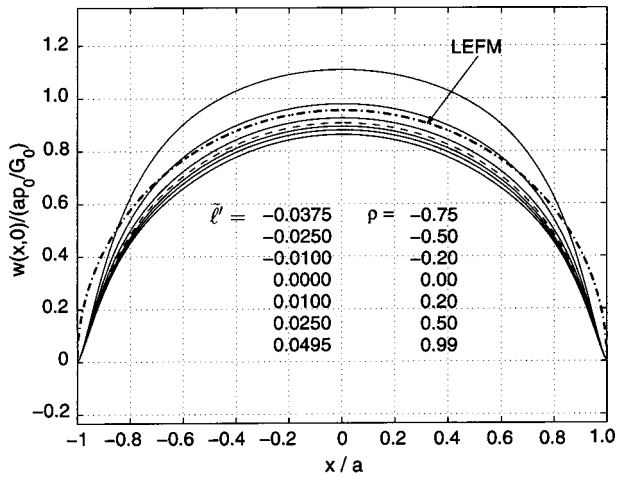


Fig. 8 Crack surface displacement profiles under uniform crack surface shear loading $\sigma_{yz}(x,0)=-p_0$ and shear modulus $G(y)=G_0e^{\gamma y}$ with choice of (normalized) $\tilde{\ell}=0.05$, $\tilde{\gamma}=0.1$, and various $\tilde{\ell}'$. The values of $\tilde{\ell}'$ (and $\rho=\ell/\ell'$) are listed in the same order as the solid-line and dashed-line ($\rho=0$) curves representing the strain gradient results.

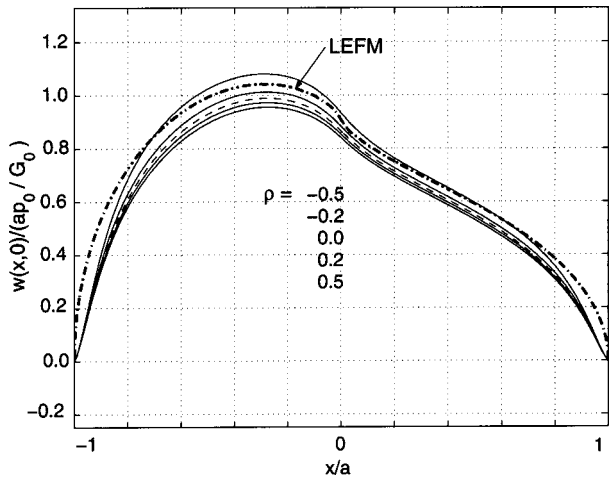


Fig. 9 Crack surface displacement profiles under discontinuous loading $p(x/a)=-1+0.5\ \mathrm{sgn}(x/a)$ and shear modulus $G(y)=G_0e^{\gamma y}$ with choice of (normalized) $\tilde{\ell}=0.05$, $\tilde{\gamma}=0.2$, and various $\rho=\ell/\ell'$. The values of ρ are listed in the same order as the solid-line and dashed-line ($\rho=0$) curves representing the strain gradient results.

parison between Figs. 5 and 6 permits to assess the influence of the gradient parameters (ℓ,ℓ') on the displacement solution. Moreover, as γ increases the displacement magnitude decreases, which is consistent with similar results by Erdogan and Ozturk [28] using classical elasticity to model mode III cracks in functionally graded materials (FGMs).

Figure 7 shows crack displacement profiles for $\tilde{\ell}'=0.05$, $\tilde{\gamma}=0.10$ and various $\tilde{\ell}$. As $\tilde{\ell}$ increases, the displacement diminishes monotonically, or alternatively the crack becomes stiffer, in comparison to the classical elasticity theory.

Figure 8 shows crack displacement profiles for $\tilde{\ell}=0.05$, $\tilde{\gamma}=0.10$ and various $\tilde{\ell}'$. As is apparent from this figure, by maintaining the values of the relative volume energy parameter $\tilde{\ell}$ constant, the crack stiffening effect becomes more pronounced as the relative surface energy parameter $\tilde{\ell}'$ increases in the range $[0,\tilde{\ell})$. It is worth mentioning that, from energy considerations, the parameter $\tilde{\ell}'$ can take negative values, [39]. Note from Fig. 8 that the effect of a negative $\tilde{\ell}'$ leads to a more compliant crack. In general, this is a desirable property of the mathematical model in regards to describing experimental results and data.

Table 4 Convergence of (normalized) generalized stress intensity factors (SIFs) for a mode III crack

	$\tilde{\gamma}=0,\ \tilde{\ell}=0.05$				$\tilde{\gamma}=0.30,\ \tilde{\ell}=0.05$			
	$\tilde{\ell}'=0;\ \rho=0$		$\tilde{\ell}'=0.01;\ \rho=0.20$		$\tilde{\ell}'=0;\ \rho=0$		$\tilde{\ell}'=0.01;\ \rho=0.20$	
N	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$	Cond. Num.	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$	Cond. Num.	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$	Cond. Num.	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$	Cond. Num.
11	0.97292	9.888	0.99640	17.018	0.89258	15.223	0.90773	15.142
21	0.97467	83.559	0.97375	1.669e+02	0.88381	1.509e+02	0.88337	1.478e+02
31	0.97467	3.555e+02	0.97355	7.131e+02	0.88376	6.437e+02	0.88287	6.314e+02
41	0.97467	1.032e+03	0.972256	2.059e+03	0.88336	1.859e+03	0.88133	1.823e+03
51	0.97467	2.395e+03	0.97109	4.754e+03	0.88301	4.293e+03	0.87999	4.206e+03
61	0.97467	4.802e+03	0.97113	9.501e+03	0.88301	8.577e+03	0.87996	8.406e+03

Table 5 Normalized generalized stress intensity factors (SIFs) for a mode III crack at various values of $\tilde{l}$, $\tilde{l}'$, and $\tilde{\gamma}$

	$\tilde{\ell}=0.05,\ \tilde{\ell}'=0$	$\tilde{\ell}=0.05,\ \tilde{\ell}'=0.01$	$\tilde{\ell}=0.2,\ \tilde{\ell}'=0$	$\tilde{\ell}=0.2,\ \tilde{\ell}'=0.04$
$\tilde{\gamma}$	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$	$\frac{K_{III}(-a)}{p_0\sqrt{\pi a}}$
-2.00	1.42126	1.41617	1.28917	1.26783
-1.00	1.21749	1.21301	1.10392	1.08610
-0.50	1.10374	1.09965	1.00377	0.98768
-0.10	1.00271	0.99903	0.91696	0.90236
0.00	0.97467	0.97113	0.89338	0.87921
0.10	0.94423	0.94086	0.86819	0.85450
0.50	0.82566	0.82282	0.76878	0.75671
1.00	0.70597	0.70324	0.66261	0.65169
2.00	0.54916	0.54592	0.50894	0.49937

Figure 9 shows crack displacement profiles considering discontinuous loading

$$p(x) = -1 + 0.5\,\mathrm{sgn}(x)$$

and $\tilde{\ell} = 0.05$, $\tilde{\gamma} = 0.2$, and various $\rho = \ell'/\ell$. Similar comments to those regarding Fig. 8 can be made with respect to Fig. 9. Moreover, qualitatively the results displayed in Figs. 7 to 9 are in agreement with those of Vardoulakis et al. [11] for homogeneous materials.

Table 4 shows a convergence study for (normalized) generalized SIFs (see Eqs. (80), (81), and (83), (84)) involving nongraded ($\tilde{\gamma} = 0$) and graded ($\tilde{\gamma} \neq 0$) gradient elastic materials considering both $\tilde{\ell}' = 0$ and $\tilde{\ell}' \neq 0$ ($\tilde{\ell}' > 0$). Note that as the number of collocation points (N) increases, the generalized SIF results converge for both materials (i.e., nongraded and graded). However, the convergence is worse for the case $\tilde{\ell}' \neq 0$ than for the case $\tilde{\ell}' = 0$. The condition number for all the examples investigated is always satisfactory.

Table 5 lists the generalized SIFs (see Eqs. (80), (81)) for gradient elastic materials considering various values of the material parameter γ and using $N = 61$ collocation points in the numerical solution. Notice that the SIF monotonically decreases as γ increases, which is in full agreement with the early results for classical elasticity considering nonhomogeneous materials (see Table 3). Consider, for example, the case $\tilde{\gamma} = 0$. In this case, the crack stiffening is due to the characteristic material lengths $\tilde{\ell}$ and $\tilde{\ell}'$ ($\tilde{\ell}' > 0$) of the structured medium which are responsible for lower generalized SIFs (< 1.0) and, consequently, lower energy release rates during crack propagation. The results indicate that a higher external load, as compared to that of the classical case, must be applied on the crack surfaces (or on the remote boundaries) to propagate it in a material with microstructure.

A few comments about the determination of characteristic lengths in continua with microstructure are in order. Shi et al. [40] have presented a brief discussion on determination of such lengths in the context of Fleck and Hutchinson's [3] strain gradient theory, which is a generalization of Mindlin's higher-order continuum theory, [41,42]. Experimental work in the field include, for example, micro-torsion by Fleck et al. [43], microbending by Stolkens and Evans [44], and microindentation by Nix [45]. The characterization of actual materials, with respect to strain gradient length-scale(s), is an ongoing research topic of much interest and impact in the field of applied mechanics.

11 Concluding Remarks

This paper has presented a theoretical framework and corresponding computational implementation for modeling antiplane shear cracks in functionally graded materials (FGMs) using strain gradient elasticity (Casal's continuum), which includes both volumetric and surface energy terms. The characteristic lengths (ℓ and ℓ', respectively) associated to these terms are assumed to be constant, and the material shear modulus is assumed to vary exponentially (see Eq. (18)). In this study, the crack is considered to be perpendicular to the material gradient. The present hypersingular integrodifferential equation approach leads to a numerically tractable solution of the fracture problem, and relevant fracture parameters have been investigated. These results include, for example, crack displacement profiles and generalized stress intensity factors. A parametric study including various gradation parameters (γ) and strain gradient parameters ($\tilde{\ell}, \tilde{\ell}'$) has been conducted and discussed. A natural extension of this work is the solution of an antiplane shear crack where the crack is parallel to the material gradation. Another potential extension consists of investigating the mode I fracture problem.

Acknowledgments

We acknowledge the support from the USA National Science Foundation (NSF) through grants CMS-9996378 (previously CMS-9713798) from the Mechanics & Materials Program, and DMS-9600119 from the Applied Mathematics Program. The first author would like to thank Prof. Y. F. Dafalias, form the University of California at Davis, for his encouragement and valuable suggestions to this work.

Appendix A

The Regular Kernel. The regular kernel $N(\xi,0)$ described in Eq. (60) can be expressed as the fraction $P(\xi)/Q(\xi)$. $Q(\xi)$ is given by

$$Q(\xi) = -i\xi\left(\sqrt{\xi^2+\gamma^2/4+1/\ell^2}+\sqrt{\xi^2+\gamma^2/4}+\gamma+\ell'/\ell^2\right). \tag{86}$$

$P(\xi)$ can be expressed as

$$P(\xi) = P_4(\xi)+P_3(\xi)+P_2(\xi)+P_1(\xi)+P_0(\xi) \tag{87}$$

in which

$$P_4(\xi) = \ell^2\xi^2\times\left(\sqrt{\xi^2+\gamma^2/4+1/\ell^2}\sqrt{\xi^2+\gamma^2/4}+\xi^2 - |\xi|\sqrt{\xi^2+\gamma^2/4+1/\ell^2}-|\xi|\sqrt{\xi^2+\gamma^2/4}\right), \tag{88}$$

$$P_3(\xi) = \tfrac{1}{2}\left(\gamma\ell^2+\ell'\right)\xi^2\left(\sqrt{\xi^2+\gamma^2/4+1/\ell^2}+\sqrt{\xi^2+\gamma^2/4}\right) - \left(\gamma\ell^2+\ell'\right)|\xi|^3, \tag{89}$$

$$P_2(\xi) = \left[1+\gamma\left(\gamma\ell^2+\ell'\right)\right]\sqrt{\xi^2+\gamma^2/4+1/\ell^2}\sqrt{\xi^2+\gamma^2/4} + \left[1+\frac{1}{4}\gamma^2\ell^2-\frac{1}{2}\left(\frac{\ell'}{\ell}\right)^2-\frac{1}{2}\gamma\ell'\right]\xi^2 - \left[1+\frac{5}{8}\gamma^2\ell^2-\left(\frac{\ell'}{2\ell}\right)^2+\frac{1}{4}\gamma\ell'\right] \times|\xi|\left(\sqrt{\xi^2+\gamma^2/4+1/\ell^2}+\sqrt{\xi^2+\gamma^2/4}\right), \tag{90}$$

$$P_1(\xi) = \frac{1}{2}\gamma\left(1+\gamma^2\ell^2+\gamma\ell'\right)\sqrt{\xi^2+\gamma^2/4+1/\ell^2} + \left[\frac{\gamma}{2}\left(1+\gamma^2\ell^2+\gamma\ell'\right)+\frac{\ell'}{\ell^2}\right]\sqrt{\xi^2+\gamma^2/4} - \left(\gamma+\frac{\ell'}{\ell^2}\right)\left[1+\frac{5}{8}\gamma^2\ell^2-\left(\frac{\ell'}{2\ell}\right)^2+\frac{1}{4}\gamma\ell'\right]|\xi|, \tag{91}$$

$$P_0(\xi) = \frac{1}{4}\ell^2\gamma^4+\frac{3}{4}\gamma^2+\frac{1}{4}\gamma^3\ell'+\frac{1}{2}\frac{\gamma\ell'}{\ell^2}. \tag{92}$$

Appendix B

Singular and Hypersingular Integrals. Closed-form solutions for evaluating singular and hypersingular integrals are provided here and can also be found in Chan et al. [34]. Those integrals are interpreted in the finite-part sense.

The solution of the crack boundary value problem requires the following formulas. Thus for $|r| < 1$, we have

$$\frac{1}{\pi}\int_{-1}^{1}\frac{U_n(s)\sqrt{1-s^2}}{s-r}\,ds = -T_{n+1}(r), \quad n \geq 0, \tag{93}$$

$$\frac{1}{\pi}\int\!\!\!\!\!\!=_{-1}^{1}\frac{U_n(s)\sqrt{1-s^2}}{(s-r)^2}\,ds = -(n+1)U_n(r), \quad n \geq 0 \tag{94}$$

$$\frac{1}{\pi}\oint_{-1}^{1}\frac{U_n(s)\sqrt{1-s^2}}{(s-r)^3}ds=\begin{cases}-1, & n=0,\\ [(n^2+n)U_{n+1}(r)-(2n^2+3n+2)U_{n-1}(r)]/[4(1-r^2)], & n\geq 1.\end{cases} \tag{95}$$

The calculation of stress intensity factors requires the following formulas. Thus, for $|r|>1$, we have

$$\frac{1}{\pi}\int_{-1}^{1}\frac{U_n(s)\sqrt{1-s^2}}{s-r}ds=-\left(r-\frac{|r|}{r}\sqrt{r^2-1}\right)^{n+1},\quad n\geq 0 \tag{96}$$

$$\frac{1}{\pi}\int_{-1}^{1}\frac{U_n(s)\sqrt{1-s^2}}{(s-r)^2}ds=-(n+1)\left(1-\frac{|r|}{\sqrt{r^2-1}}\right)\times\left(r-\frac{|r|}{r}\sqrt{r^2-1}\right)^{n},\quad n\geq 0 \tag{97}$$

$$\frac{1}{\pi}\int_{-1}^{1}\frac{U_n(s)\sqrt{1-s^2}}{(s-r)^3}ds=\frac{-1}{2}(n+1)\left(r-\frac{|r|}{r}\sqrt{r^2-1}\right)^{n-1}\times\left[n\left(1-\frac{|r|}{\sqrt{r^2-1}}\right)^2+\frac{r-\frac{|r|}{r}\sqrt{r^2-1}}{\sqrt{r^2-1^3}}\right],\quad n\geq 0. \tag{98}$$

References

[1] Eringen, A. C., 1999, *Microcontinuum Field Theories I. Foundations and Solids*, Springer-Verlag, New York.

[2] Wu, C. H., 1992, "Cohesive Elasticity and Surface Phenomena," Q. Appl. Math., **50**(1), pp. 73–103.

[3] Fleck, N. A., and Hutchinson, J. W., 1997, "Strain Gradient Plasticity," Adv. Appl. Mech., **33**, pp. 295–361.

[4] Lakes, R. S., 1983, "Size Effects and Micromechanics of a Porous Solid," J. Mater. Sci., **18**, pp. 2572–2580.

[5] Lakes, R. S., 1986, "Experimental Microelasticity of Two Porous Solids," Int. J. Solids Struct., **22**, pp. 55–63.

[6] Smyshlyaev, V. P., and Fleck, N. A., 1996, "The Role of Strain Gradients in the Grain Size Effect for Polycrystals," J. Mech. Phys. Solids, **44**(4), pp. 465–495.

[7] Van Vliet, M. R. A., and Van Mier, J. G. M., 1999, "Effect of Strain Gradients on the Size Effect of Concrete in Uniaxial Tension," Int. J. Fract., **95**, pp. 195–219.

[8] Fannjiang, A. C., Chan, Y.-S., and Paulino, G. H., 2001, "Strain Gradient Elasticity for Antiplane Shear Cracks: A Hypersingular Integrodifferential Equation Approach," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **62**(3), pp. 1066–1091.

[9] Paulino, G. H., Fannjiang, A. C., and Chan, Y.-S., 1999, "Gradient Elasticity Theory for a Mode III Crack in a Functionally Graded Material," Mater. Sci. Forum, **308–311**, pp. 971–976.

[10] Exadaktylos, G., Vardoulakis, I., and Aifantis, E., 1996, "Cracks in Gradient Elastic Bodies With Surface Energy," Int. J. Fract., **79**(2), pp. 107–119.

[11] Vardoulakis, I., Exadaktylos, G., and Aifantis, E., 1996, "Gradient Elasticity With Surface Energy: Mode-III Crack Problem," Int. J. Solids Struct., **33**(30), pp. 4531–4559.

[12] Aifantis, E., 1992, "On the Role of Gradients in the Localization of Deformation and Fracture," Int. J. Eng. Sci., **30**, pp. 1279–1299.

[13] Zhang, L., Huang, Y., Chen, J. Y., and Hwang, K. C., 1998, "The Mode III Full-Field Solution in Elastic Materials With Strain Gradient Effects," Int. J. Fract., **92**(4), pp. 325–348.

[14] Hwang, K. C., Cuo, T. F., Huang, Y., and Chen, J. Y., 1998, "Fracture in Strain Gradient Elasticity," Met. Mater., **4**(4), pp. 593–600.

[15] Hutchinson, J. W., and Evans, A. G., 2000, "Mechanics of Materials: Top-Down Approaches to Fracture," Acta Mater., **48**, pp. 125–135.

[16] Carrillo-Heian, E. M., Carpenter, R. D., Paulino, G. H., Gibeling, J. C., and Munir, Z. A., 2001, "Dense Layered $MoSi_2$/SiC Functionally Graded Composites Formed by Field-Activated Synthesis," J. Am. Ceram. Soc., **84**(5), pp. 962–968.

[17] Jin, Z.-H., and Paulino, G. H., 2001, "Transient Thermal Stress Analysis of an Edge Crack in a Functionally Graded Material," Int. J. Fract., **107**(1), pp. 73–98.

[18] Carrillo-Heian, E. M., Unuvar, C., Gibeling, J. C., Paulino, G. H., and Munir, Z. A., 2001, "Simultaneous Synthesis and Densification of Niobium Silicide/Niobium Composites," Scr. Mater., **45**(4), pp. 405–412.

[19] Carpenter, R. D., Liang, W. W., Paulino, G. H., Gibeling, J. C., and Munir, Z. A., 1999, "Fracture Testing and Analysis of a Layered Functionally Graded Ti/TiB Beam in 3-Point Bending," Mater. Sci. Forum, **837–842**, pp. 971–976.

[20] Markworth, A. J., Ramesh, K. S., and Parks, Jr., W. P., 1995, "Review Modelling Studies Applied to Functionally Graded Materials," J. Mater. Sci., **30**, pp. 2183–2193.

[21] Erdogan, F., 1995, "Fracture Mechanics of Functionally Graded Materials," Composites Eng., **5**(7), pp. 753–770.

[22] Hirai, T., 1996, "Functional Gradient Materials," *Materials Science and Technology*, (Vol. 17B of *Processing of Ceramics, Part 2*), R. J. Brook, ed., VCH Verlagsgesellschaft mbH, Weinheim, Germany, pp. 292–341.

[23] Suresh, S., and Mortensen, A., 1998, *Fundamentals of Functionally Graded Materials*, ASM International and the Institute of Materials, IOM Communications Ltd., London.

[24] Casal, P., 1961, "La Capillarite Interne," Cah. Groupe Fr. Etud. Rheol., **6**(3), pp. 31–37.

[25] Casal, P., 1963, "Capillarite Interne en Mecanique," C.R. Acad. Sci., **256**, pp. 3820–3822.

[26] Casal, P., 1972, "La théorie du second gradient et la capillarité," C.R. Acad. Sci. Paris Sér. A, **274**, pp. 1571–1574.

[27] Chan, Y.-S., Paulino, G. H., and Fannjiang, A. C., 2003, "Change of Constitutive Relations due to Interaction Between Strain Gradient Effect and Material Gradation," to be submitted.

[28] Erdogan, F., and Ozturk, M., 1992, "Diffusion Problems in Bonded Nonhomogeneous Materials With an Interface Cut," Int. J. Eng. Sci., **30**(10), pp. 1507–1523.

[29] Sneddon, I. N., 1972, *The Use of Integral Transforms*, McGraw-Hill, New York.

[30] Martin, P. A., 1991, "End-Point Behavior of Solutions to Hypersingular Integral Equations," Proc. R. Soc. London, Ser. A, **432**(1885), pp. 301–320.

[31] Erdogan, F., and Gupta, G. D., 1972, "On the Numerical Solution of Singular Integral Equations," Q. Appl. Math., **30**, pp. 525–534.

[32] Erdogan, F., Gupta, G. D., and Cook, T. S., 1973, "Numerical Solution of Singular Integral Equations," *Mechanics of Fracture*, G. C. Sih, Ed., Vol. 1, Noordhoff, Leyden, The Netherlands, pp. 368–425.

[33] Chan, Y.-S., Paulino, G. H., and Fannjiang, A. C., 2001, "The Crack Problem for Nonhomogeneous Materials Under Antiplane Shear Loading—A Displacement Based Formulation," Int. J. Solids Struct., **38**(17), pp. 2989–3005.

[34] Chan, Y.-S., Fannjiang, A. C., and Paulino, G. H., 2003, "Integral Equations With Hypersingular Kernels—Theory and Applications to Fracture Mechanics," Int. J. Eng. Sci., **41**(7), pp. 683–720.

[35] Kaya, A. C., and Erdogan, F., 1987, "On the Solution of Integral Equations With Strongly Singular Kernels," Q. Appl. Math., **45**(1), pp. 105–122.

[36] Folland, G. B., 1992, *Fourier Analysis and Its Applications*, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

[37] Stroud, A. H., and Secrest, D., 1996, *Gaussian Qudrature Formulas*, Prentice-Hall, New York.

[38] Barenblatt, G. I., 1962, "The Mathematical Theory of Equilibrium Cracks in Brittle Fracture," Adv. Appl. Mech., **7**, pp. 55–129.

[39] Vardoulakis, I., and Sulem, J., 1995, *Bifurcation Analysis in Geomechanics*, Blackie Academic and Professional, Glasgow.

[40] Shi, M. X., Huang, Y., and Hwang, K. C., 2000, "Fracture in a Higher-Order Elastic Continuum," J. Mech. Phys. Solids, **48**(12), pp. 2513–2538.

[41] Mindlin, R. D., 1964, "Micro-Structure in Linear Elasticity," Arch. Ration. Mech. Anal., **16**, pp. 51–78.

[42] Mindlin, R. D., 1965, "Second Gradient of Strain and Surface-Tension in Linear Elasticity," Int. J. Solids Struct., **1**, pp. 417–438.

[43] Fleck, N. A., Muller, G. M., Asby, M. F., and Hutchinson, J. W., 1994, "Strain Gradient Plasticity: Theory and Experiments," Acta Metall. Mater., **42**, pp. 475–487.

[44] Stolken, J. S., and Evans, A. G., 1998, "A Microbend Test Method for the Plasticity Length Scale," Acta Mater., **46**, pp. 5109–5115.

[45] Nix, W. D., 1997, "Elastic and Plastic Properties of Thin Films on Substrates: Nanoindentation Techniques," Mater. Sci. Eng., A, **234/236**, pp. 37–44.

L. J. Gray

T. Kaplan

Computer Science and Mathematics Division,
Oak Ridge National Laboratory,
P.O. Box 2008, Building 6012,
Oak Ridge, TN 37831-6367

J. D. Richardson

Department of Mechanical Engineering,
Tennessee Technological University,
P.O. Box 5014,
Cookeville, TN 38505

G. H. Paulino[1]

Department of Civil and Environmental Engineering,
University of Illinois at Urbana-Champaign,
Newmark Laboratory,
205 North Mathews Avenue,
Urbana, IL 61801
e-mail: paulino@uiuc.edu
Mem. ASME

Green's Functions and Boundary Integral Analysis for Exponentially Graded Materials: Heat Conduction

Free space Green's functions are derived for graded materials in which the thermal conductivity varies exponentially in one coordinate. Closed-form expressions are obtained for the steady-state diffusion equation, in two and three dimensions. The corresponding boundary integral equation formulations for these problems are derived, and the three-dimensional case is solved numerically using a Galerkin approximation. The results of test calculations are in excellent agreement with exact solutions and finite element simulations. [DOI: 10.1115/1.1485753]

1 Introduction

Functionally graded materials (FGMs) are an important area of materials science research, with potentially many important applications, e.g., super-heat resistance materials for thermal barrier coatings and furnace liners, vehicle and personal body armor, electromagnetic sensors, and graded refractive index materials for optical applications. In an ideal FGM, the material properties may vary smoothly in one dimension (e.g., are constant in (x, y) but vary with z). As an example, having a smooth transition region between a pure metal and pure ceramic would result in a material that combines the desirable high temperature properties and thermal resistance of a ceramic, with the fracture toughness of a metal. Comprehensive reviews of current FGM research may be found in the articles by Hirai [1], Markworth et al. [2] and Paulino et al. [3], and the book by Suresh and Mortensen [4].

Computational analysis can be an effective method for designing specific FGM systems, and for understanding FGM behavior. For homogeneous media, boundary integral equation methods (e.g., [5]) have been used extensively. However, the reformulation in terms of integral equations relies upon having, as either a closed form or a computable expression, a fundamental solution (Green's function) of the partial differential equation. Application of the boundary integral technique has therefore been limited, almost exclusively, to homogeneous, or piecewise homogeneous, media.

The fundamental solutions traditionally employed in boundary integral analysis for homogeneous materials are ''free space'' Green's functions: They satisfy the appropriate differential equation everywhere in space, except at the site where a point load driving force is applied. Derivations for some of the basic Green's functions can be found in [5,6]. There has also been work in the direction of deriving Green's functions for a general nonhomogeneous material ([7–11]). Steady-state heat conduction with an arbitrary spatially varying conductivity has recently been investigated ([12,13]) and has generated some debate in the literature ([14,15]). In most cases, exact Green's functions are only obtained under certain restrictions.

In the present paper, we derive free space fundamental solutions for both the two-dimensional and three-dimensional FGM Laplace equation, assuming that the thermal conductivity varies exponentially. The corresponding boundary integral equation formulation, which turns out to be somewhat different from the homogeneous media case, is also obtained, and numerical results based upon a Galerkin approximation are presented. Relatively little attention has been paid to obtaining Green's functions for the special case of graded materials: A Green's function for a special type of elastodynamics problem was obtained by Vrettos [16], and exponential grading was also considered in [11]. The two-dimensional Green's function results have appeared in conjunction with a convective heat transfer problem in a *homogeneous* material ([17,18]), and moreover [19] essentially contains the Green's functions derived herein (obtained in a different manner). However, the analysis employed in this paper for heat conduction in an exponential FGM will carry over to the important case of elasticity ([20]), and thus it is deemed useful to present this alternate derivation in detail.

This paper is organized as follows. The three-dimensional Laplace equation is treated in Section 2.1, and the two-dimensional case in Section 2.2. Section 3 discusses some test results from a Galerkin numerical implementation of the boundary integral formulation, and Section 4 contains some concluding remarks. Finally, in the Appendix it is shown that the integral equations and Green's functions can be suitably modified to allow for a Symmetric-Galerkin implementation. Complete formulas for the three-dimensional reformulated fundamental solutions and their first and second derivatives, for the case that the thermal conductivity is real, are also given in this Appendix.

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Dec. 14, 2000; final revision, Oct. 30, 2001. Associate Editor: M.-J. Pindera. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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 Green's Functions

Steady-state isotropic heat conduction in a solid is governed by the equation

$$\nabla \cdot (k \nabla \phi) = 0. \tag{1}$$

Here $\phi = \phi(x,y,z)$ is the temperature function, and we assume the functionally graded material is defined by the thermal conductivity

$$k(x,y,z) = k(z) = k_0 e^{-2i\alpha z}, \tag{2}$$

where α is real. This assumption of a purely imaginary exponent is apparently necessary for the derivation that follows. However, once the solution is obtained, it is readily seen to be valid for any complex α. Substituting Eq. (2) into Eq. (1), one obtains that the temperature satisfies

$$\nabla^2 \phi - 2i\alpha \phi_z = 0, \tag{3}$$

where ϕ_z denotes the derivative with respect to z.

The Green's function equation can be derived by constructing the integral equation corresponding to Eq. (3). Following the standard procedure, Eq. (3) is multiplied by an arbitrary function $f(x,y,z) = f(Q)$ and integrated over a bounded volume V. Integrating by parts, and denoting the boundary of V by Σ, one obtains

$$\begin{aligned} 0 &= \int_V f(Q)(\nabla^2 \phi(Q) - 2i\alpha \phi_z(Q)) dV_Q . \\ &= \int_\Sigma \left\{ f(Q) \frac{\partial}{\partial n} \phi(Q) - \phi(Q) \frac{\partial}{\partial n} f(Q) \right. \\ &\quad \left. - 2i\alpha n_z(Q) \phi(Q) f(Q) \right\} dQ + \int_V \phi(Q)(\nabla^2 f(Q) \\ &\quad + 2i\alpha f_z(Q)) dV_Q , \end{aligned} \tag{4}$$

where $\mathbf{n}(Q) = (n_x, n_y, n_z)$ is the unit outward normal for Σ. If $f(Q) = G(P,Q)$ satisfies the Green's function equation (the adjoint to Eq. (3))

$$\nabla^2 G(P,Q) + 2i\alpha G_z(P,Q) = -\delta(Q-P), \tag{5}$$

where δ is the Dirac delta function, the remaining volume integral becomes simply $-\phi(P)$. Thus we obtain the governing boundary integral equation

$$\begin{aligned} \phi(P) &+ \int_\Sigma \phi(Q) \left(\frac{\partial}{\partial n} G(P,Q) + 2i\alpha n_z G(P,Q) \right) dQ \\ &= \int_\Sigma G(P,Q) \frac{\partial}{\partial n} \phi(Q) dQ, \end{aligned} \tag{6}$$

which differs in form from the usual integral statements by the presence of the additional term multiplying $\phi(Q)$. With obvious changes (e.g., line integrals instead of surface integrals), the above equations are equally valid for two dimensions. We first derive the Green's function for three dimensions.

2.1 Three Dimensions. Let $\hat{f}(\boldsymbol{\omega})$ denote the Fourier transform of a function $\mathcal{F}(Q)$,

$$\hat{f}(\boldsymbol{\omega}) = \int_{\mathcal{R}^3} \mathcal{F}(Q) e^{-i\boldsymbol{\omega} \cdot Q} dQ \tag{7}$$

where $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ is the transform variable and the dot represents the inner product. Transforming Eq. (5) and solving for $\hat{G}(\boldsymbol{\omega})$ (the transform of G with respect to Q), yields

$$\hat{G}(\boldsymbol{\omega}) = \frac{e^{-i\boldsymbol{\omega} \cdot P}}{\omega^2 + 2\alpha\omega_z}, \tag{8}$$

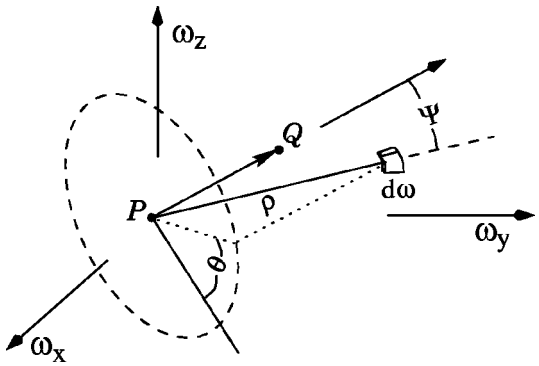


Fig. 1 Spherical coordinate system for evaluating the ω integral

where $\omega^2 = \boldsymbol{\omega} \cdot \boldsymbol{\omega}$. Applying the inverse transform, one obtains

$$G(P,Q) = \frac{1}{(2\pi)^3} \int_{\mathcal{R}^3} \frac{e^{i\boldsymbol{\omega} \cdot (Q-P)}}{\omega^2 + 2\alpha\omega_z} dw, \tag{9}$$

where dw is shorthand for the three-dimensional differential element, i.e., $dw = d\omega_x d\omega_y d\omega_z$. Changing variables

$$\omega_z \to \omega_z - \alpha \tag{10}$$

and setting $R = Q - P$, $R_z = Q_z - P_z$, we obtain

$$G(P,Q) = \frac{1}{(2\pi)^3} e^{-i\alpha R_z} \int_{\mathcal{R}^3} \frac{e^{i\boldsymbol{\omega} \cdot R}}{\omega^2 - \alpha^2} d\omega, \tag{11}$$

which can be conveniently split into two terms,

$$G(P,Q) = \frac{e^{-i\alpha R_z}}{(2\pi)^3} \left[\int_{\mathcal{R}^3} \frac{e^{i\boldsymbol{\omega} \cdot R}}{\omega^2} d\omega + \alpha^2 \int_{\mathcal{R}^3} \frac{e^{i\boldsymbol{\omega} \cdot R}}{\omega^2 (\omega^2 - \alpha^2)} d\omega \right]. \tag{12}$$

The first integral is Eq. (9) with $\alpha = 0$, and is therefore recognized as the Green's function for the Laplace equation (constant k), the point source potential:

$$\frac{e^{-i\alpha R_z}}{(2\pi)^3} \int_{\mathcal{R}^3} \frac{e^{i\boldsymbol{\omega} \cdot R}}{\omega^2} d\omega = \frac{e^{-i\alpha R_z}}{4\pi r}, \tag{13}$$

where $r = \|R\| = \|Q - P\|$ is the distance between the source point P and the field point Q.

To evaluate the second term in Eq. (12), it is convenient to employ spherical coordinates (ρ, θ, ψ), with, however, the axis defining the pole $\psi = 0$ taken as the direction R/r instead of the z-axis (see Fig. 1). The integration limits are $0 < \rho < \infty$, $0 \leq \psi \leq \pi$, and $0 \leq \theta \leq 2\pi$; however, for the residue calculations to follow, it will be much more convenient to have $-\infty < \rho < \infty$ and $0 \leq \psi \leq \pi/2$. With the standard limits, the residue calculation must shift half-planes depending upon the sign of $\cos(\psi)$; more importantly, starting at $\rho = 0$ would force consideration of contours along the imaginary axis, necessary to work with the imaginary part of the exponential. In comparison, if ρ varies over the entire real axis, a simple semicircle in the upper half-plane suffices. To this end, *if* the function f satisfies $f(\rho, \psi) = f(-\rho, \pi - \psi)$, then

$$\begin{aligned} &\int_0^\infty \int_0^\pi f(\rho,\psi) d\psi d\rho \\ &\quad = \int_0^\infty \int_0^{\pi/2} f(\rho,\psi) d\psi d\rho + \int_0^\infty \int_{\pi/2}^\pi f(\rho,\psi) d\psi d\rho \\ &\quad = \int_0^\infty \int_0^{\pi/2} f(\rho,\psi) d\psi d\rho + \int_0^\infty \int_0^{\pi/2} f(\rho,\pi-\psi) d\psi d\rho \end{aligned}$$

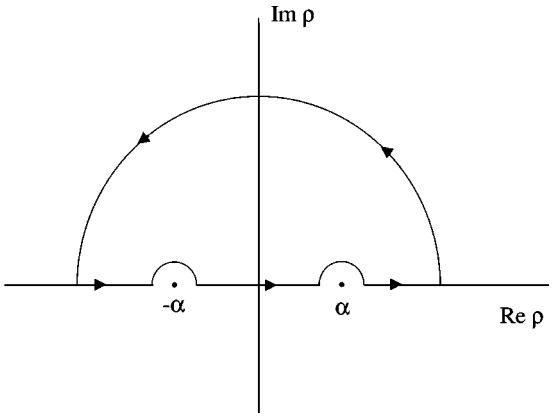


Fig. 2 Contour in the complex plane used to compute the ρ integration

$$= \int_0^\infty \int_0^{\pi/2} f(\rho,\psi)\,d\psi d\rho + \int_{-\infty}^0 \int_0^{\pi/2} f(-\rho,\pi-\psi)\,d\psi d\rho$$

$$= \int_{-\infty}^\infty \int_0^{\pi/2} f(\rho,\psi)\,d\psi d\rho. \tag{14}$$

It will turn out that the function to be integrated satisfies the above constraint, and thus the modified limits of integration for ρ and ψ can be employed. As mentioned above, this greatly simplifies the residue procedures for the ρ integration.

Noting that $\boldsymbol{\omega}\cdot R = \rho r\cos(\psi)$ and that, other than this exponential, the integrand is a function of ω^2 and independent of θ, this second term therefore becomes

$$\frac{\alpha^2 e^{-i\alpha R_z}}{(2\pi)^2}\int_0^{\pi/2}\sin(\psi)\,d\psi\int_{-\infty}^{\infty}\frac{e^{i\rho r\cos(\psi)}}{\rho^2-\alpha^2}\,d\rho. \tag{15}$$

Using the contour shown in Fig. 2, the ρ integration is a straightforward exercise in residue calculus, yielding

$$\int_{-\infty}^{\infty}\frac{e^{i\rho r\cos(\psi)}}{\rho^2-\alpha^2}\,d\rho = -\frac{\pi}{\alpha}\sin(\alpha r\cos(\psi)). \tag{16}$$

The final integration,

$$-\frac{\pi}{\alpha}\int_0^{\pi/2}\sin(\psi)\sin(\alpha r\cos(\psi))\,d\psi, \tag{17}$$

follows from a simple change of variables, and thus the second term is seen to be

$$\frac{e^{-i\alpha R_z}\cos(\alpha r)}{4\pi r} - \frac{e^{-i\alpha R_z}}{4\pi r}. \tag{18}$$

Including Eq. (13), we find the simple result

$$G(P,Q) = \frac{e^{-i\alpha R_z}\cos(\alpha r)}{4\pi r}. \tag{19}$$

Although this result was derived assuming that α is real, it is a simple matter to check by direct calculation that Eq. (19) satisfies Eq. (5) for any complex α. It is useful, especially for the discussion of the two-dimensional case that follows, to observe that

$$G(P,Q) = e^{-i\alpha R_z}\frac{e^{-i\alpha r}}{4\pi r} \tag{20}$$

is an equally valid solution of Eq. (5) for α real. Moreover, the added $\sin(\alpha r)/r$ term is regular as $r\to 0$, and thus does not alter the delta function at $Q=P$. Replacing α by $i\beta_0$, where β_0 is real, we obtain

$$G(P,Q) = \frac{e^{\beta_0(r+R_z)}}{4\pi r} \tag{21}$$

as the Green's function for $k(z)=e^{2\beta_0 z}$.

In the derivation of the boundary integral equation, a sphere S_ε of radius ε centered at the interior point P would be removed from V, and the integration over Σ would include the surface of this sphere. The limit as $\varepsilon\to 0$ of the integral

$$\int_{S_\varepsilon}\left\{G(P,Q)\frac{\partial}{\partial n}\phi(Q) - \phi(Q)\frac{\partial}{\partial n}G(P,Q) - 2i\alpha(\phi(Q)G(P,Q))n_z\right\}dQ \tag{22}$$

must therefore be considered. However, for $r\to 0$,

$$\frac{\partial}{\partial n}G(P,Q) \approx \frac{\partial}{\partial n}\frac{1}{4\pi r} \tag{23}$$

and the $\varepsilon=0$ limit does indeed produce the correct value $-\phi(P)$.

Finally, it is useful to note that Eq. (18) can, from the point of view of the singularity at $r=0$, be considered as a remainder term. That is, the singularity for the FGM Green's function is entirely contained within Eq. (13), the homogeneous steady-state solution, as Eq. (18) is regular at $r=0$.

2.2 Two Dimensions. The Green's function $g(x_Q,z_Q;x_P,z_P)$ for the two-dimensional equation,

$$\phi_{xx}+\phi_{zz}-2i\alpha\phi_z = 0, \tag{24}$$

is expected to behave as $\log(r)$, and as this function does not die off at infinity, the above Fourier transform approach is doomed to fail. However, this fundamental solution can be viewed as the response seen at the point $(x_Q,0,z_Q)$ to a uniform distribution of charge along the y-axis. This response should be the result of integrating the three-dimensional Green's function over this axis, which for the homogeneous case takes the form

$$\frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{dy_P}{((x_Q-x_P)^2+y_P^2+(z_Q-z_P)^2)^{1/2}}. \tag{25}$$

The fact that the integral doesn't exist is a minor inconvenience that is remedied by doing the analysis for $\partial G/\partial x_Q$ ([21]). The integral of this function with respect to y_P does exist, and followed by an integration over x_Q, the correct $\log(r)$ result is obtained, where r is now the two-dimensional distance.

With this framework in mind, we observe that the three-dimensional functionally graded material (FGM) Green's function, in the form of Eq. (20), is $e^{-i\alpha R_z}$ times the fundamental solution for the Helmholtz Eq. (3). Since this prefactor is independent of y_P, integrating out this coordinate as in Eq. (25), we expect that the two-dimensional FGM Green's function is given by

$$g(x_Q,z_Q;x_P,z_P) = \frac{i}{4}e^{-i\alpha R_z}H_0^1(\alpha r). \tag{26}$$

Here, H_0^1 is the zeroth-order first kind Hankel function ([22]), well known to be the solution of the Helmholtz equation in two dimensions. This expectation can be established simply by differentiating Eq. (26) and checking that

$$g_{xx}+g_{zz}+2i\alpha g_z = 0, \tag{27}$$

for $Q\neq P$ (this is the two-dimensional analogue of the Green's function equation, Eq. (5)). That this differentiation also yields a delta function at $Q=P$ follows from the known behavior of H_0^1 for the Helmholtz equation. Finally, it should be noted that the two-dimensional boundary integral equation becomes

$$\phi(P)+\int_\Sigma \phi(Q)\left(\frac{\partial}{\partial n}g(P,Q)+2i\alpha n_z g(P,Q)\right)dQ$$

$$=\int_\Sigma g(P,Q)\frac{\partial}{\partial n}\phi(Q)dQ, \tag{28}$$

which corresponds to Eq. (6) with $G(P,Q)$ (three-dimensional case) replaced by $g(P,Q)$ (two-dimensional case).

2.3 Extensions. As it may be useful to have the material properties vary in more than one component ([23]), it is worth noting that the above analysis extends to a more general exponential variation for k,

$$\mathrm{k}(x,y,z)=\mathrm{k}_0 e^{-2i\boldsymbol{\alpha}\cdot Q}, \tag{29}$$

where $\boldsymbol{\alpha}=(\alpha_x,\alpha_y,\alpha_z)$. The three-dimensional Green's function is now given by

$$G_{xyz}(P,Q)=\frac{e^{-i\boldsymbol{\alpha}\cdot R}\cos((\boldsymbol{\alpha}\cdot\boldsymbol{\alpha})r)}{4\pi r}. \tag{30}$$

Comparing this with Eq. (19), it is not surprising that the two-dimensional result in this case (again dropping out the y-coordinate) becomes

$$g_{xz}(x_Q,z_Q;x_P,z_P)=\frac{i}{4}e^{-i\boldsymbol{\alpha}\cdot R}H_0^1((\boldsymbol{\alpha}\cdot\boldsymbol{\alpha})r). \tag{31}$$

2.4 Galerkin Approximation. The numerical results presented in the next section utilize the Galerkin approximation ([5]) to reduce the integral equation to a finite system of equations. Here we briefly review this technique, starting by rewriting Eq. (6) as

$$\mathcal{P}(P)\equiv\phi(P)+\int_\Sigma \phi(Q)\left(\frac{\partial}{\partial n}G(P,Q)+2i\alpha n_z G(P,Q)\right)dQ$$

$$-\int_\Sigma G(P,Q)\frac{\partial}{\partial n}\phi(Q)dQ=0. \tag{32}$$

As is usual, basis shape functions $\psi_j(Q)$ are used to interpolate the boundary from the element nodal coordinates, and to approximate the surface potential and flux in terms of nodal values, i.e.,

$$\Sigma(\eta,\xi)=\sum_j (x_j,y_j,z_j)\psi_j(\eta,\xi)$$

$$\phi(Q)=\sum_j \phi_j\psi_j(Q) \tag{33}$$

$$\frac{\partial\phi}{\partial n}(Q)=\sum_j\left(\frac{\partial\phi}{\partial n}\right)_j\psi_j(Q).$$

The numerical results reported herein employ a six-noded quadratic triangular element, defined using the right triangle parameter space (η,ξ), $\eta\geqslant 0$, $\xi\geqslant 0$, $\eta+\xi\leqslant 1$. The shape functions are given by

$$\psi_1(\eta,\xi)=(1-\eta-\xi)(1-2\eta-2\xi) \qquad \psi_4(\eta,\xi)=4\eta(1-\eta-\xi)$$

$$\psi_2(\eta,\xi)=\eta(2\eta-1) \qquad \psi_5(\eta,\xi)=4\eta\xi \tag{34}$$

$$\psi_3(\eta,\xi)=\xi(2\xi-1) \qquad \psi_6(\eta,\xi)=4\xi(1-\eta-\xi).$$

In a Galerkin approximation, these shape functions are employed as weighting functions for enforcing Eq. (32) "on average," i.e.,

$$\int_\Sigma \psi_k(P)\int_\Sigma \mathcal{P}(P)dP=0. \tag{35}$$

When the approximations in Eq. (33) are incorporated into this equation, the resulting finite system of equations can be discretized and solved numerically.

It should also be noted that, unlike the Green's function $1/(4\pi r)$ for the Laplace equation (homogeneous problem), neither Eq. (20) nor Eq. (26) is a symmetric function of P and Q. It would therefore appear impossible to have a symmetric-Galerkin approximation ([24–28]), as this formulation demands a symmetric Green's function. However, as shown in the Appendix, a slight reworking of the equations and the kernel functions restores all of the necessary symmetry properties. This Appendix also provides formulas for all of the kernel functions: temperature and flux equations in two and three dimensions.

3 Numerical Examples

The three-dimensional steady-state fundamental solution has been incorporated into a boundary element method (BEM) algorithm. As noted above, the integral Eq. (6) is numerically approximated via the (nonsymmetric) Galerkin method (see Eq. (35)), together with standard six-node isoparametric quadratic triangular elements to interpolate the boundary and boundary functions. For the numerical examples, the conservation Eq. (1) will be taken as energy conservation in a functionally graded media under the condition of steady-state heat conduction without volumetric generation. To validate the numerical implementation, solutions to two test problems are presented below: In the first, the domain is a simple cube and the exact solution is known; the second involves a curved geometry which may be more representative of an actual systems component.

3.1 Unit Cube: Linear Heat Flux. For the first example problem, the geometry is a unit cube with the origin of a Cartesian system fixed at one corner. The thermal conductivity in this example is taken to be

$$\mathrm{k}(z)=\mathrm{k}_0 e^{2\beta z}=5e^{3z}. \tag{36}$$

The cube is insulated on the faces $[y=0]$ and $[y=1]$, while uniform heat fluxes of 5000 [POWER/AREA] are added and removed, respectively, at the $[x=1]$ and $[x=0]$ faces. In addition, the $[z=0]$ face is specified to have an x-dependent temperature distribution $T=1000\ x$ deg and at $[z=1]$ a normal heat flux of $q=15000\ x$ is removed. The analytic solution for this problem is

$$T=1000xe^{-3z}$$

$$\mathbf{q}=-5000\hat{\mathbf{i}}+15000x\hat{\mathbf{k}} \tag{37}$$

where $\hat{\mathbf{i}}$ is a unit vector in the x-direction.

The results of the numerical simulations for the temperature distributions along an edge are shown in Fig. 3. The plot also includes the results obtained from a finite element method (FEM) simulation using a commercial package. In the FEM simulation,

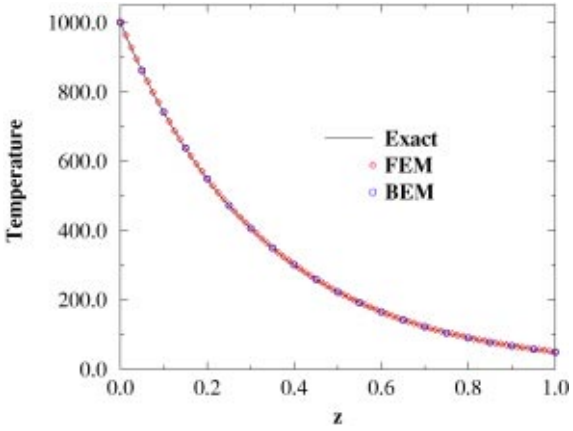


Fig. 3 Temperature distribution in the functionally graded material (FGM) unit cube along the edge $[x=1, y=1]$

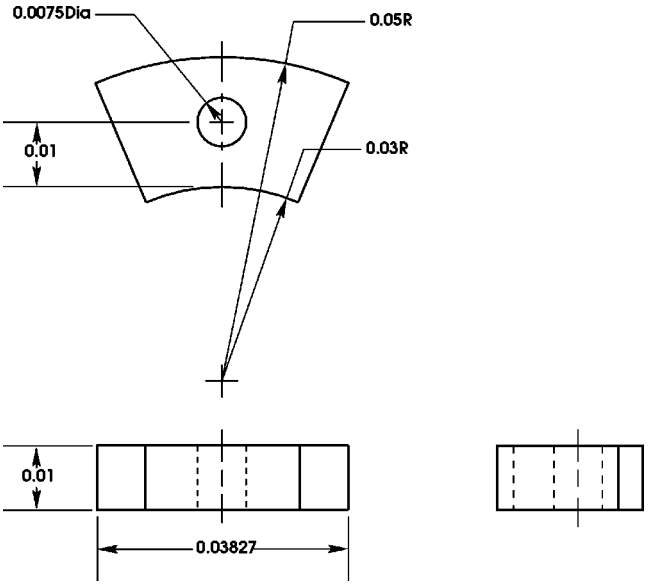


Fig. 4 Geometry of the functionally graded rotor

40 homogeneous layers were used to approximate the continuous grading; the conductivity of each layer was computed from Eq. (36) where z was taken as the z-coordinate of the layer's centroid. The FEM elements used were 20-node quadratic brick elements and each of the 40 layers contained 400 brick elements, resulting in a total of 69,720 nodes. In the boundary element method (BEM) solution, a uniform grid consisting of isosceles right triangles, with each leg having length 0.1, was employed, resulting in a total of 1200 elements and 2646 nodes.

3.2 Functionally Graded Material (FGM) Rotor. The second numerical example is a rotor with eight mounting holes. Due to the eightfold symmetry, only one-eighth of the rotor is modeled, as drawn in Fig. 4. The grading direction for the rotor is parallel to its line of symmetry, which is taken as the z-axis, and the thermal conductivity for the rotor varies according to

$$\mathrm{k}(z) = 20e^{330z}\frac{\mathrm{W}}{\mathrm{m\,K}}. \tag{38}$$

A schematic for the thermal boundary conditions is shown in Fig. 5. The temperature is specified along the inner and outer radii and a uniform heat flux of 5×10^5 W/m^2 is added on the bottom surface where $z=0$. All other surfaces are insulated as shown.

The BEM solution is compared with an FEM solution obtained from the same package used in the previous example using ten-

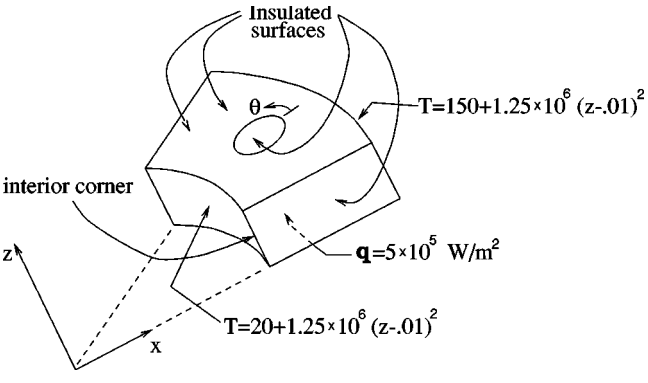


Fig. 5 Thermal boundary conditions on the rotor

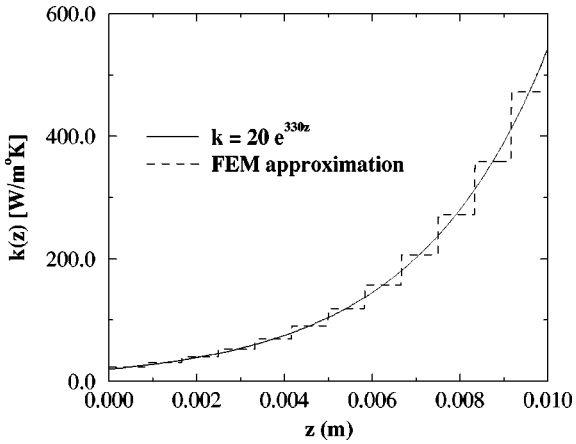


Fig. 6 Thermal conductivity profiles for the computational models of the rotor

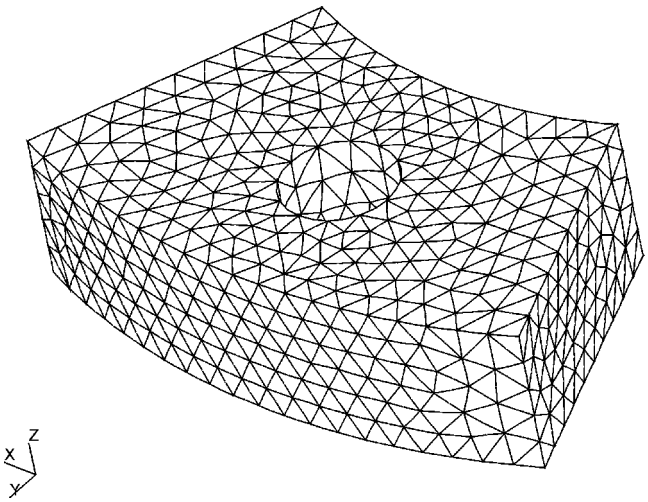


Fig. 7 Surface mesh employed on the functionally graded rotor

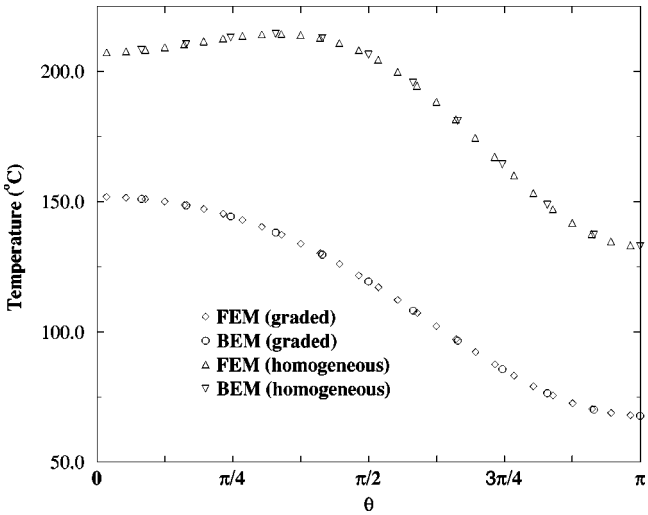


Fig. 8 Temperature distribution around the hole on the $z=0.01$ surface

node tetrahedral elements to handle the geometric complexity of the rotor. Due to resource limitations, the FEM model was limited to 12 layers which resulted in the rather crude conductivity profile shown in Fig. 6. Even so, the FEM mesh required 95,880 nodes,

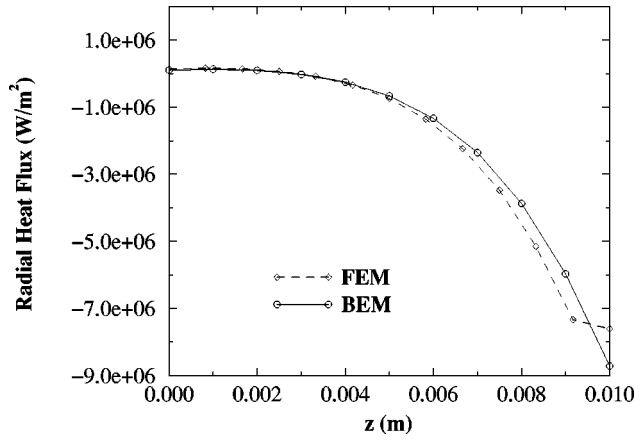


Fig. 9 Radial heat flux along the inside corner

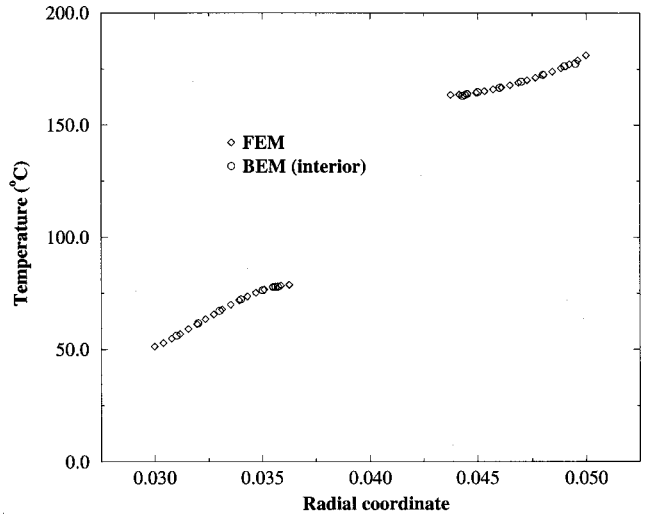


Fig. 10 Computed interior temperature values in the graded rotor

whereas the BEM mesh employed 3252. The mesh employed for the boundary integral analysis is shown in Fig. 7.

The temperature distribution around the hole is shown in Fig. 8. The angle θ is measured from a line passing through the line of symmetry for the geometry and the center line of the hole. Though surface nodal positions in the two models were not coincident in general, the plot shows a strong agreement in the two solutions. To see the effects of the grading upon the solution, the corresponding results for the *ungraded* rotor, $\beta=0$ $(\mathrm{k}(z)\equiv 20)$, are also shown.

The radial heat flux along the line shown as the interior corner in Fig. 5 is plotted in Fig. 9. The negative sign indicates that the flow of heat is toward the interior of the rotor. A limitation on the use of piecewise constant conductivities in FEM models may be evident in the plot where the FEM nodal value at $z=0.01$ seems to fall out of line with the other values on the curve. The behavior should be fully expected, however, given the local error associated with the piecewise constant approximation seen near $z=0.01$ in Fig. 6. As should also be expected, the nodal flux values from the BEM solution seem to fall onto a single curve even in the region of the steepest conductivity gradient. This is *not* to say that BEM is necessarily better than FEM for graded analysis: The finite element method is not restricted to using the discontinuous piecewise constant approximation presently available in existing packages. It is possible to incorporate continuous grading within individual elements, as demonstrated recently by Kim and Paulino [29] using a generalized isoparametric formulation.

As a final test, Fig. 10 displays a comparison between the FEM interior temperature values, and corresponding values computed from the BEM solution (in a post-processing calculation). The values are shown for a line of points on the mid-z $(z=0.005)$ plane in the radial direction, passing through the middle of the hole. Again, the BEM and FEM results agree quite well.

4 Conclusions

The primary conclusion of this work is that boundary integral analysis, for the most part limited to applications involving homogeneous or piecewise homogeneous media, can be successfully applied to exponentially graded materials. Although the simplest case, namely the Laplace equation, has been treated herein, it is expected that other applications, including transient diffusion ([30]) and elasticity ([20]), can also be addressed. Note that a specific elastodynamics problem has already been addressed by Vrettos [16].

The numerical results presented in this paper have shown that it is simple to implement the functionally graded material (FGM) Green's function in a standard boundary integral (Galerkin) approximation, and that accurate results are obtained. For graded materials, this offers the possibility of efficient and accurate solution of those types of problems for which a boundary integral analysis is particularly advantageous, such as shape optimization, moving boundaries, and small-scale structures.

Acknowledgments

This research was supported in part by the Applied Mathematical Sciences Research Program of the Office of Mathematical, Information, and Computational Sciences, U.S. Department of Energy under contract DE-AC05-00OR22725 with UT-Battelle, LLC. Additional support was provided by the Laboratory Directed Research and Development Program of the Oak Ridge National Laboratory. G. H. Paulino acknowledges the support from the National Science Foundation under grant CMS-9996378 (Mechanics and Materials Program). The authors would like to thank Profs. J. Berger and P. A. Martin for useful discussions, and Prof. Luiz Wrobel for pointing out the paper by Li and Evans [17].

Appendix

Symmetric Kernels. The symmetric-Galerkin method ([25–28]) is a highly effective numerical technique for boundary integral analysis. As the name implies, it utilizes the Galerkin approximation to induce a symmetric coefficient matrix. The symmetry comes about because of the symmetry properties of the kernel functions in the integral equations for surface temperature and for surface flux. Note that for the homogeneous Laplace equation, the fundamental solution is symmetric, $G(P,Q)=G(Q,P)$, but the functionally graded material (FGM) Green's function, Eq. (21), is not. Thus it would appear that a symmetric-Galerkin approximation is not possible.

In this section, the FGM boundary integral equations are reformulated to allow a symmetric numerical implementation. In addition, formulas for all of the required FGM kernel functions for $\mathrm{k}(z)$ real,

$$\mathrm{k}(z)=\mathrm{k}_0 e^{2\beta_0 z}, \tag{39}$$

are conveniently summarized.

To obtain a symmetric matrix, the equations have to be written in terms of the surface flux,

$$\mathcal{F}(Q)=-\mathrm{k}(z_Q)\frac{\partial}{\partial n}\phi(Q) \tag{40}$$

rather than the normal derivative. The equation for surface temperature $\phi(P)$ is therefore

$$\phi(P)+\int_\Sigma F(P,Q)\phi(Q)dQ=\int_\Sigma G_S(P,Q)\mathcal{F}(Q)dQ, \quad (41)$$

and in three dimensions the kernel functions are

$$G_S(P,Q)=-\frac{G(P,Q)}{\mathrm{k}(z_Q)}=-\frac{1}{4\mathrm{k}_0\pi}\frac{e^{\beta_0(r-z_Q-z_P)}}{r}$$

$$F(P,Q)=\frac{\partial}{\partial n}G(P,Q)-2\beta_0 n_z G(P,Q)$$

$$=-\frac{e^{\beta_0(r+R_z)}}{4\pi}\left(\frac{\mathbf{n}\cdot R}{r^3}-\beta_0\frac{\mathbf{n}\cdot R}{r^2}+\beta_0\frac{n_z}{r}\right). \quad (42)$$

Most importantly, note that $G_S(P,Q)$, unlike G, is symmetric with respect to P and Q. This is the first of three conditions needed for symmetry. The other two conditions involve the flux equation. Differentiating Eq. (41) with respect to P, dotting with $\mathbf{N}=\mathbf{N}(P)$, and multiplying by $-\mathrm{k}(z_P)$ yields the corresponding equation for surface flux

$$\mathcal{F}(P)+\int_\Sigma W(P,Q)\phi(Q)dQ=\int_\Sigma S(P,Q)\mathcal{F}(Q)dQ. \quad (43)$$

The kernel functions, again for three dimensions, are computed to be

$$S(P,Q)=-\mathrm{k}(z_P)\frac{\partial}{\partial N}G_S(P,Q)$$

$$=-\frac{e^{\beta_0(r-R_z)}}{4\pi}\left(-\frac{\mathbf{N}\cdot R}{r^3}+\beta_0\frac{\mathbf{N}\cdot R}{r^2}+\beta_0\frac{N_z}{r}\right). \quad (44)$$

and

$$W(P,Q)=-\mathrm{k}(z_P)\frac{\partial}{\partial N}F(P,Q)=\frac{\mathrm{k}_0}{4\pi}e^{\beta_0(r+z_Q+z_P)}$$

$$\times\left(3\frac{(\mathbf{n}\cdot R)(\mathbf{N}\cdot R)}{r^5}-3\beta_0\frac{(\mathbf{n}\cdot R)(\mathbf{N}\cdot R)}{r^4}\right.$$

$$+\frac{\beta_0^2(\mathbf{n}\cdot R)(\mathbf{N}\cdot R)-\beta_0(N_z\mathbf{n}-n_z\mathbf{N})\cdot R-\mathbf{n}\cdot\mathbf{N}}{r^3}$$

$$\left.+\beta_0\frac{\beta_0(N_z\mathbf{n}-n_z\mathbf{N})\cdot R+\mathbf{n}\cdot\mathbf{N}}{r^2}-\beta_0^2\frac{N_z n_z}{r}\right). \quad (45)$$

The additional symmetry requirements are that W must be symmetric, $W(P,Q)=W(Q,P)$, and that $S(P,Q)=F(Q,P)$. Interchanging Q and P implies replacing $\mathbf{N}(P)$ with $\mathbf{n}(Q)$ and changes the sign of R, and thus both conditions are seen to hold.

References

[1] Hirai, T., 1996, "Functionally Graded Materials," *Materials Science and Technology*, R. J. Brook, ed., Vol. 17B of *Processing of Ceramics*, Part 2, VCH Verlagsgesellschaft mbH, Weinheim, Germany, pp. 292–341.

[2] Markworth, A. J., Ramesh, K. S., and Parks, W. P. Jr., 1995, "Modelling Studies Applied to Functionally Graded Materials," J. Mater. Sci., **30**, pp. 2183–2193.

[3] Paulino, G. H., Jin, Z.-H., and Dodds, R. H., 2003, "Failure of Functionally Graded Materials," *Comprehensive Structural Integrity*, B. Karihaloo, R. Ritchie and I. Milne, eds., **2**, Elsevier, New York, Chap. 13.

[4] Suresh, S., and Mortensen, A., 1998, *Fundamentals of Functionally Graded Materials*, The Institute of Materials, IOM Communications Ltd., London.

[5] Bonnet, M., 1995, *Boundary Integral Equation Methods for Solids and Fluids*, John Wiley and Sons, New York.

[6] Barton, G., 1999, *Elements of Green's Functions and Propagation*, Oxford University Press, Oxford, UK.

[7] Ang, W. T., Kusuma, J., and Clements, D. L., 1996, "A Boundary Element Method for a Second Order Elliptic Partial Differential Equation With Variable Coefficients," Eng. Anal. Boundary Elem., **18**, pp. 311–316.

[8] Clements, D. L., 1980, "A Boundary Integral Equation Method for the Numerical Solution of a Second Order Elliptic Equation With Variable Coefficients," J. Aust. Math. Soc. B, Appl. Math., **22**, pp. 218–228.

[9] Shaw, R. P., 1994, "Green's Functions for Heterogeneous Media Potential Problems," Eng. Anal. Boundary Elem., **13**, pp. 219–221.

[10] Shaw, R. P., and Makris, N., 1992, "Green's Functions for Helmholtz and Laplace Equations in Heterogeneous Media," Eng. Anal. Boundary Elem., **10**, pp. 179–183.

[11] Shaw, R. P., and Manolis, G. D., 2000, "A Generalized Helmholtz Equation Fundamental Solution Using a Conformal Mapping and Dependent Variable Transformation," Eng. Anal. Boundary Elem., **24**, pp. 177–188.

[12] Divo, E., and Kassab, A. J., 1998, "Generalized Boundary Integral Equation for Heat Conduction in Non-homogeneous Media: Recent Developments on the Sifting Property," Eng. Anal. Boundary Elem., **22**, pp. 221–234.

[13] Kassab, A. J., and Divo, E., 1996, "A Generalized Boundary Integral Equation for Isotropic Heat Conduction With Spatially Varying Thermal Conductivity," Eng. Anal. Boundary Elem., **18**, pp. 273–286.

[14] Bonnet, M., and Guiggiani, M., 1998, Comments about the paper entitled "A Generalized Boundary Integral Equation for Isotropic Heat Conduction With Spatially Varying Thermal Conductivity," by A. J. Kassab and E. Divo, Eng. Anal. Boundary Elem., **22**, pp. 235–240.

[15] Power, H., 1997, "On the Existence of Kassab and Divo's Generalized Boundary Integral Equation Formulation for Isotropic Heterogeneous Steady State Heat Conduction Problems," Eng. Anal. Boundary Elem., **20**, pp. 341–345.

[16] Vrettos, C., 1991, "Surface Green's Functions for Continuously Nonhomogeneous Soil," *Computer Methods and Advances in Geomechanics*, Beer, Booker, and Carter, eds., Rotterdam, Balkema, pp. 801–804.

[17] Li, B. Q., and Evans, J. W., 1991, "Boundary Element Solution of Heat Convection-Diffusion Problems," J. Comput. Phys., **93**, pp. 255–272.

[18] Wrobel, L. C., personal communication.

[19] Shaw, R. P., and Gipson, G. S., 1995, "Interrelated Fundamental Solutions for Various Heterogeneous Potential, Wave and Advective-Diffusive Problems," Eng. Anal. Boundary Elem., **16**, pp. 29–34.

[20] Martin, P. A., Richardson, J. D., Gray, L. J., and Berger, J., 2002, "On Green's Function for a Three-Dimensional Exponentially-Graded Elastic Solid," Proc. Royal Soc. London A, in press.

[21] Kellogg, O. D., 1953, *Foundations of Potential Theory*, Dover, New York.

[22] Olver, F. W. J., 1972, "Bessel Functions of Integer Order," *Handbook of Mathematical Functions*, M. Abromowitz and I. A. Stegun, eds., National Bureau of Standards, Washington, D.C., Chap. 9, pp. 355–434.

[23] Konda, N., and Erdogan, F., 1994, "The Mixed Mode Crack Problem in a Nonhomogeneous Elastic Medium," Eng. Fract. Mech., **47**, pp. 533–545.

[24] Bonnet, M., Maier, G., and Polizzotto, C., 1998, "Symmetric Galerkin Boundary Element Method," ASME Appl. Mech. Rev., **51**, pp. 669–704.

[25] Hartmann, F., Katz, C., and Protopsaltis, B., 1985, "Boundary Elements and Symmetry," Ing.-Arch., **55**, pp. 440–449.

[26] Maier, G., Diligenti, M., and Carini, A., 1991, "A Variational Approach to Boundary Element Elastodynamic Analysis and Extension to Multidomain Problems," Comp. Meth. Appl. Eng., **92**, pp. 193–213.

[27] Sirtori, S., 1979, "General Stress Analysis Method by Means of Integral Equations and Boundary Elements," Meccanica, **14**, pp. 210–218.

[28] Sirtori, S., Maier, G., Novati, G., and Miccoli, S., 1992, "A Galerkin Symmetric Boundary Element Method in Elasticity: Formulation and Implementation," Int. J. Numer. Methods Eng., **35**, pp. 255–282.

[29] Kim, J.-H., and Paulino, G. H., 2002, "Finite Element Evaluation of Mixed Mode Stress Intensity Factors in Functionally Graded Materials," Int. J. Numer. Methods Eng., **53**, pp. 1903–1935.

[30] Sutradhar, A., Paulino, G. H., and Gray, L. J., 2002, "Transient Heat Conduction in Homogeneous and Non-homogeneous Materials by the Laplace Transform Galerkin Boundary Element Method," Eng. Anal. Boundary Elem., **26**, pp. 119–132.

T. L. Yip

C. O. Ng[1]
Mem. ASME
e-mail: cong@hku.hk

Department of Mechanical Engineering,
The University of Hong Kong,
Pokfulam Road,
Hong Kong

Dual-Species Transport Subject to Sorptive Exchange in Pipe Flow

The transport in pipe flow of a chemical species can be materially affected by the presence of solid suspension if the species is capable of partitioning into a solute phase and a solid phase sorbed onto the suspended particles. An asymptotic analysis is used in this work to deduce the effective transport equations for the two phases, with kinetic sorptive exchange taken into account. The effects of sorption on the advection and dispersion of a sorbing chemical are discussed and illustrated with a numerical example.
[DOI: 10.1115/1.1576805]

1 Introduction

The concept of dispersion, first introduced by Taylor [1,2], is required to account for the longitudinal spreading of substances such as pollutants in a pipe. The spreading is at a rate much faster than would result from diffusion (molecular or turbulent) alone. The Taylor dispersion mechanism is one in which the transverse variation of longitudinal velocity and transverse diffusion interact to result in an overall longitudinal mixing process that appears Fickian. This enables the effective transport in a pipe to be described by a one-dimensional advection-dispersion equation, in which the dispersion coefficient is expressible in terms of the flow and mixing characteristics.

Taylor [2] pioneered the work on dispersion in turbulent flow through a pipe. Since then, the problem has been advanced considerably. Nevertheless, existing analyses are mostly concerned with passive or chemically inert dispersants; a relatively fewer investigations have studied the phase change effects on the transport processes. Some authors (e.g., Smith [3,4] and Purnama [5]) have demonstrated the boundary absorption/retention effects on dispersion in shear flow. Here we intend to show that the presence of suspended particles also can have nontrivial effects on the dispersion of a sorbing solute in turbulent pipe flow.

Sorption is a chemical process that may be in the form of adsorption, chemisorption, ion exchange, or absorption. It is a reaction by which a chemical species is partitioned into a solute phase that is miscible with the fluid and a solid phase that is sorbed onto some solid matter in contact with the fluid. The phase exchange is often reversible; the chemical changes forward and backward between the two phases depending on the concentration difference. Since typically the solid phase is immobile or moves more slowly than the fluid, such phase partitioning will lead to a diminished advection speed, or retardation, of the chemical in the flow. Other processes like biodegradation, radioactive decay, and precipitation will reduce the concentration but may not slow down the movement of the chemical as effectively as sorption. In this connection, sorption is accountable for the separation of the clouds of two miscible chemicals if they have vastly different affinities for a particular type of solid matter, which may exist on the pipe wall or suspended particles. As multispecies transport is very common in practice, it is of practical value to look into this and other effects of sorption on the transport; this motivates the present study.

Our specific objective here is to study the effects of sorptive exchange on the transport of a sorbing solute in a pipe carrying turbulent flow laden with suspended particles. To this end, we employ an asymptotic method to deduce the effective (i.e., section-averaged) transport equations for the solute and the sorbed phases. The starting point is the conventional advection-diffusion equation, in which the diffusive flux is assumed to be linearly proportional to the concentration gradient. The turbulent diffusion coefficient, or eddy diffusivity, is a function of the flow. Despite being phenomenological, such a classical approach in modeling turbulent mixing is still greatly favored in engineering applications nowadays. More advanced methods, such as the Monte Carlo techniques based on the statistical theory, have met some success in some specific applications. These techniques are however typically computationally intensive and unwieldy for practical use. By contrast, the method applied here, as presented below, can yield closed-form analytical expressions for the transport coefficients, thereby facilitating the discussion on the physical effects.

Two previous works by the authors provide the basis for the present study. In Ng and Yip [6], the transport in open-channel flow of a chemical species under the influence of kinetic sorptive exchange was investigated. It was demonstrated that the sorption kinetics can have nontrivial effects on the advection and dispersion of a solute cloud, which are found to be functions of space and time depending on the local concentration of solid suspension. In Ng [7], the classical problem of longitudinal dispersion of passive heavy particles in turbulent pipe flow was re-examined. On pointing out that Elder's, [8], theory is defective, a formal expression for the dispersion coefficient has been deduced with a systematic asymptotic analysis. The present work, which is further described below, is essentially an extension of these two previous works.

2 Problem Formulation

Consider radially symmetric steady flow in a horizontal circular pipe of radius a. Cylindrical coordinates (x,r,θ) are defined such that x points downstream along the pipe axis, r is the radial distance from the axis, and θ is the angle measured clockwise, viewed downstream, from the vertical. The flow carries a chemical species as well as solid particles in suspension. The chemical exists primarily as a solute phase miscible with the fluid, and a solid phase sorbed onto the particles. The particle radius is denoted by αa, where $\alpha<1$ is the ratio of the radius of the particles to that of the pipe. Because of their small yet finite size, the particles can have access only to a cross section of the pipe of radius $(1-\alpha)a$. It is assumed that the presence of the particles and the solute do not materially affect the flow. The solid particles settle under gravity. The fall velocity w_f is assumed to be constant. Turbulence in the flow is strong enough to entrain any deposited particles, and the particles stay in suspension mostly all the time. The particles are also assumed to contain high amounts

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Sept. 6, 2001; final revision, Oct. 24, 2002. Associate Editor: D. A. Siginer. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

of sorbents, which are materials such as organic matter providing a site onto which the chemical can be sorbed. The fractions of mass of the chemical in the two phases can therefore be comparable with each other. For analytical simplicity, we further introduce the approximation that the eddy diffusivity is constant and isotropic, and is the same for the particles and the solute.

Conservation of mass gives the following transport equation for the suspended particles:

$$\frac{\partial \zeta}{\partial t}+u\frac{\partial \zeta}{\partial x}-\frac{1}{r}\frac{\partial}{\partial r}(rw_f\cos\theta\zeta)+\frac{1}{r}\frac{\partial}{\partial\theta}(w_f\sin\theta\zeta)$$
$$=\frac{\partial}{\partial x}\left(E\frac{\partial \zeta}{\partial x}\right)+\frac{E}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \zeta}{\partial r}\right)+\frac{E}{r^2}\frac{\partial^2\zeta}{\partial\theta^2}, \tag{1}$$

where t is the time, $\zeta(x,r,\theta,t)$ is the particle concentration (mass of suspended particles per bulk volume), $u(r)$ is the time-smoothed fluid velocity in the axial direction, and E is the eddy diffusivity. The net flux vanishes on the pipe wall, so the boundary conditions are

$$w_f\cos\theta\zeta+E\frac{\partial \zeta}{\partial r}=0 \quad \text{at } r=(1-\alpha)a. \tag{2}$$

For the chemical that is partitioned between a dissolved phase and a sorbed phase, the total mass concentration $C_{\text{tot}}(x,r,\theta,t)$ (total mass of chemical per bulk volume) is

$$C_{\text{tot}}=C+C_s\zeta, \tag{3}$$

where $C(x,r,\theta,t)$ is the solute concentration (mass of dissolved phase per bulk volume) and $C_s(x,r,\theta,t)$ is the sorbate concentration (mass of solid phase sorbed onto unit mass of suspended particles). The transport of the chemical is governed by

$$\frac{\partial C_{\text{tot}}}{\partial t}+u\frac{\partial C_{\text{tot}}}{\partial x}-\frac{1}{r}\frac{\partial}{\partial r}(rw_f\cos\theta C_s\zeta)+\frac{1}{r}\frac{\partial}{\partial\theta}(w_f\sin\theta C_s\zeta)$$
$$=\frac{\partial}{\partial x}\left(E\frac{\partial C_{\text{tot}}}{\partial x}\right)+\frac{E}{r}\frac{\partial}{\partial r}\left(r\frac{\partial C_{\text{tot}}}{\partial r}\right)+\frac{E}{r^2}\frac{\partial^2 C_{\text{tot}}}{\partial\theta^2}, \tag{4}$$

and the zero-flux boundary conditions

$$w_f\cos\theta C_s\zeta+E\frac{\partial C_{\text{tot}}}{\partial r}=0 \quad \text{at } r=(1-\alpha)a. \tag{5}$$

The range $0\leqslant\theta\leqslant 2\pi$ for one period of phase angle is considered. By continuity, all concentrations and their derivatives have the same values at the two limits of θ; that is,

$$(\zeta,C,\ldots)_{\theta=0}=(\zeta,C,\ldots)_{\theta=2\pi}. \tag{6}$$

Sorption is a reversible reaction between the dissolved and sorbed phases of the chemical. The overall sorption rate can be described by a linear kinetic sorption model

$$\frac{\partial C_s}{\partial t}=k_fC-kC_s, \tag{7}$$

where k_f and k are, respectively, the forward (sorption) and backward (desorption) rate constants for the sorption reaction. When the steady state is attained or the reaction is fast compared with other processes, the two phases will be in chemical equilibrium. Then the ratio of their concentrations will be given by

$$(C_s/C)_{\text{equilib}}=k_f/k\equiv K_d, \tag{8}$$

where K_d, defined as the sorption partition coefficient, is a ratio of the forward and backward rate constants for the processes. The value of this coefficient tends to be large for a chemical which has a strong affinity for the solid matter. On substituting Eq. (8) into Eq. (7), the first-order kinetics can be written as

$$\frac{\partial C_s}{\partial t}=k(K_dC-C_s), \tag{9}$$

or the rate of change of the sorbate concentration is linearly proportional to the departure from local equilibrium. The backward rate constant k will be simply referred to as the sorption rate constant.

For the perturbation analysis, the order of magnitude of individual terms in the governing equations and boundary conditions (1), (2), (4), (5), and (9) need be estimated. A similar exercise has been carried out previously by Ng and Yip [6] for transport of a sorbing solute in open-channel flow. Citing their results, we may express these equations below, with a small ordering parameter $\epsilon\ll 1$ inserted to indicate the relative order of the terms. The small parameter ϵ can be taken as the ratio of the turbulent diffusivity and longitudinal dispersion coefficient.

Now, the particle transport equation is

$$\epsilon\frac{\partial \zeta}{\partial t}+\epsilon u\frac{\partial \zeta}{\partial x}-\frac{1}{r}\frac{\partial}{\partial r}(rw_f\cos\theta\zeta)+\frac{1}{r}\frac{\partial}{\partial\theta}(w_f\sin\theta\zeta)$$
$$=\epsilon^2E\frac{\partial^2\zeta}{\partial x^2}+\frac{E}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \zeta}{\partial r}\right)+\frac{E}{r^2}\frac{\partial^2\zeta}{\partial\theta^2} \text{ in } 0<r<(1-\alpha)a,$$
$$0<\theta<2\pi, \quad -\infty<x<\infty, \quad t>0, \tag{10}$$

with the boundary condition

$$w_f\cos\theta\zeta+E\frac{\partial \zeta}{\partial r}=0 \quad \text{at } r=(1-\alpha)a. \tag{11}$$

The chemical transport equation is

$$\epsilon\frac{\partial C_{\text{tot}}}{\partial t}+\epsilon u\frac{\partial C_{\text{tot}}}{\partial x}-\frac{1}{r}\frac{\partial}{\partial r}(rw_f\cos\theta C_s\zeta)+\frac{1}{r}\frac{\partial}{\partial\theta}(w_f\sin\theta C_s\zeta)$$
$$=\epsilon^2E\frac{\partial^2 C_{\text{tot}}}{\partial x^2}+\frac{E}{r}\frac{\partial}{\partial r}\left(r\frac{\partial C_{\text{tot}}}{\partial r}\right)+\frac{E}{r^2}\frac{\partial^2 C_{\text{tot}}}{\partial\theta^2}$$
$$\text{in } 0<r<(1-\alpha)a, \quad 0<\theta<2\pi, \quad -\infty<x<\infty, \quad t>0, \tag{12}$$

with the boundary condition

$$w_f\cos\theta C_s\zeta+E\frac{\partial C_{\text{tot}}}{\partial r}=0 \quad \text{at } r=(1-\alpha)a. \tag{13}$$

The sorption kinetics equation is

$$\epsilon\frac{\partial C_s}{\partial t}=k(K_dC-C_s). \tag{14}$$

Two pertinent time scales are to be accounted for in the present problem: $T_1=L/\overline{u}$ for advection along the pipe, and $T_2=L^2/\mathcal{D}=O(\epsilon^{-1}T_1)$ for longitudinal dispersion along the pipe, where $\overline{u}$ the discharge velocity, E the eddy diffusivity, and $\mathcal{D}$ the dispersion coefficient. It is noted that the longitudinal dispersion is effective only at a time scale one order of magnitude longer than that for the advection.

Perturbation equations are now obtainable when the following multiple-scale expansions of the variables are substituted into Eqs. (10)–(14):

$$(\zeta,C,C_s)\rightarrow(\zeta_0,C_0,C_{s0})+\epsilon(\zeta_1,C_1,C_{s1})+\epsilon^2(\zeta_2,C_2,C_{s2})+O(\epsilon^3), \tag{15}$$

$$\partial/\partial t\rightarrow\partial/\partial t_1+\varepsilon\partial/\partial t_2. \tag{16}$$

In the following sections, we will denote the cross-sectional average by an overbar. For example, for any function $g(r,\theta)$,

$$\overline{g}\equiv\frac{1}{\pi(1-\alpha)^2a^2}\int_0^{2\pi}\int_0^{(1-\alpha)a}grdrd\theta. \tag{17}$$

3 Transport of Suspended Particles

The leading-order $O(1)$ particle concentration ζ_0 can be written as

$$\zeta_0(x,r,\theta,t)=f(r,\theta)\overline{\zeta}(x,t), \tag{18}$$

where $\overline{\zeta}$ is the cross-sectional-average particle concentration, and $f(r,\theta)$ is the probability density function or particle distribution function given by

$$f(r,\theta)=f_0\exp\left(-\frac{w_f}{E}r\cos\theta\right), \tag{19}$$

which results from a balance between turbulent mixing and particle settling. The constant f_0 is determined by the condition that the average of the distribution function f over the effective cross section of radius $(1-\alpha)a$ is equal to unity (i.e., $\overline{f}=1$):

$$f_0=\frac{\delta(1-\alpha)}{2I_1[\delta(1-\alpha)]}, \tag{20}$$

where

$$\delta=w_f a/E, \tag{21}$$

is a form of the suspension number and I_1 is the modified Bessel function of the first kind of order one. The particles will be fine enough to stay in suspension largely all the time; this is realistic when δ is sufficiently small (Sumer [9]).

At $O(\epsilon)$, the perturbation equation after substituting Eq. (18) is

$$\frac{\partial\zeta_0}{\partial t_1}+u\frac{\partial\zeta_0}{\partial x}-\frac{1}{r}\frac{\partial}{\partial r}(rw_f\cos\theta\zeta_1)+\frac{1}{r}\frac{\partial}{\partial\theta}(w_f\sin\theta\zeta_1)=\frac{E}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\zeta_1}{\partial r}\right)+\frac{E}{r^2}\frac{\partial^2\zeta_1}{\partial\theta^2}, \tag{22}$$

and the boundary condition is

$$w_f\cos\theta\zeta_1+E\frac{\partial\zeta_1}{\partial r}=0 \quad \text{at } r=(1-\alpha)a. \tag{23}$$

On taking an average over the cross section and using the boundary conditions (23) and (6), Eq. (22) yields the leading order transport equation for the suspended particles

$$\frac{\partial\overline{\zeta}}{\partial t_1}+u_s\frac{\partial\overline{\zeta}}{\partial x}=0, \tag{24}$$

where

$$u_s=\overline{fu} \tag{25}$$

is the effective advection velocity of the particles, which is a cross-sectional-average velocity weighted by the particle distribution factor f. As expected, only advection appears in this leading-order equation.

One can easily show that

$$\overline{u}-u_s=-\overline{(f-1)(u-\overline{u})}. \tag{26}$$

The difference between the discharge velocity and the effective advection speed of the particles is related to the cross-sectional covariance between the particle concentration and the fluid velocity, which in general is negative. Hence, the particles are advected at a speed slower than the discharge velocity.

On eliminating the unsteady term from Eqs. (24) and (22), we get

$$f(u-u_s)\frac{\partial\overline{\zeta}}{\partial x}=\frac{1}{r}\frac{\partial}{\partial r}\left[r\left(w_f\cos\theta\zeta_1+E\frac{\partial\zeta_1}{\partial r}\right)\right]+\frac{1}{r}\frac{\partial}{\partial\theta}\left[-w_f\sin\theta\zeta_1+\frac{E}{r}\frac{\partial\zeta_1}{\partial\theta}\right]. \tag{27}$$

It follows from Eqs. (27) and (23) that the following form for the first-order particle concentration $\zeta_1(x,r,\theta,t)$ can be suggested:

$$\zeta_1(x,r,\theta,t)=N\frac{\partial\zeta_0}{\partial x}=Nf\frac{\partial\overline{\zeta}}{\partial x}, \tag{28}$$

where $N=N(r,\theta)$ is a cell function that describes the cross-sectional variation for the $O(\epsilon)$ correction to the particle concentration and is governed by

$$f(u-u_s)=\frac{1}{r}\frac{\partial}{\partial r}\left[r\left(w_f\cos\theta fN+E\frac{\partial fN}{\partial r}\right)\right]+\frac{1}{r}\frac{\partial}{\partial\theta}\left[-w_f\sin\theta fN+\frac{E}{r}\frac{\partial fN}{\partial\theta}\right] \quad \text{in } 0<r<(1-\alpha)a, \quad 0<\theta<2\pi, \tag{29}$$

with the boundary conditions

$$w_f\cos\theta fN+E\frac{\partial fN}{\partial r}=0 \quad \text{at } r=(1-\alpha)a, \tag{30}$$

$$N \text{ is finite} \quad \text{at } r=0. \tag{31}$$

At $O(\epsilon^2)$, the perturbation equation is

$$\frac{\partial\zeta_0}{\partial t_2}+\frac{\partial\zeta_1}{\partial t_1}+u\frac{\partial\zeta_1}{\partial x}-\frac{1}{r}\frac{\partial}{\partial r}(rw_f\cos\theta\zeta_1)+\frac{1}{r}\frac{\partial}{\partial\theta}(w_f\sin\theta\zeta_1)=E\frac{\partial^2\zeta_0}{\partial x^2}+\frac{E}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\zeta_2}{\partial r}\right)+\frac{E}{r^2}\frac{\partial^2\zeta_2}{\partial\theta^2}, \tag{32}$$

and the boundary condition is

$$w_f\cos\theta\zeta_2+E\frac{\partial\zeta_2}{\partial r}=0 \quad \text{at } r=(1-\alpha)a. \tag{33}$$

Using the boundary conditions (33) and (6), the cross-sectional average of Eq. (32) gives

$$\frac{\partial\overline{\zeta}}{\partial t_2}+\frac{\partial\overline{\zeta_1}}{\partial t_1}+u\overline{\frac{\partial\zeta_1}{\partial x}}=E\frac{\partial^2\overline{\zeta}}{\partial x^2}. \tag{34}$$

Further substituting (28) for ζ_1 and (24) for $\partial\overline{\zeta}/\partial t_1$, the $O(\epsilon^2)$ effective equation becomes

$$\frac{\partial\overline{\zeta}}{\partial t_2}=\mathcal{D}_s\frac{\partial^2\overline{\zeta}}{\partial x^2}, \tag{35}$$

where the diffusion coefficient $\mathcal{D}_s$ is

$$\mathcal{D}_s=E+\mathcal{D}_{Ts}, \tag{36}$$

in which the first component is the longitudinal eddy diffusivity E and the second component is a Taylor dispersion coefficient $\mathcal{D}_{Ts}$ formally given by

$$\mathcal{D}_{Ts}=-\overline{Nf(u-u_s)}, \tag{37}$$

which will be determined in Section 5. Typically $\mathcal{D}_{Ts}\gg E$, or the dispersion dominates over the turbulent mixing in controlling the spreading of the particles along the pipe.

Finally, we may combine Eqs. (24) and (35) to get the effective transport equation for the particle concentration $\overline{\zeta}(x,t)$, which is correct to $O(\epsilon)$:

$$\frac{\partial\overline{\zeta}}{\partial t}+u_s\frac{\partial\overline{\zeta}}{\partial x}=\mathcal{D}_s\frac{\partial^2\overline{\zeta}}{\partial x^2}. \tag{38}$$

4 Transport of a Sorbing Chemical Species

The preceding procedures of deduction are now applied to the transport of a sorbing chemical. Without repeating the steps in

detail, only the key results are presented below. From Eq. (14), the phase partitioning between the dissolved and sorbed phases is in local equilibrium only at the leading order:

$$C_{s0}=K_dC_0, \tag{39}$$

where C_{s0} and C_0 are, respectively, the leading-order sorbate and solute concentrations. Hence the $O(1)$ total solute concentration can be written as

$$C_{\text{tot }0}=C_0(1+K_d\zeta_0)=RC_0, \tag{40}$$

where

$$R(x,r,\theta,t)=1+K_d\zeta_0=1+K_d\overline{\zeta}(x,t)f(r,\theta)\geqslant 1 \tag{41}$$

is the retardation factor resulting from the phase partitioning. Physically the retardation factor is the ratio of the total concentration to the solute concentration, and reflects the extent of sorption.

The $O(1)$ solute transport Eq. (12) and boundary condition (13) give that C_0 is independent of r and θ, or

$$C_0=C_0(x,t) \quad \text{and} \quad C_{s0}=C_{s0}(x,t)=K_dC_0(x,t). \tag{42}$$

At the leading order, the solute and sorbate concentrations are locally uniformly distributed across the section of pipe.

At $O(\epsilon)$, Eq. (14) gives

$$\frac{\partial C_{s0}}{\partial t}=k(K_dC_1-C_{s1}). \tag{43}$$

Hence the rate of change of the $O(1)$ sorbate concentration is driven by the departure from local equilibrium between the $O(\epsilon)$ concentrations. The leading-order transport equation for the solute is obtained after taking the cross-sectional average of the $O(\epsilon)$ terms of Eq. (12):

$$\frac{\partial C_0}{\partial t_1}+u_c\frac{\partial C_0}{\partial x}=0, \tag{44}$$

where $C_0(x,t)$ is the leading-order solute concentration, and

$$u_c(x,t)=\overline{Ru}/\overline{R} \tag{45}$$

is the leading-order effective advection velocity for the dissolved phase. This is essentially an average velocity weighted by the retardation factor R. One may infer from Eqs. (24) and (44) that the particles and the solute are in general advected at different speeds because

$$u_c=u_s+(\overline{u}-u_s)/\overline{R}=\overline{u}-(\overline{R}-1)(\overline{u}-u_s)/\overline{R}. \tag{46}$$

The above relations give that the solute moves effectively faster than the particles, but will be slower than the discharge velocity under the influence of sorption. If the chemical is nonsorbing or the particle concentration is zero, R becomes unity and $u_c=u$.

Similar to ζ_1 in Eq. (28), the following form for C_1 can be suggested:

$$C_1(x,r,\theta,t)=P\frac{\partial C_0}{\partial x}. \tag{47}$$

The function $P(x,r,\theta,t)$, which is the $O(\epsilon)$ correction of the solute concentration, is given by

$$\frac{E}{r}\frac{\partial}{\partial r}\left(rR\frac{\partial P}{\partial r}\right)+\frac{1}{r}\frac{\partial}{\partial\theta}\left(\frac{ER}{r}\frac{\partial P}{\partial\theta}\right)=R(u-u_c) \text{ in } 0<r<(1-\alpha)a, \quad 0<\theta<2\pi, \tag{48}$$

with the boundary conditions

$$R\frac{\partial P}{\partial r}=0, \quad \text{at } r=(1-\alpha)a, \tag{49}$$

$$P \text{ is finite at } r=0. \tag{50}$$

Now, we may obtain from Eqs. (28), (39), (43), (44), and (47) the following expressions for the $O(\epsilon)$ correction to the sorbate concentration and the total chemical concentration:

$$C_{s1}(x,r,\theta,t)=K_d\left(P+\frac{u_c}{k}\right)\frac{\partial C_0}{\partial x}, \tag{51}$$

and

$$C_{\text{tot }1}=\left(RP+\frac{K_du_c}{k}\zeta_0\right)\frac{\partial C_0}{\partial x}+K_dNf\frac{\partial\overline{\zeta}}{\partial x}C_0. \tag{52}$$

In order to ensure that $\overline{C_{\text{tot }0}}$ and $\overline{\zeta_0}$ are correct to $O(\epsilon)$, or their first-order corrections are zero: $\overline{C_{\text{tot }1}}=\overline{\zeta_1}=0$, we add the condition that

$$\overline{C_{\text{tot }1}}=0, \tag{53}$$

which can be satisfied if, evident from Eq. (52),

$$\overline{RP+K_d\zeta_0u_c/k}=0 \quad \text{and} \quad \overline{Nf}=0. \tag{54}$$

Conditions similar to (54) have been applied by Chatwin [10], Mei [11], and Ng and Yip [6].

Subsequently, the $O(\epsilon^2)$ solute transport equation becomes

$$\frac{\partial C_0}{\partial t_2}+u_c'\frac{\partial C_0}{\partial x}=\frac{\partial}{\partial x}\left(\mathcal{D}_c\frac{\partial C_0}{\partial x}\right), \tag{55}$$

where

$$u_c'(x,t)=-\frac{K_d}{R}(\mathcal{D}_s+\mathcal{D}_c)\frac{\partial\overline{\zeta}}{\partial x} \tag{56}$$

is the $O(\epsilon)$ correction to the advection velocity in which $\mathcal{D}_s$ is the dispersion coefficient for the particles as given by Eq. (36), and

$$\mathcal{D}_c(x,t)=E+\mathcal{D}_{Tc}+\mathcal{D}_{Kc} \tag{57}$$

is the effective dispersion coefficient for the solute. Similar to $\mathcal{D}_s$, the first component of $\mathcal{D}_c$ is the longitudinal eddy diffusivity E, the second component is the Taylor dispersion coefficient

$$\mathcal{D}_{Tc}(x,t)=-\overline{RP(u-u_c)}/\overline{R}, \tag{58}$$

and the third component, $\mathcal{D}_{Kc}$, is the sorption-kinetics-induced dispersion coefficient formally given by

$$\mathcal{D}_{Kc}(x,t)=-\frac{K_d\overline{\zeta}u_c}{k\overline{R}^2}\overline{(f-1)(u-\overline{u})}, \tag{59}$$

which is first obtained by Ng [12].

On combining Eqs. (44) and (55), we get an effective transport equation for the chemical with an error of $O(\epsilon^2)$:

$$\frac{\partial C_0}{\partial t}+(u_c+u_c')\frac{\partial C_0}{\partial x}=\frac{\partial}{\partial x}\left(\mathcal{D}_c\frac{\partial C_0}{\partial x}\right). \tag{60}$$

It is remarkable that despite the strictly uniform flow both the advection velocity and dispersion coefficient are functions of x and t through dependence on the local particle concentration. They are also functions of the hydrodynamic and sorption effects.

5 Taylor Dispersion Coefficients

The Taylor dispersion coefficients $\mathcal{D}_{Ts}$ and $\mathcal{D}_{Tc}$, formally given by Eqs. (37) and (58), rely on the functions N and P, which are governed by the boundary value problems (29)–(31) and (48)–(50). These two problems can be solved, at least numerically, for any given velocity profile $u(r)$.

The problem (29)–(31) can be simplified to

$$\hat{\nabla}\cdot(f\hat{\nabla}N)\equiv\frac{1}{\hat{r}}\frac{\partial}{\partial\hat{r}}\left(\hat{r}f\frac{\partial N}{\partial\hat{r}}\right)+\frac{1}{\hat{r}^2}\frac{\partial}{\partial\theta}\left(f\frac{\partial N}{\partial\theta}\right)=\frac{fa}{\hat{E}}(\hat{u}-\hat{u}_s) \text{ in } 0<\hat{r}<1-\alpha, \quad 0<\theta<2\pi, \tag{61}$$

$$\frac{\partial N}{\partial \hat{r}}=0 \quad \text{at } \hat{r}=1-\alpha, \tag{62}$$

and

$$N \text{ is finite} \quad \text{at } \hat{r}=0, \tag{63}$$

where the following normalized quantities (distinguished by a caret) have been introduced:

$$\hat{r}=r/a, \quad \hat{u}=u/u_*, \quad \hat{E}=E/u_* a, \tag{64}$$

in which u_* is the shear velocity, which equals the square root of the wall shear stress divided by the fluid density. The function N can be called a cell function, since it governs the cross-sectional variations for the first-order correction to the mean concentration. Using Eqs. (61) and (62), one can show that the dispersion coefficient $\mathcal{D}_{Ts}$ is always positive:

$$\hat{\mathcal{D}}_{Ts} \equiv \mathcal{D}_{Ts}/u_* a = -\overline{Nf(\hat{u}-\hat{u}_s)}/a = \hat{E}\overline{f|\hat{\nabla}N|^2}/a^2 > 0. \tag{65}$$

To seek a solution for N, we first expand the particle distribution function f, defined by Eq. (19), into the following series (Abramowitz and Stegun [[13], p. 376]):

$$f=f_0 \exp(-\delta\hat{r}\cos\theta)=f_0\left[I_0(-\delta\hat{r})+2\sum_{k=1}^{\infty} I_k(-\delta\hat{r})\cos k\theta\right], \tag{66}$$

where f_0 and δ are, respectively, given by Eqs. (20) and (21), and I_k $(k=0,1,2,\ldots)$ is the modified Bessel function of the first kind of order k. For simplicity, we shall from here on unless stated otherwise write the modified Bessel functions without specifying their argument, which is understood to be $(-\delta\hat{r})$. If the function $N(\hat{r},\theta)$, which should also be an even function of θ, is represented by its Fourier series:

$$N(\hat{r},\theta)=a\sum_{k=0}^{\infty} N_k(\hat{r})\cos k\theta, \tag{67}$$

the dispersion coefficient can then be found from Eq. (65):

$$\hat{\mathcal{D}}_{Ts}=-\frac{2f_0}{(1-\alpha)^2}\int_0^{1-\alpha}\left(\sum_{k=0}^{\infty} I_k N_k\right)(\hat{u}-\hat{u}_s)\hat{r}\,d\hat{r}. \tag{68}$$

A differential equation for the function $N_k(\hat{r})$ can be derived by plugging Eqs. (66) and (67) into (61) followed by matching terms of the k-harmonics. For this work, we shall be content with a solution containing up to the third harmonics. Hence, omitting terms of higher harmonics, one may get after some algebra

$$\frac{1}{\hat{r}}\frac{\mathrm{d}}{d\hat{r}}\left[\hat{r}I_0\frac{dN_0}{d\hat{r}}\right]=\frac{I_0}{\hat{E}}(\hat{u}-\hat{u}_s)-\frac{1}{\hat{r}}\frac{d}{d\hat{r}}\left[\hat{r}\left(I_1\frac{dN_1}{d\hat{r}}+I_2\frac{dN_2}{d\hat{r}}+I_3\frac{dN_3}{d\hat{r}}\right)\right], \tag{69}$$

$$\frac{1}{\hat{r}}\frac{d}{d\hat{r}}\left[\hat{r}(I_0+I_2)\frac{dN_1}{d\hat{r}}\right]-\frac{(I_0-I_2)N_1}{\hat{r}^2}=\frac{2I_1}{\hat{E}}(\hat{u}-\hat{u}_s)-\frac{1}{\hat{r}}\frac{\mathrm{d}}{d\hat{r}}\left[\hat{r}\left(2I_1\frac{dN_0}{d\hat{r}}+(I_1+I_3)\frac{dN_2}{d\hat{r}}+I_2\frac{dN_3}{d\hat{r}}\right)\right]+\frac{1}{\hat{r}^2}[2(I_1-I_3)N_2+3I_2N_3], \tag{70}$$

$$\frac{1}{\hat{r}}\frac{d}{d\hat{r}}\left[\hat{r}I_0\frac{dN_2}{d\hat{r}}\right]-\frac{4I_0N_2}{\hat{r}^2}=\frac{2I_2}{\hat{E}}(\hat{u}-\hat{u}_s)-\frac{1}{\hat{r}}\frac{\mathrm{d}}{d\hat{r}}\left[\hat{r}\left(2I_2\frac{dN_0}{d\hat{r}}+(I_1+I_3)\frac{dN_1}{d\hat{r}}+I_1\frac{dN_3}{d\hat{r}}\right)\right]+\frac{1}{\hat{r}^2}[2(I_1-I_3)N_1+6I_1N_3], \tag{71}$$

and

$$\frac{1}{\hat{r}}\frac{\mathrm{d}}{d\hat{r}}\left[\hat{r}I_0\frac{dN_3}{d\hat{r}}\right]-\frac{9I_0N_3}{\hat{r}^2}=\frac{2I_3}{\hat{E}}(\hat{u}-\hat{u}_s)-\frac{1}{\hat{r}}\frac{d}{d\hat{r}}\left[\hat{r}\left(2I_3\frac{dN_0}{d\hat{r}}+I_2\frac{dN_1}{d\hat{r}}+I_1\frac{dN_2}{d\hat{r}}\right)\right]+\frac{1}{\hat{r}^2}[3I_2N_1+6I_1N_2]. \tag{72}$$

The problem (61)–(63) admits an arbitrary constant of integration. For simplicity, we let N equal zero at $\hat{r}=0$. Then the above equations are subject to the boundary conditions

$$N_k(0)=0, \quad N_k'(1-\alpha)=0 \quad k=0,1,2,3. \tag{73}$$

Equations (69)–(72) are linear but coupled second-order ordinary differential equations, in which δ, α, and $\hat{E}$ are the parameters. As in Elder [8], the following defect law for the velocity distribution is adopted:

$$\hat{u}(0)-\hat{u}(r)=F(\hat{r}), \tag{74}$$

where F is a universal distribution function for smooth or rough turbulent flow in a straight pipe of circular cross-section. Empirical data for this function have been derived by Taylor [[2], Table 1]. With Eq. (74), the velocity deviation from the mean can be computed as follows:

$$\hat{u}(\hat{r})-\hat{u}_s(\hat{r})=\hat{u}(\hat{r})-\overline{f\hat{u}}=\overline{fF}-F(\hat{r}). \tag{75}$$

An iterative numerical scheme has been used to obtain solutions of (69)–(72). In each cycle of iteration, these equations are solved by finite differences, one after the other, for N_0 through N_3, respectively, where on the right-hand side of each equation the most recently updated values for the other functions are used. Convergence to the solutions is achieved within typically ten cycles of iteration.

In terms of normalized quantities introduced earlier in Eq. (64), the problem (48)–(50) can be written as

$$\hat{\nabla}\cdot(R\hat{\nabla}P)\equiv\frac{1}{\hat{r}}\frac{\partial}{\partial\hat{r}}\left(\hat{r}R\frac{\partial P}{\partial\hat{r}}\right)+\frac{1}{\hat{r}^2}\frac{\partial}{\partial\theta}\left(R\frac{\partial P}{\partial\theta}\right)=\frac{Ra}{\hat{E}}(\hat{u}-\hat{u}_c) \quad \text{in } 0<\hat{r}<1-\alpha, \quad 0<\theta<2\pi, \tag{76}$$

$$\frac{\partial P}{\partial\hat{r}}=0 \quad \text{at } \hat{r}=1-\alpha, \tag{77}$$

and

$$P \text{ is finite} \quad \text{at } \hat{r}=0. \tag{78}$$

As for N, we solve the problem for $P(\hat{r},\theta)$ by first expanding this function and $R(\hat{r},\theta)$ into Fourier series

$$P(\hat{r},\theta)=a\sum_{k=0}^{\infty} P_k(\hat{r})\cos k\theta, \tag{79}$$

$$R(\hat{r},\theta)=1+K_d\overline{\zeta f}(r,\theta)$$

$$=[1+\beta\hat{\bar{\zeta}}f_0I_0]+2\sum_{k=1}^{\infty}[\beta\hat{\bar{\zeta}}f_0I_k]$$

$$=M_0(-\delta\hat{r})+2\sum_{k=1}^{\infty}[M_k(-\delta\hat{r})\cos k\theta], \tag{80}$$

where

$$\hat{\bar{\zeta}}=\bar{\zeta}/\tilde{\zeta}, \quad \beta=K_d\tilde{\zeta}, \tag{81}$$

and $\tilde{\zeta}$ is a characteristic scale for the particle concentration. The parameter β is the bulk solid-fluid distribution ratio of the chemical.

Clearly, the functions $P_k(\hat{r})$ for $k=0,1,2,3$ can be solved using the same set of Eqs. (69)–(73) where $N_k(\hat{r})$, $I_k(-\delta\hat{r})$ and $(\hat{u}-\hat{u}_s)$ are replaced by $P_k(\hat{r})$, $M_k(-\delta\hat{r})$ and $(\hat{u}-\hat{u}_c)$, respectively. One should note that, while N_k, I_k, and $\hat{u}_s$ are pure functions of $\hat{r}$, their counterparts P_k, M_k, and $\hat{u}_c$ are functions of x and t as well because they depend on $\zeta(x,t)$.

The Taylor dispersion coefficient for the solute can then be evaluated using the following integral:

$$\hat{\mathcal{D}}_{Tc}=-\frac{\overline{RP(\hat{u}-\hat{u}_c)}}{\bar{R}a}=-\frac{2}{\bar{R}(1-\alpha^2)}\int_0^{1-\alpha}\left(\sum_{k=0}^{\infty}M_kP_k\right)(\hat{u}-\hat{u}_c)\hat{r}d\hat{r}. \tag{82}$$

Also, using Eqs. (76) and (77), one may show that $\hat{\mathcal{D}}_{Tc}$ is always positive:

$$\hat{\mathcal{D}}_{Tc}=\hat{E}\overline{R|\hat{\nabla}P|^2}/a^2>0. \tag{83}$$

As noted above, while $\hat{\mathcal{D}}_{Ts}$ is a constant, $\hat{\mathcal{D}}_{Tc}$ can vary with x and t.

Recall that a positive δ means settling particles while a negative one means rising particles. Since I_0 and I_2 are even functions and I_1 and I_3 are odd functions and therefore M_0 and M_2 are even functions and M_1 and M_3 are odd functions correspondingly, it can be deduced that with respect to δ, P_0, and P_2 are even functions while P_1 and P_3 are odd functions as well. Consequently, the dispersion coefficient $\hat{\mathcal{D}}_{Tc}$ is also an even function of δ, which confirms that the solute dispersion is symmetrical with respect to the falling and rising of the particles.

As remarked earlier, the retardation factor and hence the advection speed and dispersion coefficient for the solute depend on the local particle concentration, which varies in general with the axial distance and time. By contrast, the advection speed and dispersion coefficient for the particles are independent of these variables. The solution sequence is to first find the particle distribution as a function of time, by which the advection speed and dispersion coefficient for the solute can be evaluated, and then the chemical concentration can be solved also as a function of space and time. A numerical example is presented in the next section.

6 Numerical Example—A Continuous Discharge of a Sorbing Chemical With a Pulse Input of Particles

To illustrate our theory, we consider an example of transport in pipe flow of two phases subject to the effects of sorptive exchange. This is a case in which a source discharges continuously and steadily a soluble sorbing chemical well upstream into a section of long horizontal pipe carrying fully developed turbulent flow. Before the front of the chemical concentration is about to reach a point down gradient, a finite amount of particles is released as a pulse input into the flow at the point. Practically, the continuous source of chemical may correspond to leakage from, for example, a chemical drum or a buried waste tank. Before the polluted fluid reaches downstream, some particles onto which the chemical can be sorbed are released ahead of the pollutant front. The particles move slower than the flow and hence will be overtaken by the chemical front. As the chemical front passes through the particle cloud, sorption takes effect and will affect the breakthrough characteristics of the chemical.

We define $x=0$ and $t=0$ as the point and the time at which the particles are released. By virtue of the particle transport Eq. (38) which has constant coefficients, the particle concentration can be described by a Gaussian distribution in the frame $(\xi=x-u_st)$ which moves at the speed of the center of mass of the particle cloud:

$$\bar{\zeta}(\xi,t)=\frac{m}{(4\pi D_st)^{1/2}}\exp\left(-\frac{\xi^2}{4\mathcal{D}_st}\right) \quad \text{for } -\infty<\xi<\infty, \quad t>0, \tag{84}$$

where m is the total mass of particle divided by the cross-sectional area of the pipe, which by conservation of mass is the area under the concentration distribution curve at any time:

$$\int_{-\infty}^{\infty}\bar{\zeta}d\xi=m \quad \text{for all } t>0. \tag{85}$$

Transforming from (x,t) to (ξ,t) and using Eq. (46), the solute transport Eq. (60) becomes

$$\frac{\partial C}{\partial t}+\left[\frac{\bar{u}-u_s}{\bar{R}}+u_c'\right]\frac{\partial C}{\partial\xi}=\frac{\partial}{\partial\xi}\left(D_c\frac{\partial C}{\partial\xi}\right), \tag{86}$$

where the leading order subscript has been suppressed. This is the rate of change of solute concentration as observed from the center of the particle cloud. Because of the apparent advection (inside the square brackets) which is mostly positive, the solute front will catch up and eventually pass the particle cloud.

For a continuous discharge from upstream, the boundary conditions are

$$C(\xi,t)\to C_{-\infty}=\text{constant as } \xi\to-\infty, \tag{87}$$

$$C(\xi,t)\to 0 \text{ as } \xi\to\infty \text{ for finite } t, \tag{88}$$

and

$$C(\xi,t)\to C_{-\infty} \text{ for finite } \xi \text{ as } t\to\infty. \tag{89}$$

The chemical concentration is maintained at a constant value $C_{-\infty}$ far upstream. Further suppose that when the particles are released, or at $t=0$, the chemical source has been operating for a period of time t_c. An initial chemical concentration distribution satisfying Eqs. (86)–(89) can then be obtained as follows:

$$C(\xi,0)=\frac{C_{-\infty}}{2}\text{erfc}\left(\frac{\xi-\xi_0}{(4D_ct_c)^{1/2}}\right), \tag{90}$$

where erfc is the complementary error function, and ξ_0 is the initial position of the center of the front at $t=0$, which is the point where the concentration is half the maximum value. The front is initially far upstream from the particle discharge point $x=0$, and therefore $\xi_0<0$.

Let us recall Eq. (64) and introduce the following additional normalized quantities (distinguished by a caret):

$$(\xi,\xi_0)=L(\hat{\xi},\hat{\xi}_0), \quad (t,t_c)=(L^2/u_*a)(\hat{t},\hat{t}_c), \quad \bar{\zeta}=(m/L)\hat{\bar{\zeta}},$$

$$(C,C_{\text{tot}})=C_{-\infty}(\hat{C},\hat{C}_{\text{tot}}), \quad (\mathcal{D}_s,\mathcal{D}_c)=u_*a(\hat{\mathcal{D}}_s,\hat{\mathcal{D}}_c),$$

$$(\bar{u},u_s,u_c,u_c')=u_*(\hat{\bar{u}},\hat{u}_s,\hat{u}_c,\hat{u}_c'), \quad \beta=K_dm/L, \tag{91}$$

where u_* is the shear velocity, and L is a longitudinal length scale for the transport. In terms of the normalized quantities, the equations can be expressed as follows. The particle concentration distribution is

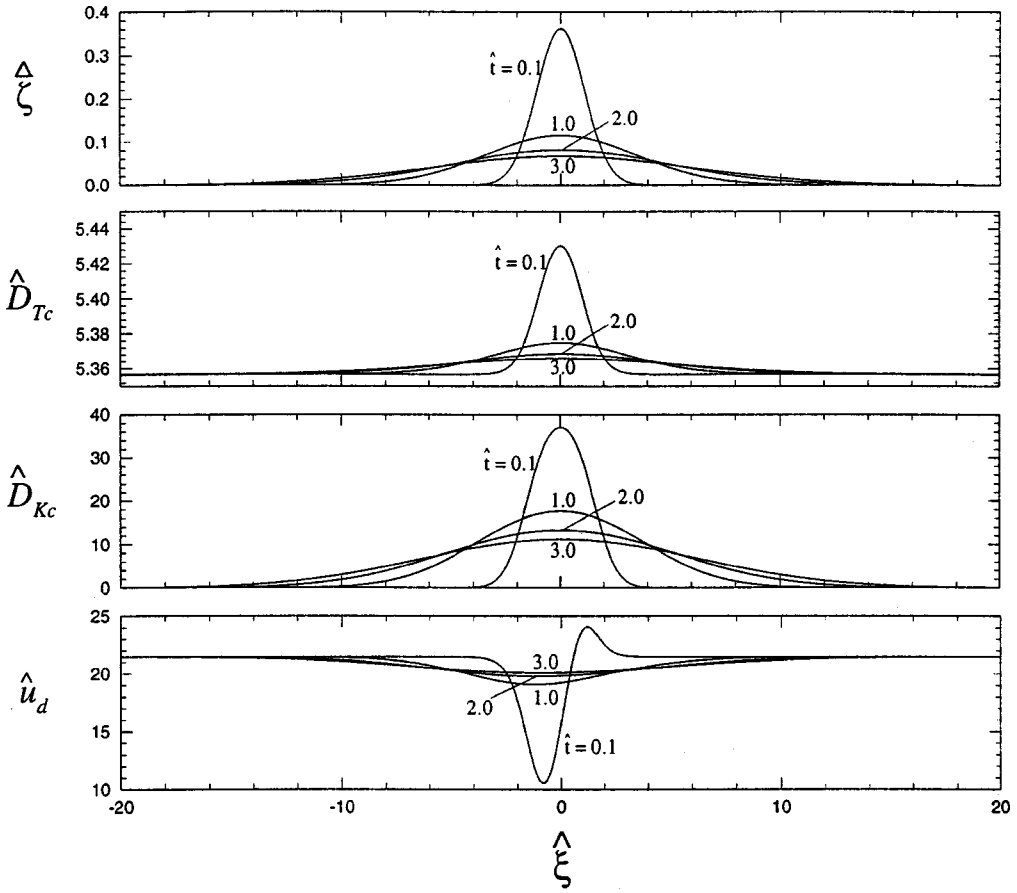


Fig. 1 Distributions of the particle concentration $\hat{\bar{\zeta}}$, solute Taylor dispersion coefficient $\hat{D}_{Tc}$, sorption-kinetics-induced dispersion coefficient $\hat{D}_{Kc}$, and drifting velocity $\hat{u}_d$ for Case 1 (β=1)

$$\hat{\bar{\zeta}}(\hat{\xi},\hat{t})=(4\pi\hat{\mathcal{D}}_s\hat{t})^{-1/2}\exp\left(-\frac{\hat{\xi}^2}{4\hat{\mathcal{D}}_s\hat{t}}\right), \tag{92}$$

while the particle distribution factor f is

$$f(\hat{r},\theta)=f_0\exp(-\delta\hat{r}\cos\theta). \tag{93}$$

The retardation factor R is

$$R(\hat{\xi},\hat{r},\theta,\hat{t})=1+\beta f\hat{\bar{\zeta}} \quad \text{and} \quad \overline{R}(\hat{\xi},\hat{t})=1+\beta\hat{\bar{\zeta}}. \tag{94}$$

Transport of the chemical is governed by

$$\frac{\partial\hat{C}}{\partial\hat{t}}+\hat{u}_d\frac{\partial\hat{C}}{\partial\hat{\xi}}=\frac{\partial}{\partial\hat{\xi}}\left(\hat{\mathcal{D}}_c\frac{\partial\hat{C}}{\partial\hat{\xi}}\right), \tag{95}$$

where $\hat{u}_d$ is the drifting velocity given by

$$\hat{u}_d(\hat{\xi},\hat{t})=-\frac{L}{aR}\overline{(f-1)(\hat{u}-\hat{\bar{u}})}-\frac{\beta}{R}(\hat{\mathcal{D}}_s+\hat{\mathcal{D}}_c)\frac{\partial\hat{\bar{\zeta}}}{\partial\hat{\xi}}, \tag{96}$$

which has been normalized with respect to au_*/L, and describes how fast the dissolved phase of the chemical moves relative to the particle cloud. Ignoring the components due to the longitudinal eddy diffusivity, the dispersion coefficients are computed using the formulas: $\hat{\mathcal{D}}_s=\hat{\mathcal{D}}_{Ts}$ is given by Eq. (68), and

$$\hat{\mathcal{D}}_c(\hat{\xi},\hat{t})=\hat{\mathcal{D}}_{Tc}+\hat{\mathcal{D}}_{Kc}, \tag{97}$$

where $\hat{\mathcal{D}}_{Tc}(\hat{\xi},\hat{t})$ is given by Eq. (82), and

$$\hat{\mathcal{D}}_{Kc}(\hat{\xi},\hat{t})=-\beta\gamma\hat{\bar{\zeta}}\hat{u}_c\,\overline{(f-1)(\hat{u}-\hat{\bar{u}})}/\overline{R}^2, \tag{98}$$

where

$$\gamma=u_*/ka \tag{99}$$

is the sorption kinetics parameter, which should be of order unity or greater.

For convenience, we may choose that initially $4\hat{\mathcal{D}}_c\hat{t}_c=1$, and therefore the initial condition (90) is simplified to

$$\hat{C}(\hat{\xi},0)=\frac{1}{2}\operatorname{erfc}(\hat{\xi}-\hat{\xi}_0). \tag{100}$$

The integral in Eq. (85) now becomes

$$\int_{-\infty}^{\infty}\hat{\bar{\zeta}}d\hat{\xi}=1, \quad \text{for all } \hat{t}>0, \tag{101}$$

which provides a check for computational accuracy.

Equation (95) is solved numerically using a standard second-order implicit scheme of forward-time and central space differences. Approximation of the differential equation by implicit finite differences produces a tridiagonal linear system which can be solved with a simple routine. The condition in (101) is checked by numerical integration at each time step to ensure that mass conservation is observed throughout the computation. Sufficiently small spatial discretizations ($\Delta\hat{\xi}=0.01$, $\Delta\hat{r}=0.0005$) and time step ($\Delta\hat{t}=0.002$) have been used so that the maximum error is kept below 1%. General considerations (e.g., numerical dispersion and artificial oscillation) in applying a numerical scheme to this kind of problem can be found in Wood [14] and Zheng and Bennett [[15], Chap. 6].

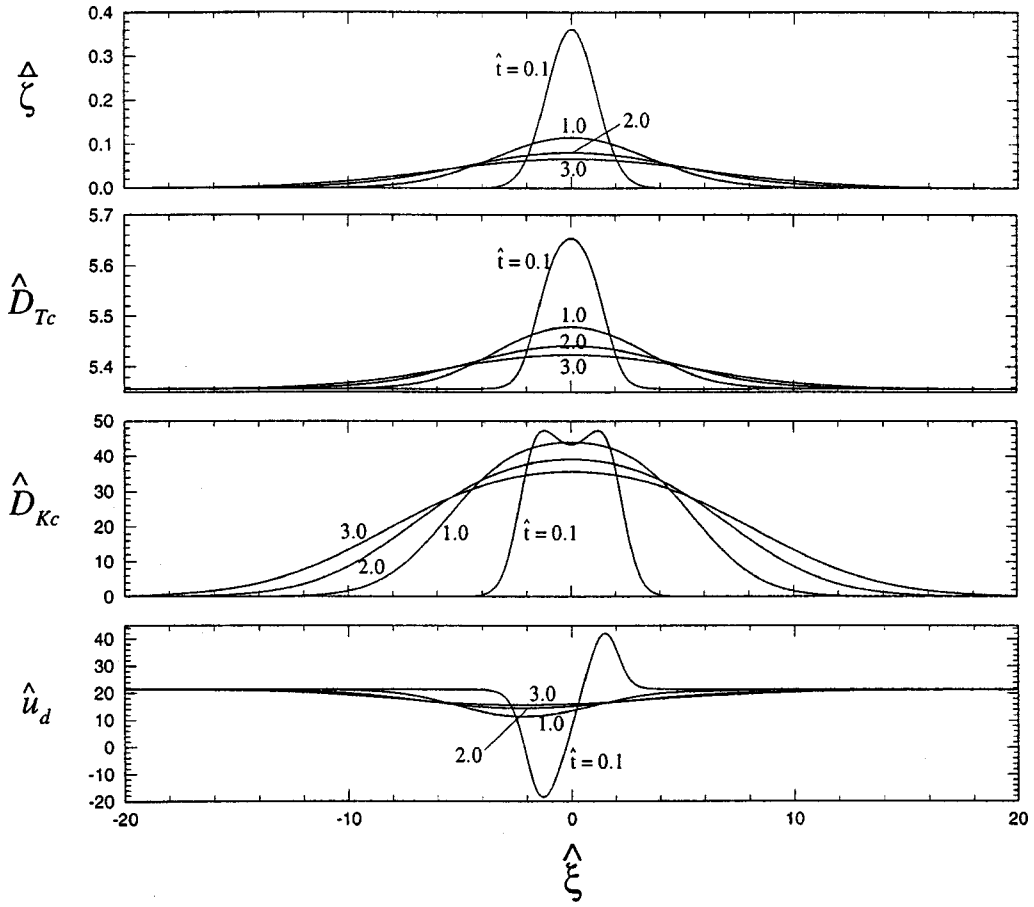


Fig. 2 As Fig. 1, but for Case 2 (β=5)

7 Discussion

To examine the effects of sorption, we have generated computational results for three different values of the bulk solid-fluid distribution ratio, which are β=1 (Case 1), β=5 (Case 2), and β=10 (Case 3). Cases 1 through 3 correspond to an increasingly large fraction of mass of the chemical in the sorbed form. For the other parameters, the following values are chosen for all the three cases: δ=1, γ=50, $L/a=100$, $\hat{\xi}_0=-10$, $\hat{E}=0.1$, and α=0. The fairly large γ=50 means a rather strong sorption kinetics or a small sorption rate constant. The vanishing particle-to-pipe radius ratio, α=0, means that the particle size is negligibly small compared with the pipe radius.

Based on some empirical data for the eddy diffusivity as a function of r, Taylor [2] evaluated that the dispersion coefficient for an inert solute in turbulent flow through a pipe was $10.06u_*a$. However, should one use the classical logarithmic velocity defect law, one would obtain a value of $5.11u_*a$ for the dispersion coefficient. In this work, it has been assumed for the sake of analyticity that the eddy diffusivity is constant. A value of $\hat{E}=0.1$, which corresponds to the core value in a parabolic-constant distribution for the eddy diffusivity, is chosen for the present example. This choice of eddy diffusivity yields a dispersion coefficient of $5.36u_*a$ for an inert solute, which is somewhat different from Taylor's value, but close to the one obtained with the velocity defect law. Anyhow, since our interest is in the relative changes brought about by sorptive exchange to the dispersion and advection processes, the absolute value for a dispersion coefficient is immaterial to the discussions here.

Also, the suspension number δ=0.1 yields a dispersion coefficient $\hat{D}_s=5.94u_*a$ for the suspended particles, which is slightly greater than that for a solute (i.e., a neutrally buoyant phase). The effect of the suspension number δ on the dispersion coefficient D_s has been discussed by Ng [7]. By and large, the dispersion coefficient will increase with δ as long as $\delta<2$ (i.e., relatively fine particles), and beyond this limit the dispersion coefficient will decrease with increasing δ. To be consistent with the assumption that the particles will remain in suspension mostly all the time, a small value of δ=0.1 has been used for this numerical example. The calculated values for $\overline{u}$ and u_s are, respectively, $17.74u_*$ and $17.53u_*$, so the particle advection velocity is only 1% lower than the discharge velocity.

Figures 1–3 show the spatial distributions of $\hat{\zeta}$, $\hat{\mathcal{D}}_{Tc}$, $\hat{\mathcal{D}}_{Kc}$, and $\hat{u}_d$ at instants $\hat{t}=0.1$, 1.0, 2.0, and 3.0 for the three values of β. One can see from these plots how the dispersion coefficients and the drifting velocity for the chemical, which are functions of the particle concentration, vary with distance along the pipe and time. In the absence of particles at large $|\hat{\xi}|$, $\hat{\mathcal{D}}_{Tc}$ reduces to the value for an inert solute, while $\hat{\mathcal{D}}_{Kc}$ drops to zero. In all cases, $\hat{\mathcal{D}}_{Tc}$ is only slightly increased by the presence of particles (<10%). The distribution and magnitude of $\hat{\mathcal{D}}_{Kc}$ is much affected by the degree of sorptive exchange, as is expected. We observe that for a sufficiently large value of γ and β, the sorption-kinetics-induced dispersion coefficient $\hat{\mathcal{D}}_{Kc}$ always dominates over the Taylor dispersion coefficient $\hat{\mathcal{D}}_{Tc}$ as long as the particle concentration is finite. In these three cases, $\hat{\mathcal{D}}_{Kc}$ can be as much as one order of magnitude larger than $\hat{\mathcal{D}}_{Tc}$ at the center of the particle cloud. It is also worth noting that in Cases 2 and 3, there exist two local maxima of $\hat{\mathcal{D}}_{Kc}$ at early interaction of the solute and particles. The particle concentration gradient leads to a distinct distribution of the drifting velocity $\hat{u}_d$, which exhibits a maximum downstream and a minimum upstream of the center of the particle cloud. The mini-

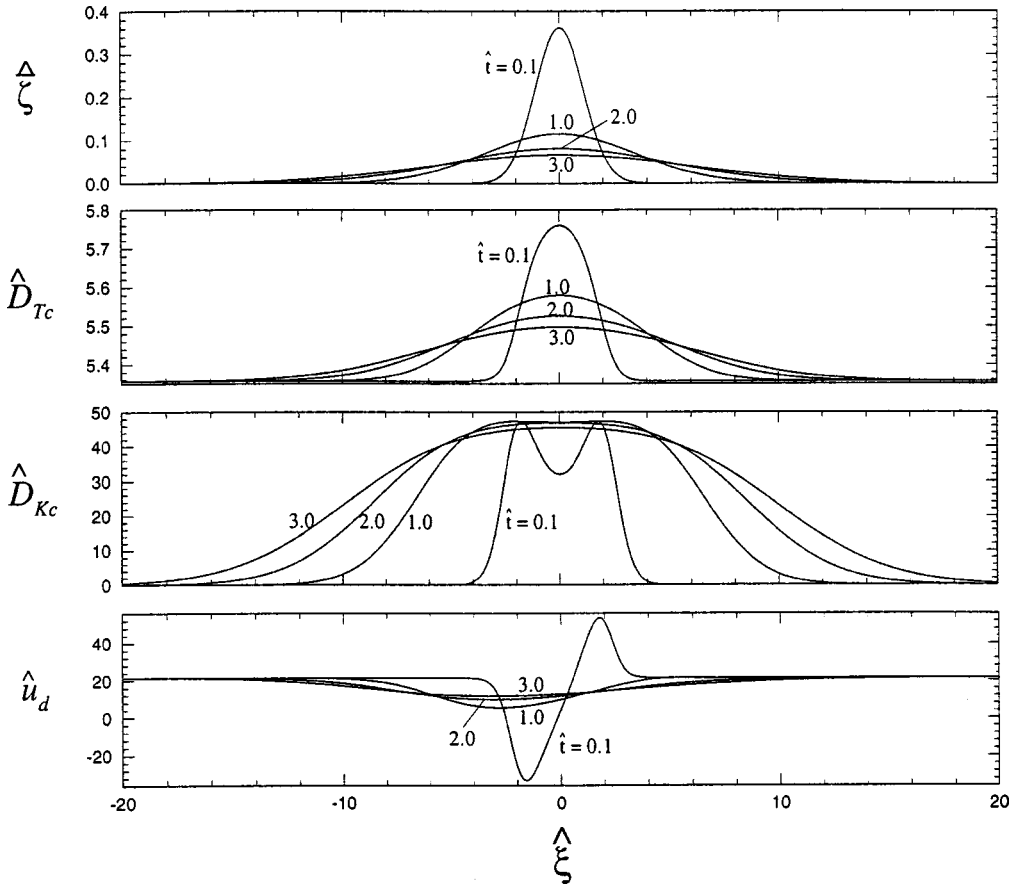


Fig. 3 As Fig. 1, but for Case 3 (β=10)

mum $\hat{u}_d$ can even be negative at early stages in Cases 2 and 3. Physically it means that the chemical is being advected at a faster rate if downstream from the particle cloud center, but a slower rate (even in the opposite direction) if upstream. This will effectively cause additional dispersion on the spreading of the chemical front as it passes through the particle cloud.

Snapshots of the concentration distributions for $\hat{C}$ (solid curves) and $\hat{\zeta}$ (dashed curves) are plotted in Fig. 4 for the three cases. This figure shows the changes in the profile of the chemical concentration front in the course of the chemical overtaking the particle cloud. It is obvious that in Cases 2 and 3 the chemical front changes its shape from concave down initially to concave up as it passes through the particle cloud. In such cases, the sorption kinetics is strong while the sorption partition is high: The spreading of the chemical front is greatly influenced by the presence and the distribution of the particles. Of course, when the front has largely passed through the particle cloud, its profile recovers an S-shape.

The following statistical parameters (i) the location of the center of the front $\hat{\xi}_c$, (ii) the variance σ^2, and (iii) the skew coefficient χ can be used to characterize the distribution of the chemical concentration front. Their definitions are as follows:

$$\hat{\xi}_c = \frac{\mu_1}{\mu_0}, \tag{102}$$

$$\sigma^2 = \frac{1}{\mu_0}\int_{-\infty}^{\infty} (\hat{\xi}-\hat{\xi}_c)^2 \left(-\frac{d\hat{C}}{d\hat{\xi}}\right) d\hat{\xi} = \frac{\mu_2}{\mu_0} - \frac{\mu_1^2}{\mu_0^2}, \tag{103}$$

and

$$\chi = \frac{1}{(\sigma^2)^{3/2}\mu_0}\int_{-\infty}^{\infty} (\hat{\xi}-\hat{\xi}_c)^3 \left(-\frac{d\hat{C}}{d\hat{\xi}}\right) d\hat{\xi}$$

$$= \frac{1}{(\sigma^2)^{3/2}}\left[\frac{\mu_3}{\mu_0} - 3\frac{\mu_1}{\mu_0}\frac{\mu_2}{\mu_0} + 2\frac{\mu_1^3}{\mu_0^3}\right], \tag{104}$$

where μ_n is the nth integral moment of $-d\hat{C}/d\hat{\xi}$:

$$\mu_n = \int_{-\infty}^{\infty} \hat{\xi}^n \left(-\frac{d\hat{C}}{d\hat{\xi}}\right) d\hat{\xi}. \tag{105}$$

We remark that the rate of increase of the variance σ^2 gives the rate of broadening of the front. In particular, for an S-curve given by the complementary error function (90), $\hat{\xi}_c$ is the location of the center where the function value is half the maximum value and the slope is the steepest. Also, the skew coefficient χ of this function is zero because of its symmetry in slope about the center.

The three parameters are plotted as functions of time for the three cases in Fig. 5. For comparison, the corresponding values for a nonsorbing chemical is also plotted. On comparing with this limiting case, one can readily observe that the sorptive exchange can cause the following effects as soon as the chemical front interacts with the suspension cloud. First, the rate of movement of the front is reduced. This would delay the arrival of the maximum impact of the chemical at a certain point in the pipe downstream. Second, the rate of broadening of the front is increased. Third, the S-symmetry of the front is destroyed. The skew coefficient increases from zero initially to a positive value as the leading part of the front (i.e., the part ahead of the inflexion point) disperses more extensively than the trailing part (i.e., the part behind the inflexion

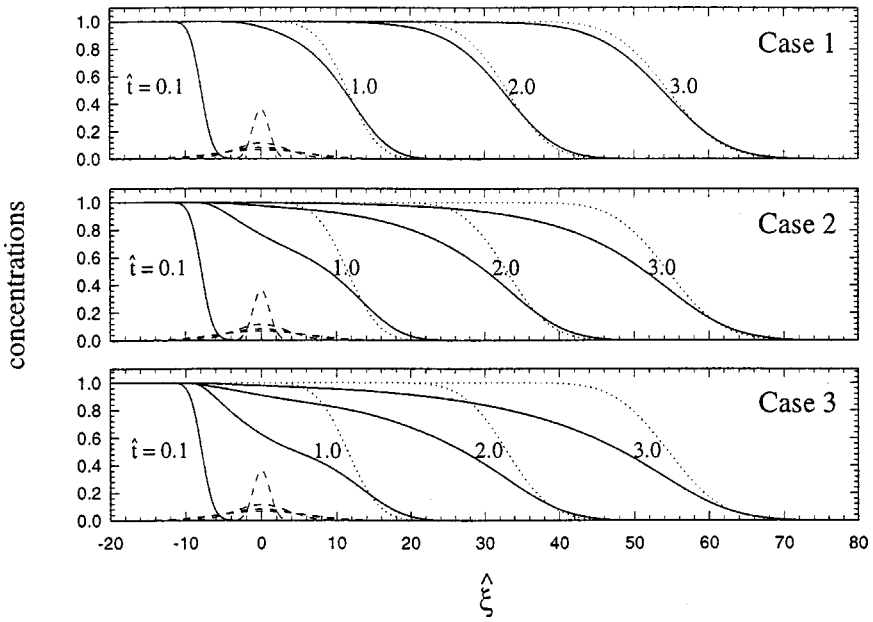


Fig. 4 Snapshots of the distributions of the solute concentration $\hat{C}(\hat{\xi},\hat{t})$ (solid lines) and the particle concentration $\hat{\bar{\xi}}(\hat{\xi},\hat{t})$ (dashed lines) for Cases 1, 2, and 3. The dotted lines represent the limiting case when the chemical is non-sorbing or the sorptive exchange is nil.

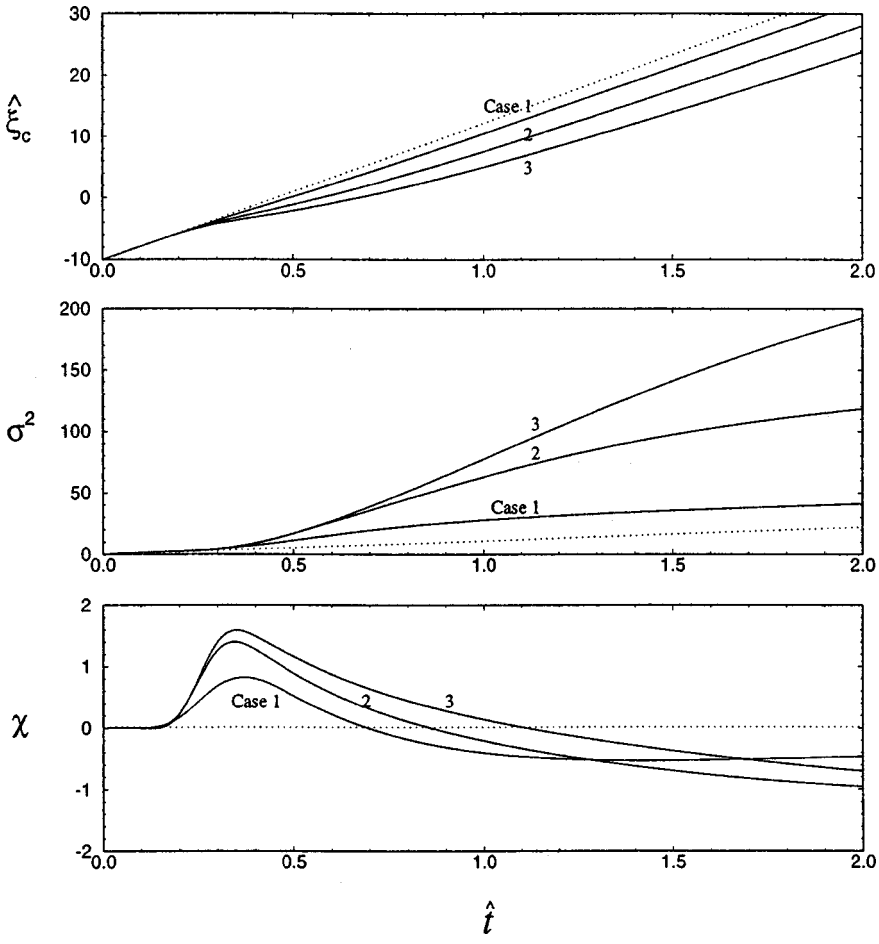


Fig. 5 The location of the center $\hat{\xi}_c$, the variance σ^2, and the skew coefficient χ for the chemical front as a function of time $\hat{t}$ for Cases 1, 2, and 3. The dotted lines represent the limiting case when the chemical is nonsorbing or the sorptive exchange is nil.

point). The skew coefficient soon turns to negative as the inflexion point of the front overtakes the suspension cloud center. The skewness then stays negative for some time before gradually getting back to zero, implying a long tail of concentration defect on the trailing part of the front. The concentration does not reach the maximum level until the center of the front has long passed the particle cloud. All these above-mentioned effects are the most extensive for Case 3, in which the sorption kinetics is the strongest and the mass fraction of chemical sorbed on the particles is the largest.

8 Summary and Concluding Remarks

In this paper we have used an asymptotic method to obtain effective transport Eqs. (38) and (60) for suspended particles and a sorbing solute in a steady fully developed turbulent pipe flow. The key assumptions for the analysis include (i) the pipe radius is much smaller than the longitudinal length-scale of transport; (ii) the time-scale for transverse turbulent diffusion is much shorter than that for longitudinal advection; (iii) the height of distribution of suspended particles is comparable to the pipe diameter; (iv) the time-scales for sorption kinetics and transverse turbulent diffusion are comparable to each other; (v) there is a finite fraction of mass of the chemical in sorbed form; and (vi) the sorptive exchange is described by a first-order kinetics relation. Because of phase partitioning and sorption kinetics, the advection velocity and dispersion coefficient for the chemical transport are shown to be functions of longitudinal distance and time via dependence on the local particle concentration. The settling of particles renders the advection speed of suspended cloud slower than that of the chemical, which in turn is smaller than the discharge velocity. The dispersion coefficient has two components: (i) $\mathcal{D}_{Tc}$, a Taylor dispersion coefficient, which because of phase partitioning is greater than the clear water value; and (ii) $\mathcal{D}_{Kc}$, a sorption-kinetics-induced dispersion coefficient, which is proportional to the cross-sectional covariance between the fluid velocity and particle concentration profiles, and inversely proportional to the sorption rate constant. These coefficients are essentially controlled by three parameters: suspension number (δ), bulk solid-fluid distribution ratio (β) and the sorption kinetics parameter (γ).

With a numerical example, we have examined the effects of kinetic sorptive exchange on the concentration front distribution of a sorbing chemical discharged continuously into a pipe, as the front interacts with a finite cloud of suspension. When sorption effects are strong, the sorption-kinetics-induced dispersion coefficient $\mathcal{D}_{Kc}$ dominates and the longitudinal variations of the advection velocity can add to broaden the distribution. As a result, the concentration front deviates from an error function type of profile soon after interacting with the solid suspension cloud. In summary, the sorptive exchange can be very influential in governing the advection speed, the rate of broadening and the extent of departure from symmetry of the distribution of a sorbing solute, if the suspended particles are rich in sorbents or sites for sorption.

The use of the above-mentioned effects in controlling pollutant transport in pipe flow is of great potential value and deserves more in-depth future investigations including experiments. It is also worth extending the present analysis to the case when the particles are so heavy that they fall out of suspension soon after release into the flow, and thereafter form a layer of immobile reactive sediment on the pipe wall.

Acknowledgments

The research was sponsored by the Hong Kong Research Grants Council under Grants HKU 7117/99E and HKU 7081/02E. The first author was supported financially by a Post-Doctoral Fellowship for the Area of Excellence in Harbor and Coastal Environment Studies.

References

[1] Taylor, G. I., 1953, ''Dispersion of Soluble Matter in Solvent Flowing Slowly Through a Tube,'' Proc. R. Soc. London, Ser. A, **219**, pp. 186–203.
[2] Taylor, G. I., 1954, ''The Dispersion of Matter in Turbulent Flow Through a Pipe,'' Proc. R. Soc. London, Ser. A, **223**, pp. 446–468.
[3] Smith, R., 1983, ''Effect of Boundary Absorption Upon Longitudinal Dispersion in Shear Flow,'' J. Fluid Mech., **134**, pp. 161–177.
[4] Smith, R., 1995, ''How Far Can a Boundary Coating Material be Carried Downstream in Turbulent Pipe Flow?'' J. Eng. Math., **29**, pp. 51–62.
[5] Purnama, A., 1988, ''Boundary Retention Effects Upon Contaminant Dispersion in Parallel Flows,'' J. Fluid Mech., **195**, pp. 393–412.
[6] Ng, C. O., and Yip, T. L., 2001, ''Effects of Kinetic Sorptive Exchange on Solute Transport in Open-Channel Flow,'' J. Fluid Mech., **446**, pp. 321–345.
[7] Ng, C. O., 2002, ''On the Longitudinal Dispersion of Heavy Particles in a Horizontal Circular Pipe,'' Int. J. Eng. Sci., **40**, pp. 239–250.
[8] Elder, J. W., 1959, ''The Dispersion of Marked Fluid in Turbulent Shear Flow,'' J. Fluid Mech., **5**, pp. 544–560.
[9] Sumer, B. M., 1974, ''Mean Velocity and Longitudinal Dispersion of Heavy Particles in Turbulent Open-Channel Flow,'' J. Fluid Mech., **65**, pp. 11–28.
[10] Chatwin, P. C., 1970, ''The Approach to Normality of the Concentration Distribution of a Solute in a Solvent Flowing Through a Straight Pipe,'' J. Fluid Mech., **43**, pp. 321–352.
[11] Mei, C. C., 1992, ''Method of Homogenization Applied to Dispersion in Porous Media,'' Transp. Porous Media, **9**, pp. 261–274.
[12] Ng, C. O., 2000, ''Dispersion in Particle-Laden Stream Flow,'' J. Eng. Mech., **126**(8), pp. 779–786.
[13] Abramowitz, M., and Stegun, I. A., 1972, *Handbook of Mathematical Functions*, Dover, New York.
[14] Wood, W. L., 1993, *Introduction to Numerical Methods for Water Resources*, Oxford University Press, New York.
[15] Zheng, C., and Bennett, G. D., 1995, *Applied Contaminant Transport Modeling*, Van Nostrand Reinhold, New York.

G. M. L. Gladwell
Department of Civil Engineering,
University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada

M. M. Khonsari
Fellow ASME

Y. M. Ram
Mem. ASME

Department of Mechanical Engineering,
Louisiana State University,
Baton Rouge, LA 70806

Stability Boundaries of a Conservative Gyroscopic System

Depending on the speed of rotation, a gyroscopic system may lose or gain stability. The paper characterizes the critical angular velocities at which a conservative gyroscopic system may change from a stable to an unstable state, and vice versa, in terms of the eigenvalues of a high-order matrix pencil. A numerical method for evaluation of all possible candidates for such critical velocities is developed. [DOI: 10.1115/1.1574062]

1 Introduction

Problems involving infinitesimal oscillations of particles and bodies attached to rotating frames lead to the quadratic eigenvalue problem

$$(\lambda^2\mathbf{M}+\omega\lambda\mathbf{G}+\mathbf{K}-\omega^2\mathbf{R})\mathbf{v}=\mathbf{o}, \quad \mathbf{M},\mathbf{G},\mathbf{K},\mathbf{R}\in\Re^{n\times n}, \quad \mathbf{v}\neq\mathbf{o}\in C^n \tag{1}$$

where $\mathbf{M}$, $\mathbf{K}$, and $\mathbf{R}$ are symmetric positive definite matrices, $\mathbf{G}=-\mathbf{G}^T$ is a skew-symmetric matrix, ω is the angular velocity of the rotating frame, and n denotes the number of degrees-of-freedom in the system.

The system is said to be (*weakly*) *stable* if there exist no initial conditions causing the response of the system to increase with time without bound. Denote the two-variable matrix pencil

$$\mathbf{P}(\lambda,\omega)=\lambda^2\mathbf{M}+\omega\lambda\mathbf{G}+\mathbf{K}-\omega^2\mathbf{R}, \tag{2}$$

and let λ_i, $i=1,2,\ldots,2n$, be the eigenvalues of (1), satisfying

$$P(\lambda_i,\omega)\triangleq\det(\mathbf{P}(\lambda_i,\omega))=0. \tag{3}$$

Then the system (1) is stable if all of its eigenvalues are purely imaginary and distinct, i.e.,

$$\mathrm{Re}(\lambda_k)=0, \quad k=1,2,\ldots,2n; \quad \lambda_i\neq\lambda_j \ \text{ for } \ i\neq j. \tag{4}$$

If at least one of the eigenvalues of (1) has a nonvanishing real part, or if λ_k is a purely imaginary eigenvalue of multiplicity $p>1$, and there are not p linearly independent eigenvectors associated with it, then the system is unstable.

Note that this criterion gives no indication regarding the range of ω for which the system is stable or not. The essential problem treated in this paper is the determination of the critical values of ω, for which the system (1) may change from a stable state to unstable state and vice versa. These critical values are possible *stability boundaries*, [1].

There exists a wealth of literature associated with properties and stability of gyroscopic systems containing various results involving necessary or sufficient conditions for stability, which are based on matrix properties and inequalities, see, e.g., [2–8]. Veselić [9] and Hryniv and Lancaster [10] have investigated the stability of gyroscopic system in the context of the two-parameter matrix pencil (2). They have determined conditions ensuring that gyroscopic systems are stable for all sufficiently large values of ω. The behavior of the eigenvalues of gyroscopic systems has been also studied by Seyranian and Kliem in [11] and Seyranian et al. in [12]. In an important work, Afolabi [13] has characterized the stability boundaries via explicit expansion of the characteristic polynomial. Afolabi found that the stability boundaries are determined by the roots of the discriminant of the characteristic Eq. (3), and by other values of ω which cause the absolute term of the characteristic equation (i.e., the term which is independent of λ) to vanish. However, explicit expansion of the characteristic polynomial (3) involves symbolic manipulations of the order of $n!$ operations. The results presented here may be regarded as the matrix analogues of Afolabi's criteria. We characterize the possible stability boundary in terms of the eigenvalues of a certain matrix eigenvalue problem, without requiring the expansion of the characteristic polynomial. It is shown that evaluation of all possible stability boundaries can be achieved by using numerical procedures including polynomial fitting, interpolation, and eigenvalue extraction. We note that the reduction of the stability characterization problem into such paradigms is intended for the sake of clarity rather than computational efficiency.

We remark that in this work we are interested only in the determination of possible stability boundaries. The other important problem of determining whether, while crossing the critical values of ω, the system actually changes its nature from stable to unstable and vice versa, is not studied here. Once the stability boundaries are found, determination of the system stability within a particular stability interval is obtained using criterion (4), applied to a typical system within the stability interval.

The motivation for the study is presented in Section 2. It is shown that, with certain simplifying assumptions, the motion of particles and elastic bodies in rotating frames is characterized by the quadratic eigenvalue problem (1). Although the literature contains many papers and books on gyroscopic system modeling, the stability analysis in some studies has been devoted to problems involving generalization of the eigenvalue problem (1). Such a generalization may include for example an additional skew-symmetric stiffness matrix in (1). The materials presented in next section demonstrate that the fundamental physical problem involving rotation of rigid bodies with constant angular speed is formulated by the eigenvalue problem (1). It is also apparent from the formulation that if the system rotates with variable angular speed, then an additional skew-symmetric stiffness matrix is added to the formulation (1), but in this case ω is time-dependent and the problem cannot be reduced to the form (1). In Section 3 we show that the stability boundaries are distinguished by the existence of repeated eigenvalues. We then use this criterion in

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Oct. 7, 2001; final revision, Apr. 29, 2002. Associate Editor: O. O' Reilly. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

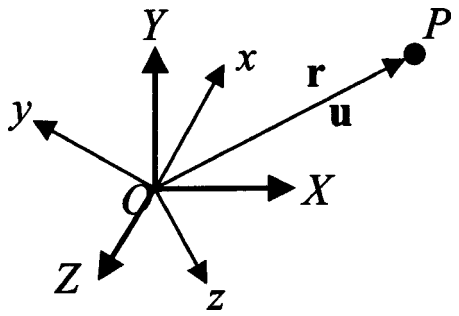


Fig. 1 The position of *P*

Section 4 to obtain a numerical method, allowing numerical evaluation of all possible stability boundaries. Examples demonstrating the results are presented in Section 5.

2 Modeling of Gyroscopic Systems

Consider an inertial Cartesian coordinate system $O(XYZ)$. Let $O(xyz)$ be a rotating coordinate system with the same origin, and let P be a moving particle of mass m. Denote the position of P in the inertial and rotating coordinate systems by the vectors $\mathbf{r}$ and $\mathbf{u}$, respectively, as shown in Fig. 1. Then

$$\mathbf{r}=\mathbf{u}. \tag{5}$$

Upon differentiating (5) with respect to t twice, we obtain

$$\frac{d^2\mathbf{r}}{dt^2}=\ddot{\mathbf{u}}+2\boldsymbol{\omega}\times\dot{\mathbf{u}}+\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{u})+\dot{\boldsymbol{\omega}}\times\mathbf{u}, \tag{6}$$

where $\boldsymbol{\omega}$ is the angular velocity vector of $O(xyz)$, time differentiation in the inertial coordinate system $O(XYZ)$ is denoted by d/dt, and dots represent time differentiation with respect to an observer in the rotating system. By virtue of (6), Newton's second law takes the form

$$\mathbf{f}=m(\ddot{\mathbf{u}}+2\boldsymbol{\omega}\times\dot{\mathbf{u}}+\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{u})+\dot{\boldsymbol{\omega}}\times\mathbf{u}) \tag{7}$$

where $\mathbf{f}$ is the resultant of forces applied to P, and $\dot{\mathbf{u}}$ and $\ddot{\mathbf{u}}$ are the relative velocity and acceleration of P with respect $O(xyz)$. Let γ_1, γ_2, and γ_3 be the direction cosines of $\boldsymbol{\omega}$ in $O(XYZ)$. Then

$$\boldsymbol{\omega}=\omega(\gamma_1 \quad \gamma_2 \quad \gamma_3)^T, \tag{8}$$

where ω is the magnitude of $\boldsymbol{\omega}$. Hence,

$$2m\boldsymbol{\omega}\times\dot{\mathbf{u}}=\omega\begin{bmatrix}0 & -2m\gamma_3 & 2m\gamma_2\\ 2m\gamma_3 & 0 & -2m\gamma_1\\ -2m\gamma_2 & 2m\gamma_1 & 0\end{bmatrix}\begin{pmatrix}\dot{u}_1\\ \dot{u}_2\\ \dot{u}_3\end{pmatrix}\triangleq\omega\mathbf{G}\dot{\mathbf{u}}, \tag{9}$$

$$m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{u})=m\omega^2\begin{bmatrix}-\gamma_2^2-\gamma_3^2 & \gamma_1\gamma_2 & \gamma_1\gamma_3\\ \gamma_1\gamma_2 & -\gamma_1^2-\gamma_3^2 & \gamma_2\gamma_3\\ \gamma_1\gamma_3 & \gamma_2\gamma_3 & -\gamma_1^2-\gamma_2^2\end{bmatrix}\begin{pmatrix}u_1\\ u_2\\ u_3\end{pmatrix}$$

$$\triangleq-\omega^2\mathbf{R}\mathbf{u}, \tag{10}$$

with the obvious definition of $\mathbf{G}$ and $\mathbf{R}$.

If the forces applied to P are proportional to the displacement of the particle relative to the moving frame, as shown in Fig. 2, where the particle is supported by springs attached to the rotating frame, then

$$\mathbf{f}=-\mathbf{K}\mathbf{u} \tag{11}$$

where $\mathbf{K}$ is a positive semi-definite symmetric stiffness matrix. Hence provided that the angular velocity of $O(xyz)$ is constant, the eigenvalue problem (1) is obtained from (7) by virtue of (9), (10), (11) and $\mathbf{M}=m\mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

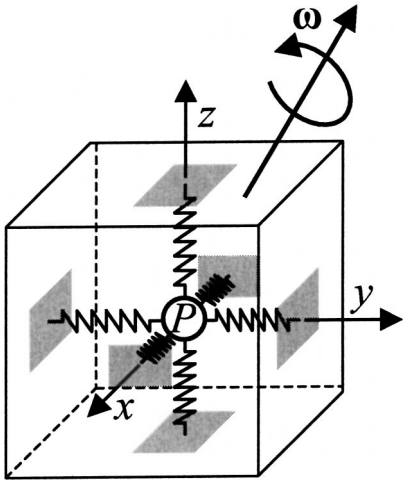


Fig. 2 A particle in a rotating frame

The eigenvalue problem (1) is also associated with motion of a multidegree-of-freedom vibrating system attached to a frame, which rotates with constant angular velocity. It generalizes naturally to a distributed parameter rotating system whose motion is governed by partial differential equations. Consider, for example, a rotating shaft with constant angular velocity $\boldsymbol{\omega}$, such as that shown in Fig. 3(*a*). Using a rotating coordinate system $O(xyz)$ we obtain from Newton's second law (7) applied to an infinitesimal element P of length dz

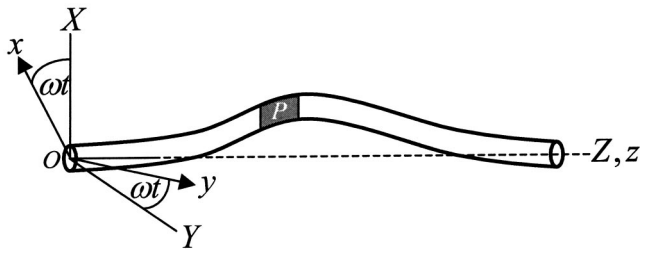
$$\mathbf{f}=(\ddot{\mathbf{u}}+2\boldsymbol{\omega}\times\dot{\mathbf{u}}+\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{u}))m dz \tag{12}$$

where m is the mass per unit length of the shaft, and $\mathbf{u}(t)=(u_1\ u_2)^T$ denotes the displacement of the element in the x and y-direction. The free-body diagram shown in Fig. 3(*b*) gives

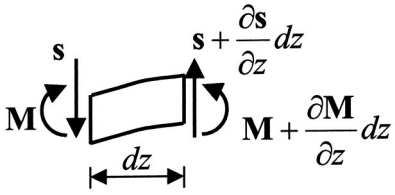
$$\mathbf{f}=\mathbf{s}'dz, \tag{13}$$

where $\mathbf{s}$ is the shear force vector,

$$\mathbf{s}=-(EIu'')', \tag{14}$$



(*a*)



(*b*)

Fig. 3 A continuous gyroscopic system: (*a*) a rotating shaft, and (*b*) a free-body diagram for a typical element

E is the modulus of elasticity, I is the moment of inertia of the shaft, and prime denotes differentiation with respect to z. Combining (12)–(14), where $\boldsymbol{\omega}=(0\ 0\ \omega)^T$, gives the equations of motion for small oscillations of the rotating shaft

$$\begin{pmatrix} EIu_1'' \\ EIu_2'' \end{pmatrix}'' + m\begin{pmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{pmatrix} + 2m\omega\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{pmatrix} \dot{u}_1 \\ \dot{u}_2 \end{pmatrix} + m\begin{bmatrix} -\omega^2 & 0 \\ 0 & -\omega^2 \end{bmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{15}$$

Finite difference approximation of this equation leads to the eigenvalue problem (1). It is important in practical applications to be able to determine the range of ω for which the system (15) is stable. This essential problem is studied here.

3 Characterization of the Stability Boundaries

Due to Duffin [5], Barkwell and Lancaster [4], Lancaster and Zizler [6], and many others, the properties of gyroscopic systems are well understood. For the sake of completion and self-sufficiency, we redevelop, state, and highlight in this section some of the properties applicable to the quadratic eigenvalue problem (1).

Proposition 1. If λ is a root of $P(\lambda,\omega)$ of multiplicity $p\geqslant 1$ then $-\lambda$, $\bar{\lambda}$ and $-\bar{\lambda}$ are roots of $P(\lambda,\omega)$ of the same multiplicity, where bar demotes complex conjugation.

Indeed,

$$P(-\lambda,\omega)=\det(\lambda^2\mathbf{M}-\lambda\omega\mathbf{G}+\mathbf{K}-\omega^2\mathbf{R}) = \det(\lambda^2\mathbf{M}-\lambda\omega\mathbf{G}+\mathbf{K}-\omega^2\mathbf{R})^T=P(\lambda,\omega), \tag{16}$$

since $\mathbf{M}$, $\mathbf{K}$, and $\mathbf{R}$ are symmetric and $\mathbf{G}=-\mathbf{G}^T$. Moreover, given that ω is real, the roots λ_i of $P(\lambda,\omega)$ must form a self-conjugate set, which completes the proof of the proposition.

The stability criterion (4) follows from Proposition 1. Moreover, since the eigenvalues are continuous functions of ω, it follows from the double symmetry property expressed by Proposition 1 that

Proposition 2. The system (3) may lose or gain stability only when λ is a root of $P(\lambda,\omega)$ of multiplicity $p>1$.

Note that the condition in Proposition 2 is necessary but not sufficient. The system will not necessarily lose or gain stability at a value of ω for which $P(\lambda,\omega)$ has a multiple root. Figure 4(*a*) shows schematically how the system may lose stability when, due to a small perturbation in ω, two pairs of repeated eigenvalues separate and leave the imaginary axis. Figure 4(*b*) shows how the system may gain stability when two pairs of eigenvalues converge to the imaginary axis. Near these values of ω the equation $P(\lambda,\omega)=0$ has eigenvalues that are near each other.

In light of Proposition 1 we may arrange the roots of $P(\lambda,\omega)$, for a fixed value of ω, such that $\lambda_i=-\lambda_{n+i}$ for $i=1,2,\ldots,n$, and obtain

$$P=\alpha\prod_{i=1}^{n}(\lambda-\lambda_i)(\lambda+\lambda_i)=\alpha\prod_{i=1}^{n}(\lambda^2-\lambda_i^2), \tag{17}$$

where α is a constant. Moreover, since $P(\lambda,\omega)=P(\lambda,-\omega)$, we have the following:

Proposition 3. The polynomial $P(\lambda,\omega)$ contains only even powers of ω and λ.

and

Proposition 4. If $\lambda=0$ is a root of $P(\lambda,\omega)$ then it is of even multiplicity.

4 Evaluation of the Stability Boundaries

We will now show how the values of ω, for which the polynomial $P(\lambda,\omega)$ has a root $\lambda=\lambda_i$ of multiplicity $p>1$, can be found.

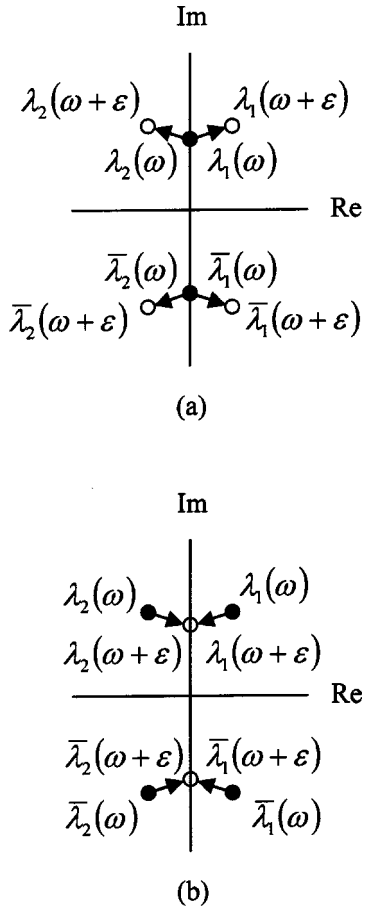


Fig. 4 Eigenvalue change due to a small perturbation in ω near the stability boundaries: (*a*) losing stability, and (*b*) gaining stability

These multiple roots determine the stability boundaries via Proposition 2.

By Proposition 3

$$P(\lambda,\omega)=\sum_{k=0}^{n}Q_{n-k}(\omega)\lambda^{2k} \tag{18}$$

where $Q_m(\omega)$ is a polynomial in ω of maximal degree $2m$. We differentiate (18) with respect to λ and obtain

$$\frac{\partial P}{\partial\lambda}=2\lambda\sum_{k=0}^{n-1}(k+1)Q_{n-k-1}(\omega)\lambda^{2k}\triangleq 2\lambda S(\lambda,\omega), \tag{19}$$

with the obvious definition of $S(\lambda,\omega)$. A necessary and sufficient condition for λ_i to be a nonsimple root of P for a fixed value of ω, is that λ_i is a root of both P and $\partial P/\partial\lambda$. Since, independent of ω, $\lambda=0$ is a root of $\partial P/\partial\lambda$, we conclude that each ω that satisfies $P(0,\omega)=0$ is a possible stability boundary. It thus follows from (2) and (3) that

Proposition 5. The positive eigenvalues λ_i of $\mathbf{K}-\lambda^2\mathbf{R}$ determine possible stability boundaries of (1).

Note that since $\mathbf{K}$ and $\mathbf{R}$ are symmetric and positive definite matrices, the roots λ^2 of $\det(\mathbf{K}-\lambda^2\mathbf{R})$ are all positive. This criterion is in principle equivalent to that of Afolabi [13], requiring the absolute term of the characteristic polynomial to vanish.

We now show how the other stability boundaries associated with the multiple roots $\lambda_i\neq 0$ of $P(\lambda,\omega)$ can be found. The problem under consideration is essentially one of solving simultaneously the two-variable polynomials $P(\lambda,\omega)=0$ and $S(\lambda,\omega)=0$ for their common roots λ_i and ω_i.

We denote the following bigradient matrix:

$$\mathbf{B}=\begin{bmatrix} Q_0 & Q_1 & \cdots & Q_n & & & & & & \\ & Q_0 & Q_1 & \cdots & Q_n & & & & & \\ & & \ddots & \ddots & \ddots & \ddots & & & & \\ & & & & Q_0 & Q_1 & Q_2 & \cdots & Q_n & \\ & & & & & nQ_0 & (n-1)Q_1 & \cdots & Q_{n-1} & \\ & & & & nQ_0 & (n-1)Q_1 & \cdots & Q_{n-1} & & \\ & & \ddots & \ddots & \ddots & \ddots & & & & \\ nQ_0 & (n-1)Q_1 & \cdots & Q_{n-1} & & & & & & \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\\a\end{matrix}}\right\} n-1 \text{ rows} \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} n \text{ rows} \end{matrix} . \tag{20}$$

Then, by elimination (see, e.g., [14]), λ is a common root of $P(\lambda,\omega)$ and $S(\lambda,\omega)$ if and only if $\det(\mathbf{B})=0$. This condition allows evaluation of all values of ω that determine the other possible stability boundaries associated with the repeated eigenvalues $\lambda_i \neq 0$.

There are two remaining issues. One involves a method of determining $Q_i(\omega)$ for $i=1,2,\ldots,n$. The other issue involves determining the values for ω which render $\det(\mathbf{B})=0$. We first address the second issue, assuming that $Q_i(\omega)$ are known.

We define the coefficients of $Q_k(\omega)$ as follows:

$$Q_k(\omega)=\sum_{j=0}^{k} q_j^{(k)}\omega^{2j}, \tag{21}$$

and denote $q_m^{(k)}=0$ when $m>k$. Then $\mathbf{B}$ can be written as

$$\mathbf{B}=\sum_{j=0}^{n} \omega^{2j}\mathbf{B}_j, \tag{22}$$

where $\mathbf{B}_j$ are constant matrices

$$\mathbf{B}_j=\begin{bmatrix} q_j^{(0)} & q_j^{(1)} & \cdots & q_j^{(n)} & & & & & \\ & q_j^{(0)} & q_j^{(1)} & \cdots & q_j^{(n)} & & & & \\ & & & \ddots & \ddots & \ddots & \ddots & & \\ & & & & & q_j^{(0)} & q_j^{(1)} & q_j^{(2)} & \cdots & q_j^{(n)} \\ & & & & & & nq_j^{(0)} & (n-1)q_j^{(1)} & \cdots & q_j^{(n-1)} \\ & & & & & nq_j^{(0)} & (n-1)q_j^{(1)} & \cdots & q_j^{(n-1)} & \\ & & & \iddots & \iddots & \iddots & \iddots & & & \\ nq_j^{(0)} & (n-1)q_j^{(1)} & \cdots & q_j^{(n-1)} & & & & & & \end{bmatrix} . \tag{23}$$

Note that the leading matrix in (22),

$$\mathbf{B}_n=\begin{bmatrix} \mathbf{O} & q_n^{(n)}\mathbf{I}_{n-1} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}, \tag{24}$$

is singular. The values of ω which lead to repeated eigenvalues $\lambda_i \neq 0$ of P are the finite eigenvalues of

$$(\mathbf{B}_0+\omega^2\mathbf{B}_1+\ldots+\omega^{2n}\mathbf{B}_n)\mathbf{v}=\mathbf{o}, \tag{25}$$

and we have the following.

Proposition 6. The positive eigenvalues λ_i be of the matrix pencil (25) determine possible stability boundaries of (1).

A first-order realization of (25) is given by

$$\begin{bmatrix} \mathbf{I} & & & \\ & \mathbf{I} & & \\ & & \ddots & \\ & & & \mathbf{B}_0 \end{bmatrix} \begin{pmatrix} \omega^{2n-2}\mathbf{v} \\ \omega^{2n-4}\mathbf{v} \\ \vdots \\ \mathbf{v} \end{pmatrix} - \omega^2 \begin{bmatrix} \mathbf{O} & \mathbf{I} & & \\ & \mathbf{O} & \ddots & \\ & & \ddots & \mathbf{I} \\ -\mathbf{B}_n & -\mathbf{B}_{n-1} & \cdots & -\mathbf{B}_1 \end{bmatrix} \begin{pmatrix} \omega^{2n-2}\mathbf{v} \\ \omega^{2n-4}\mathbf{v} \\ \vdots \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{o} \\ \mathbf{o} \\ \vdots \\ \mathbf{o} \end{pmatrix}. \tag{26}$$

Hence, the stability boundaries are determined by finding the eigenvalues of (26). Since $\mathbf{B}_n$ is singular the system (25) has some unbounded eigenvalues which can be eliminated by deflation.

The problem of evaluating the polynomials Q_k will now be addressed. We may choose an arbitrary value ω_i, for ω and solve

(1) for its eigenvalues $\lambda_1^{(i)}$, $\lambda_2^{(i)}$, . . . , $\lambda_{2n}^{(i)}$. These values determine $Q_k(\lambda_i)$, $k=0,1,\ldots,n$, via the system of linear equations:

$$\begin{bmatrix} (\lambda_1^{(i)})^{2n-2} & (\lambda_1^{(i)})^{2n-4} & \cdots & 1 \\ (\lambda_2^{(i)})^{2n-2} & (\lambda_2^{(i)})^{2n-4} & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ (\lambda_n^{(i)})^{2n-n} & (\lambda_n^{(i)})^{2n-4} & \cdots & 1 \end{bmatrix} \begin{pmatrix} Q_1(\omega_i) \\ Q_2(\omega_i) \\ \vdots \\ Q_n(\omega_i) \end{pmatrix} = \begin{pmatrix} -(\lambda_1^{(i)})^{2n} q_0^{(0)} \\ -(\lambda_2^{(i)})^{2n} q_0^{(0)} \\ \vdots \\ -(\lambda_n^{(i)})^{2n} q_0^{(0)} \end{pmatrix}. \tag{27}$$

where $q_0^{(0)}=\det(\mathbf{M})$. Repeating this process $n+1$ times for different values of ω_i allows determination of $Q_k(\omega_i)$ for $i=1,2,\ldots,n+1$. The polynomials $Q_k(\omega)$, $k=1,2,\ldots,n$ can be determined from these data by interpolation.

Note that the above method for determining $Q_k(\omega)$ is intended for the sake of clarity rather than computational accuracy. The fairly direct problem of reconstructing the coefficients of the polynomial (18) from its roots, which is expressed in terms of the Vandermonde system (27), can be determined alternatively by say the explicit Newton's formulas

$$\frac{Q_k(\omega_i)}{q_0^{(0)}}=(-1)^k N_k^{(i)}, \quad k=1,2,\ldots,n, \tag{28}$$

where $N_k^{(i)}$ is the sum of all $n!/(n-k)!k!$ products combining k factors $\lambda_j^{(i)}$ without repetition of subscripts. For large n, several extreme coefficients $Q_j(\omega_i)$ can be determined from (28) and then be used in (27) to reduce the dimension of the Vandermonde system.

5 Examples

Example 5.1. Consider a two-degree-of-freedom gyroscopic system with

$$\mathbf{M}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{G}=\begin{bmatrix} 0 & -5 \\ 5 & 0 \end{bmatrix}, \quad \mathbf{K}=\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$$

$$\text{and} \quad \mathbf{R}=\begin{bmatrix} 5 & -2 \\ -2 & 10 \end{bmatrix}. \tag{29}$$

The problem of evaluating the stability boundaries of this two-degree-of-freedom system is elementary, and can be solved analytically. We first solve the problem using basic principles, and then apply the method described in Section 4.

It follows from Eqs. (2) and (3) that

$$P(\lambda,\omega)=\det\left(\begin{bmatrix} \lambda^2-5\omega^2+4 & -5\lambda\omega+2\omega^2-2 \\ 5\lambda\omega+2\omega^2-2 & \lambda^2-10\omega^2+4 \end{bmatrix}\right)$$
$$=\lambda^4+(10\omega^2+8)\lambda^2+46\omega^4-52\omega^2+12, \tag{30}$$

and hence by Eq. (19)

$$\frac{\partial P}{\partial \lambda}=2\lambda(2\lambda^2+10\omega^2+8). \tag{31}$$

It thus follows that $\partial P/\partial\lambda=0$ if either $\lambda=0$, or $\lambda^2=-5\omega^2-4$.

Suppose $\lambda=0$. Then $P(\lambda,\omega)=0$ implies $46\omega^4-52\omega^2+12=0$, with the roots ± 0.5685, and ± 0.8985.

Suppose now that $\lambda^2=-5\omega^2-4$. Substituting this relation in (30) gives

$$P(\lambda,\omega)=21\omega^4-92\omega^2-4 \tag{32}$$

which has the roots ± 2.1033 and $\pm 0.2075i$.

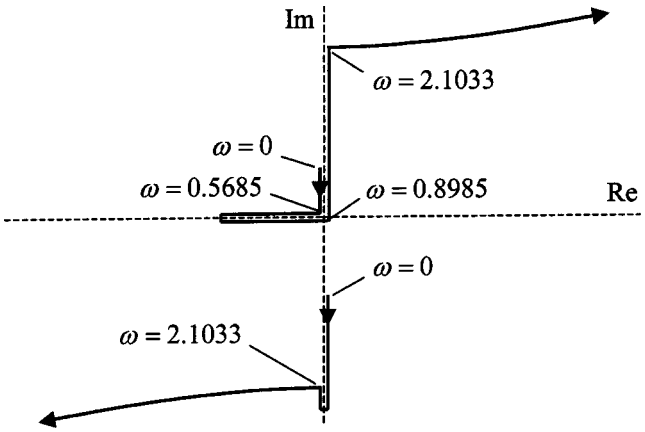


Fig. 5 Root loci for two nonconjugate eigenvalues

It thus follows from the analysis so far that the system of equations $P(\lambda,\omega)=0$ and $Q(\lambda,\omega)=0$ has three common positive roots for ω, namely,

$$0.5685, \quad 0.8985, \quad \text{and} \quad 2.1033. \tag{33}$$

These are the stability boundaries of the system. Perhaps the most important implication of this result is that for no other value of ω the system may lose or gain stability.

Root loci for two nonconjugate eigenvalues associated with this system are shown in Fig. 5. Inspection of this figure confirms that the system loses stability at $\omega=0.5685$, gains stability at $\omega=0.8985$, and again loses stability at $\omega=2.1033$.

We now solve the problem again using the method proposed in Section 4. By Proposition 5 the stability boundaries associated with the multiple eigenvalue $\lambda=0$ are determined by the eigenvalues of the pencil $\mathbf{K}-\lambda\mathbf{R}$. In our case the eigenvalues of

$$\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}-\lambda\begin{bmatrix} 5 & -2 \\ -2 & 10 \end{bmatrix} \tag{34}$$

are 0.3231 and 0.8073. Hence, two stability boundaries are $\sqrt{0.3231}=0.5685$ and $\sqrt{0.8073}=0.8985$. These are the same as the first two boundaries in (33).

Next we illustrate how to determine the other stability boundary associated with a multiple eigenvalue $\lambda\neq 0$. The first stage in the process requires evaluation of the polynomials $Q_i(\omega)$ for $i=0,1,2$. Note that by comparing (18) and (30) it becomes clear that these polynomials are

$$Q_0(\omega)=1, \quad Q_1(\omega)=10\omega^2+8,$$
$$\text{and} \quad Q_2(\omega)=46\omega^4-52\omega^2+12. \tag{35}$$

However, this solution is arrived at using analytical expansion of the determinant of $\mathbf{P}(\lambda,\omega)$. It should be noted that analytical expansion of the determinant of an $n\times n$ matrix involves an order of $n!$ basic numerical operations (i.e., multiplications and summations). Moreover, symbolic manipulation is required in our problem since the matrix pencil depends on two variables λ and ω. Consequently, such an approach cannot be applied for a practical system possessing modest dimension of say, $n=20$, for example. In what follows we illustrate how $Q_i(\omega)$ can be found numerically using the approach presented in Section 4, where no symbolic determinant expansion is required.

The zero-order polynomial is

$$Q_0=q_0=\det(\mathbf{M})=1. \tag{36}$$

For $\omega_1=0$ the eigenvalue problem (1) reduces to

$$(\lambda^2\mathbf{M}+\mathbf{K})\mathbf{v}=\mathbf{o}, \tag{37}$$

with eigenvalues $\pm\sqrt{2}i$ and $\pm\sqrt{6}i$. It follows that $\omega_1^2=0$, $(\lambda_1^{(1)})^2=-2$ and $(\lambda_2^{(1)})^2=-6$. Equation (27) for $i=1$ is, therefore,

$$\begin{bmatrix}(\lambda_1^{(1)})^2 & 1\\ (\lambda_2^{(1)})^2 & 1\end{bmatrix}\begin{pmatrix}Q_1(0)\\ Q_2(0)\end{pmatrix}=\begin{pmatrix}-(\alpha_1^{(1)})^2\\ -(\alpha_2^{(1)})^2\end{pmatrix} \tag{38}$$

or using numerical values

$$\begin{bmatrix}-2 & 1\\ -6 & 1\end{bmatrix}\begin{pmatrix}Q_1(0)\\ Q_2(0)\end{pmatrix}=\begin{pmatrix}-4\\ -36\end{pmatrix}, \tag{39}$$

which yields

$$Q_1(0)=8 \text{ and } Q_2(0)=12. \tag{40}$$

For $\omega_2=1$ the eigenvalue problem (1) is

$$(\lambda^2\mathbf{M}+\lambda\mathbf{G}+\mathbf{K}-\mathbf{R})\mathbf{v}=\mathbf{o}, \tag{41}$$

which has eigenvalues $\pm 0.5829i$ and $\pm 4.2024i$. It follows that $\omega_2=1$, $(\lambda_1^{(2)})^2=-0.3397$ and $(\lambda_2^{(2)})^2=-17.6603$, and from (27) we obtain

$$\begin{bmatrix}-0.3397 & 1\\ -17.6603 & 1\end{bmatrix}\begin{pmatrix}Q_1(1)\\ Q_2(1)\end{pmatrix}=\begin{pmatrix}-0.1154\\ -311.8846\end{pmatrix}, \tag{42}$$

which has the solution

$$Q_1(1)=18 \text{ and } Q_2(1)=6. \tag{43}$$

For $\omega_3=2$ the eigenvalue problem (1) is

$$(\lambda^2\mathbf{M}+2\lambda\mathbf{G}+\mathbf{K}-4\mathbf{R})\mathbf{v}=\mathbf{o} \tag{44}$$

with eigenvalues $\pm\sqrt{18}i$ and $\pm\sqrt{30}i$. Hence $\omega_3=2$, $(\lambda_1^{(3)})^2=-18$ and $(\lambda_2^{(3)})^2=-30$, and Eq. (27) gives

$$\begin{bmatrix}-18 & 1\\ -30 & 1\end{bmatrix}\begin{pmatrix}Q_1(2)\\ Q_2(2)\end{pmatrix}=\begin{pmatrix}-324\\ -900\end{pmatrix}, \tag{45}$$

which gives

$$Q_1(2)=48 \text{ and } Q_2(2)=540. \tag{46}$$

It is found by interpolation that the polynomial $Q_1(\omega)$ which satisfies $Q_1(0)=8$, $Q_1(1)=18$, (and $Q_1(2)=48$), is $Q_1(\omega)=10\omega^2+8$. The polynomial $Q_2(\omega)$ which satisfies $Q_2(0)=12$, $Q_2(1)=6$, and $Q_2(2)=540$, is $Q_2(\omega)=46\omega^4-52\omega^2+12$. These are the same results as obtained in (35).

Once $Q_i(\omega)$, $i=0,1,2$, is found we may evaluate the bigradient matrix $\mathbf{B}$ in (20). For our case $\mathbf{B}$ is reduced to

$$\mathbf{B}=\begin{bmatrix}Q_0 & Q_1 & Q_2\\ 0 & 2Q_0 & Q_1\\ 2Q_0 & Q_1 & 0\end{bmatrix}, \tag{47}$$

or using (35)

$$\mathbf{B}=\begin{bmatrix}1 & 10\omega^2+8 & 46\omega^4-52\omega^2+12\\ 0 & 2 & 10\omega^2+8\\ 2 & 10\omega^2+8 & 0\end{bmatrix}. \tag{48}$$

The third stability boundary is determined by the positive ω which makes $\mathbf{B}$ in (48) singular. We note in passing that analytical expansion of the determinant of $\mathbf{B}$ gives

$$\det(\mathbf{B})=-4(21\omega^4-92\omega^2-4), \tag{49}$$

a constant factor of the polynomial in (32), which confirms the result.

In order to determine numerically the values of ω which make (48) singular we write $\mathbf{B}$, as in (22) and (23),

$$\mathbf{B}=\sum_{k=0}^{2}\omega^{2k}\mathbf{B}_k=\omega^4\begin{bmatrix}0 & 0 & 46\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}+\omega^2\begin{bmatrix}0 & 10 & -52\\ 0 & 0 & 10\\ 0 & 10 & 0\end{bmatrix}+\begin{bmatrix}1 & 8 & 12\\ 0 & 2 & 8\\ 2 & 8 & 0\end{bmatrix}, \tag{50}$$

with the obvious definition of $\mathbf{B}_0$, $\mathbf{B}_1$, and $\mathbf{B}_2$. It thus follows from (26) that the third stability boundary is determine by the finite positive eigenvalue of

$$\begin{bmatrix}\mathbf{I}_3 & \mathbf{0}\\ \mathbf{0} & \mathbf{B}_0\end{bmatrix}-\omega^2\begin{bmatrix}0 & \mathbf{I}\\ -\mathbf{B}_2 & -\mathbf{B}_1\end{bmatrix}, \tag{51}$$

or explicitly

$$\begin{bmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&8&12\\0&0&0&0&2&8\\0&0&0&2&8&0\end{bmatrix}-\omega^2\begin{bmatrix}0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\\0&0&-46&0&-10&52\\0&0&0&0&0&-10\\0&0&0&0&-10&0\end{bmatrix}. \tag{52}$$

The matrix pencil (52) has four finite eigenvalues $\omega_{1,2}^2=4.4240$ and $\omega_{3,4}^2=-0.0431$. Hence, as in (33), the third stability boundary is $\sqrt{4.4240}=2.1033$.

Example 5.2. Consider the 6×6 system with

$$\mathbf{M}=\mathbf{I}_6,$$

$$\mathbf{G}=[g_{ij}]=\begin{cases}-5, & i=1,j=6\\ 5, & i=6,j=1\\ 0, & \text{otherwise,}\end{cases}$$

Table 1 Stability boundaries

Critical ω	Number of Poles: Imaginary	Real	Complex
0	12	0	0
0.4592			
0.6480	10	2	0
0.6627	8	4	0
0.7849	8	0	4
0.8234	8	4	0
0.8240	10	2	0
0.8911	6	2	4
0.8958	6	6	0
0.8986	8	4	0
0.9260	4	4	4
0.9261	4	8	0
0.9283	6	6	0
0.9384	2	6	4
0.9391	6	6	0
1.0000	4	8	0
2.3449	4	8	0
∞	0	8	4

$$\mathbf{K}=[k_{ij}]=\begin{cases} 4, & i=j, \\ -2, & i=j-1, j=2,3,\ldots,6 \\ -2, & i=j+1, j=1,2,\ldots,5 \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\mathbf{R}=[r_{ij}]=\begin{cases} 5, & i=j, j=1,2,\ldots,5 \\ -2, & i=j-1, j=2,3,\ldots,6 \\ -2, & i=j+1, j=1,2,\ldots,6 \\ 10, & i=j=6 \\ 0 & \text{otherwise.} \end{cases}$$

For this case we have

$$Q_0(\omega)=1$$

$$Q_1(\omega)=-10\omega^2+24$$

$$Q_2(\omega)=-20\omega^4-260\omega^2+220$$

$$Q_3(\omega)=180\omega^6+1320\omega^4-2380\omega^2+960$$

$$Q_4(\omega)=2021\omega^8-9624\omega^6+14736\omega^4-9184\omega^2+2016$$

$$Q_5(\omega)=-11050\omega^{10}+40368\omega^8-57932\omega^6-40608\omega^4 -13792\omega^2+1792$$

$$Q_6(\omega)=12286\omega^{12}-47316\omega^{10}+73676\omega^8-58944\omega^6 +25296\omega^4-5440\omega^2+448.$$

The critical values for ω obtained from (26) are

0.662724, 0.784963, 0.824002, 0.891162, 0.898645, 0.926033,

0.928360, 0.938422, 1.000000, 2.344929

and from Proposition 5

0.459267, 0.648012, 0.823424, 0.895866, 0.926168, 0.939133.

These values determine 17 intervals for $0<\omega<\infty$. Each interval contains a fixed number of purely imaginary poles, real poles, and complex poles. These values, shown in Table 1, indicate that the system is stable for $0<\omega<0.4592$ and unstable for other values of ω.

6 Conclusions

We have presented in this paper a numerical method for evaluating the critical values of the speed of rotation for which a conservative gyroscopic system may change its nature from stable to unstable and vice versa. Using this method one is able to determine all ranges of the angular speeds for which a gyroscopic system is stable by considering the stability of the system for a finite number of frequencies.

For an n degree-of-freedom gyroscopic system the process of evaluating the stability boundaries associated with the multiple eigenvalue $\lambda=0$ requires via Proposition 5 solving a generalized eigenvalue problem of dimension n for its eigenvalues. Evaluation of the other possible stability boundaries associated with a multiple eigenvalue $\lambda\neq 0$ requires via Proposition 6 solving an nth order matrix pencil of dimension $2n-1$ for its finite positive eigenvalues. Numerical examples have demonstrated these results.

Acknowledgment

The work of Y.M. Ram was supported in part by a National Science Foundation grant CMS-9978786. The work of M. M. Khonsari was supported in part by a Louisiana Board of Regents grant LEQSF(2000-03)-RD-B-02.

References

[1] Parker, R. J., 1998, ''On the Eigenvalues and Critical Speed Stability of Gyroscopic Continua,'' ASME J. Appl. Mech., **65**, pp. 134–140.

[2] Ahmadian, M., and Inman, D. J., 1986, ''Some Stability Results for General Linear Lumped-Parameter Dynamic Systems,'' ASME J. Appl. Mech., **53**, pp. 10–14.

[3] Barkwell, L., Lancaster, P., and Marcus, A. S., 1992, ''Gyroscopically Stabilized Systems: A Class of Quadratic Eigenvalue Problems With Real Spectrum,'' Can. J. Math., **44**, pp. 42–53.

[4] Barkwell, L., and Lancaster, P., 1992, ''Overdamped and Gyroscopic Vibrating Systems,'' ASME J. Appl. Mech., **59**, pp. 176–181.

[5] Duffin, R. J., 1955, ''The Rayleigh-Ritz Method for Dissipative or Gyroscopic Systems,'' Q. Appl. Math., **18**, pp. 215–221.

[6] Lancaster, P., and Zizler, P., 1998, ''On the Stability of Gyroscopic Systems,'' ASME J. Appl. Mech., **65**, pp. 519–522.

[7] Walker, J., 1988, ''Pseudodissipative Systems, I: Stability of Generalized Equilibria,'' ASME J. Appl. Mech., **55**, pp. 681–686.

[8] Walker, J. A., 1991, ''Stability of Linear Conservative Gyroscopic Systems,'' ASME J. Appl. Mech., **58**, pp. 229–232.

[9] Veselić, K., 1995, ''On the Stability of Rotating Systems,'' Z. Angew. Math. Mech., **75**, pp. 325–328.

[10] Hryniv, R., and Lancaster, P., 2001, ''Stabilization of Gyroscopic Systems,'' Z. Angew. Math. Mech., **81**, pp. 675–681.

[11] Seyranian, A. P., and Kliem, W., 2001, ''Bifurcations of Eigenvalues of Gyroscopic Systems With Parameters Near Stability Boundaries,'' ASME J. Appl. Mech., **68**, pp. 199–205.

[12] Seyranian, A., Stoustrup, J., and Kliem, W., 1995, ''On Gyroscopic Stabilization,'' Z. Angew. Math. Phys., **46**, pp. 255–267.

[13] Afolabi, D., 1995, ''Sylvester's Eliminant and Stability Criteria for Gyroscopic Systems,'' J. Sound Vib., **182**, pp. 229–244.

[14] Turnbull, H. W., 1944, *Theory of Equations*, Oliver and Boyd, Edinburgh, UK.

Constraint Forces and the Method of Auxiliary Generalized Speeds

S. Djerassi

H. Bamberger

Rafael,
P.O. Box 2250,
Haifa, Israel

This paper deals with noncontributing forces, usually called constraint forces or reaction forces, arising in simple, nonholonomic multibody systems. These forces are related to two kinds of constraints, namely, kinematical constraints—derived from kinematical requirements, and auxiliary constraints, introduced for the purpose of constraint forces determination. Here, the method of "auxiliary generalized speeds" is used to bring into evidence constraint forces related to the two kinds of constraints. It is shown that auxiliary generalized speeds can always be chosen in a way that gives rise to additional equations each having one measure number of a constraint force as an unknown. Motion equations can thus be generated and solved without regard to constraint forces determination; and constraint forces can be determined with no matrix inversion, at a minimal computational cost. [DOI: 10.1115/1.1572902]

Introduction

Let S be a system possessing n independent generalized speeds $u_1, \ldots, u_n$, whose motion is defined as unconstrained, and let μ simple, nonholonomic constraints be imposed on the motion of S. Also, let $R_1, \ldots, R_\mu$ be measure numbers of the related constraint forces. Then $n+\mu$ differential-algebraic equations (DAE) govern the motion of S, the unknowns being $\dot{u}_1, \ldots, \dot{u}_n$ and $R_1, \ldots, R_\mu$. Furthermore, if $u_{n-\mu+1}, \ldots, u_n$ comprise a choice of μ generalized speeds regarded as dependent on $u_1, \ldots, u_{n-\mu}$, then $\dot{u}_{n-\mu+1}, \ldots, \dot{u}_n$ and $R_1, \ldots, R_\mu$ can be eliminated, giving rise to a minimal set of $n-\mu$ ordinary differential equations (ODE) in $\dot{u}_1, \ldots, \dot{u}_{n-\mu}$ (see e.g., Shabana [1], Secs. 5.8–5.10). Generally speaking, DAE/ODE solutions become more (computationally) expensive as n and μ become larger. This is also true when such techniques as LU decomposition (Wehage and Haug [2] and Nikravesh and Haug [3]), QR decomposition (Kim and Vanderploeg [4]), zero-eigenvalue theorem (Kamman and Houston [5] and Loduha and Ravani [6]), singular value decomposition (Singh and Likins [7]), Householder transformation (Amirouche et al. [8]), or Graham-Schmidt orthogonalization process (Liang and Lance [9]) are used, in connection with DAE formulations, to deal with singular configurations through the introduction of the orthogonal complement matrix idea (first suggested by Hemami and Weimer [10]). Moreover, in accordance with the indicated state-of-the-art techniques of analytical mechanics, n and μ are frequently increased for the sole purpose of constraint forces determination, an observation raising the following question. Can the determination of these forces be done without affecting the computational cost of generation and solution of dynamical equations?

This question can be addressed effectively after two kinds of constraints are identified. One concerns *m kinematical constraints* such as constraints associated with the closure of kinematical loops, constraints associated with specified motions and constraints associated with the motion of rolling elements. The other kind concerns *M auxiliary constraints* coming into play if the associated constraint forces have to be determined. Examples are constraints associated with joints and constraints eliminating relative motions of particles of a rigid body. It may be concluded that, since $\mu = m + M$, the computational cost associated with the (unavoidable) introduction of constraints of the first kind increases by the introduction of constraints of the second kind. Many authors accept the additional cost, recognizing the importance of the determination of constraint forces—they are closely related to structural loads, and form a basis for the analytical treatment of friction forces. Moreover, a number of authors (Wehage and Haug [2], Nikravesh [11], and Amirouche et al. [8]), recommend that systems of $\overline{\nu}$ rigid bodies are temporarily regarded as "totally" unconstrained, so that $M = 6\overline{\nu} - n$. One can then determine, in conjunction with DAE/ODE-related methods, $m+M$ measure numbers of constraint forces, m associated with kinematical constraints and M associated with the joints connecting the rigid bodies to one another, making no distinction between the two kinds of constraints. This approach underlies the majority of the multibody programs presented, e.g., by Schiehlen [12]. A number of authors deal with constraint forces with less conventional methods. Papastavridis, using Appell's equations, [13], and Maggi's equations, [14], and Udvadia [15], using Gause's principle, generate constraint forces; however, from a computational point of view their approaches do not foretell a breakthrough.

The purpose of this work is to address the above question, showing that it is suffices to generate and solve n dynamical equations in conjunction with m constraint equations of the first kind without regard to constraint forces determination, and use, for this purpose, any technique helpful in avoiding singularities; and that m measure numbers of constraint forces of the first kind and M of the second kind can be exposed at a computational cost proportional to nm^2 and nM, respectively. This is done in the second and fourth sections for constraint forces of the first kind, and in the third section for constraint forces of the second kind. Two examples are used to illustrate the determination of a variety of measure numbers of constraint forces.

Constraint Forces of the First Kind—Preliminaries

Consider a simple, nonholonomic system S of ν particles P_i $(i=1, \ldots, \nu)$ of mass m_i possessing $\overline{n}$ generalized coordinates $q_1, \ldots, q_{\overline{n}}$ and n (where $n < \overline{n}$) generalized speeds $u_1, \ldots, u_n$ in N, a Newtonian reference frame. ($u_1, \ldots, u_n$ comprise linear combination of $\dot{q}_1, \ldots, \dot{q}_{\overline{n}}$, the coefficients being functions of $q_1, \ldots, q_{\overline{n}}$ and time t.) Suppose that the motion of S is *defined* as unconstrained, and that the velocity $\mathbf{v}^{P_i}$ of P_i $(i=1, \ldots, \nu)$ in N is written (Kane and Levinson [16], Sec. 2.14)

$$\mathbf{v}^{P_i} = \sum_{r=1}^{n} \mathbf{v}_r^{P_i} u_r + \mathbf{v}_t^{P_i} \quad (i=1, \ldots, \nu) \tag{1}$$

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Mar. 7, 2001; final revision, Oct. 10, 2002. Associate Editor: A. A. Ferri. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

where $\mathbf{v}_r^{P_i}$ and $\mathbf{v}_t^{P_i}$ $(i=1,\ldots,\nu, r=1,\ldots,n)$ are function of $q_1,\ldots,q_n^-$ and time t. Then its governing dynamical equations are (Kane and Levinson [16], Sec 6.1)

$$\left(\sum_{i=1}^{\nu}\mathbf{F}^{P_i}\cdot\mathbf{v}_r^{P_i}+\sum_{i=1}^{\nu}\mathbf{F}^{*P_i}\cdot\mathbf{v}_r^{P_i}=\right)F_r+F_r^*=0 \quad (r=1,\ldots,n) \tag{2}$$

where $\mathbf{F}^{P_i}$ and $\mathbf{F}^{*P_i}$ are, respectively, the resultant of all active forces and the inertia force acting on P_i, and F_r and F_r^* are the rth generalized active force and the rth generalized inertia force for S.

Let P_k be a particle of S, and let $\overline{P}_k$ be either a particle of S or a particle of R_B, a set of particles with a prescribed motion in N. Let $\mathbf{v}^{P_k}$ and $\mathbf{v}^{\overline{P}_k}$ be the velocities of P_k and $\overline{P}_k$ in N, respectively, and suppose P_k and $\overline{P}_k$ are momentarily or continuously in contact with each other. If the motion of m pairs of particles can be described similarly, then S is subject to m constraints of the first kind, indicating that

$$(\mathbf{v}^{P_k}-\mathbf{v}^{\overline{P}_k})\cdot\hat{\mathbf{a}}_k=f_k(q_1,\ldots,q_n^-,t) \quad (k=p+1,\ldots,n) \tag{3}$$

where $\hat{\mathbf{a}}_k$ $(k=p+1,\ldots,n)$ are unit vectors, f_k $(q_1,\ldots,q_n^-,t)$ $(k=p+1,\ldots,n)$ are known functions of $q_1,\ldots,q_n^-$ and t, and

$$p\triangleq n-m. \tag{4}$$

Similarly to $\mathbf{v}^{P_i}$ in Eqs. (1), $\mathbf{v}^{P_k}$ and $\mathbf{v}^{\overline{P}_k}$ are given by

$$\mathbf{v}^{P_k}=\sum_{r=1}^{n}\mathbf{v}_r^{P_k}u_r+\mathbf{v}_t^{P_k}, \quad \mathbf{v}^{\overline{P}_k}=\sum_{r=1}^{n}\mathbf{v}_r^{\overline{P}_k}u_r+\mathbf{v}_t^{\overline{P}_k}$$

(note that if $\overline{P}_k$ belongs to R_B then $\mathbf{v}_r^{\overline{P}_k}=0$ $(r=1,\ldots,n)$). Hence, Eqs. (3) become

$$\sum_{r=1}^{n}A_{kr}u_r+B_k=0 \quad (k=p+1,\ldots,n) \tag{5}$$

if A_{kr} and B_k are defined

$$A_{kr}\triangleq(\mathbf{v}_r^{P_k}-\mathbf{v}_r^{\overline{P}_k})\cdot\hat{\mathbf{a}}_k, \quad B_k\triangleq(\mathbf{v}_t^{P_k}-\mathbf{v}_t^{\overline{P}_k})\cdot\hat{\mathbf{a}}_k-f_k$$
$$(k=p+1,\ldots,n, \ r=1,\ldots,n). \tag{6}$$

Equations (5) comprise m independent linear relations between $u_1,\ldots,u_n$ called simple nonholonomic constraint equations (Kane and Levinson [16], Sec. 2.13). These equations govern the majority of cases in dynamics of constrained systems (problems discussed, e.g., by Shan [17] and Kitzka [18] represent exceptions governed by nonlinear relations between $u_1,\ldots,u_n$).

Suppose $u_1,\ldots,u_p$ and $u_{p+1},\ldots,u_n$ are regarded as independent variables and as dependent variables, respectively. Then Eqs. (5) can be solved for $u_{p+1},\ldots,u_n$, yielding

$$u_k=\sum_{r=1}^{p}C_{kr}u_r+D_k=0 \quad (k=p+1,\ldots,n). \tag{7}$$

It can be shown that, under these circumstances, the following equations govern the motion of the constrained system (Kane and Levinson [16], Secs. 4.4, 4.11, and 6.1):

$$F_r+F_r^*+\sum_{k=p+1}^{n}C_{kr}(F_k+F_k^*)=0 \quad (r=1,\ldots,p). \tag{8}$$

These equations can be solved for $\dot{u}_1,\ldots,\dot{u}_n$ in conjunction with

$$\sum_{r=1}^{n}A_{kr}\dot{u}_r+\sum_{r=1}^{n}\dot{A}_{kr}u_r+\dot{B}_k=0 \quad (k=p+1,\ldots,n), \tag{9}$$

derived from Eqs. (5). Alternatively, Eqs. (9) can be used to eliminate $\dot{u}_{p+1},\ldots,\dot{u}_n$ from Eqs. (8), leading to a minimal set of p dynamical equations in $\dot{u}_1,\ldots,\dot{u}_p$ (ODE formulations).

Constraint forces do not appear in Eqs. (8). However, they can be determined at will. To this end, reconsider Eqs. (2) and suppose that m pairs of contact forces $\mathbf{R}_k$ and $\overline{\mathbf{R}}_k$ $(k=p+1,\ldots,n)$ are exerted by $\overline{P}_k$ on P_k and by P_k on $\overline{P}_k$, respectively, in the $\hat{\mathbf{a}}_k$ direction, and that the action of these forces validates Eqs. (5). Then R_k can be defined as $R_k\triangleq\mathbf{R}_k\cdot\hat{\mathbf{a}}_k$, so that, in accordance with the law of action and reaction

$$\mathbf{R}_k=R_k\hat{\mathbf{a}}_k, \quad \overline{\mathbf{R}}_k=-R_k\hat{\mathbf{a}}_k \quad (k=p+1,\ldots,n). \tag{10}$$

These forces contribute to Eqs. (2), which become

$$F_r+F_r^*+\sum_{k=p+1}^{n}[(R_k\hat{\mathbf{a}}_k)\cdot\mathbf{v}_r^{P_k}+(-R_k\hat{\mathbf{a}}_k)\cdot\mathbf{v}_r^{\overline{P}_k}]=0$$
$$(r=1,\ldots,n)$$

(Kane and Levinson [16], Sec. 4.4) or, in view of Eqs. (6)

$$F_r+F_r^*+\sum_{k=p+1}^{n}A_{kr}R_k=0 \quad (r=1,\ldots,n). \tag{11}$$

Substitutions of $\dot{u}_1,\ldots,\dot{u}_n$ obtained from Eqs. (8) and (9) in the m last Eqs. (11) enable the determination of $R_{p+1},\ldots,R_n$. This involves the inversion of an $m\times m$ matrix having A_{kr} $(k,r=p+1,\ldots,n)$ as its entries (see Eq. (e) in the Appendix). However, this inversion is required for the generation of C_{kr} $(k=p+1,\ldots,n, r=1,\ldots,p)$ in Eqs. (7) and (8). It may thus be concluded that the determination of $R_{p+1},\ldots,R_n$ can be performed without matrix inversion. Alternatively, Eqs. (11) and (9) can be solved simultaneously for $\dot{u}_1,\ldots,\dot{u}_n$ and $R_{p+1},\ldots,R_n$ (DAE formulation).

Constraints of the first kind can be holonomic and/or simple, nonholonomic. Simple, nonholonomic constraint equations are obtained by substitutions in Eqs. (3). If constraint forces are *chosen* to be expressed as in Eqs. (10), then both f_k in Eqs. (3) and R_k in Eqs. (10) comprise measure numbers of vectors aligned with $\hat{\mathbf{a}}_k$. This choice, called by Blajer [19] "ideal," underlies the use of Eqs. (11) for the determination of $R_{p+1},\ldots,R_n$. Under these circumstances, the latter are identical to $\lambda_{p+1},\ldots,\lambda_n$, m Lagrange's multipliers (Whittaker, [20] and Pars [21]) that are introduced when the associated mathematical method (Gelfand and Fomin [22] and Lanczos [23]) is applied to constrained systems. By way of contrast, equations describing holonomic constraints are usually written $g_k(q_1,\ldots,q_n^-,t)=0$; and, although the equations $dg_k/dt=0$ are subsequently generated, no equations similar to Eqs. (3) are written, and the indicated relation between f_k and R_k $(k=p+1,\ldots,n)$ is lost. Thus, if $\mathbf{R}_k$ and $\overline{\mathbf{R}}_k$ are chosen to be parallel to $\hat{\mathbf{a}}_k'$ $(\hat{\mathbf{a}}_k'\neq\hat{\mathbf{a}}_k)$, then A_{kr} in Eqs. (11) must be replaced with A_{kr}', defined $A_{kr}'\triangleq(\mathbf{v}_r^{P_k}-\mathbf{v}_r^{\overline{P}_k})\cdot\hat{\mathbf{a}}_k'$ $(k=p+1,\ldots,n, r=1,\ldots,n)$. Hence $F_r'=\Sigma_{k=p+1}^{n}A_{kr}'R_k=\Sigma_{k=p+1}^{n}A_{kr}\lambda_k$ $(r=1,\ldots,n)$ and, with $\lambda_{p+1},\ldots,\lambda_n$ having been evaluated, the exposition of R_k $(k=p+1,\ldots,n)$ would require the inversion of an additional $m\times m$ matrix (having A_{kr}' as its entry in row k, column r). This inversion can be avoided if Eqs. (3) and (10) are also applied to holonomic constraints, a task involving no additional effort (the effort required to form velocities appearing in Eqs. (3) is comparable to that required to form the equations $dg_k/dt=0$). One can thus benefit from Eqs. (11) in connection with both holonomic and simple, nonholonomic constraints alike.

Constraint Forces of the Second Kind

Suppose that constraint forces $\mathbf{R}_k$ and $\overline{\mathbf{R}}_k$ exerted by $\overline{P}_k$ on P_k and by P_k on $\overline{P}_k$, respectively, have to be determined for $k=n+1,\ldots,n+M$. Then, in accordance with the state-of-the-art

technique of analytical dynamics, one has to reformulate the problem, introducing M additional motion variables—and M constraints—that bring into evidence the additional constraint forces. One then obtains equations structured as Eqs. (8) and (9) with $n+M$ replacing n.

The central claim of this work is closely related to Kane's method of auxiliary generalized speed (Kane and Levinson [16] Sec. 4.9) used to determine constraint forces. The claim for constraints of the second kind can be stated as follows. If constraints of the first kind have been imposed, and if, in accordance with the law of action and reaction, $\mathbf{R}_k$ and $\overline{\mathbf{R}}_k$ are expressed as

$$\mathbf{R}_k = R_k \hat{\mathbf{a}}_k \quad \overline{\mathbf{R}}_k = -R_k \hat{\mathbf{a}}_k \quad (k=n+1,\ldots,n+M), \tag{12}$$

then it is always possible to choose M auxiliary generalized speeds $\overline{u}_{n+1},\ldots,\overline{u}_{n+M}$ that give rise to M additional dynamical equations, each having one of R_k $(k=n+1,\ldots,n+M)$ as (the only) unknown. With this choice, the additional constraint equations are

$$\overline{u}_k = 0 \quad (k=n+1,\ldots,n+M) \tag{13}$$

and the associated dynamical equations are

$$F_r + R_r + F_r^* = 0 \quad (r=n+1,\ldots,n+M). \tag{14}$$

To show this, consider P_k and $\overline{P}_k$, and note that the relations between $\mathbf{v}^{P_k}$ and $\mathbf{v}^{\overline{P}_k}$, the velocities of P_k and $\overline{P}_k$, can always be written

$$(\mathbf{v}^{P_k} - \mathbf{v}^{\overline{P}_k}) \cdot \hat{\mathbf{a}}_k = f_k(q_1,\ldots,q_{\overline{n}},t) \tag{15}$$

where, as in Eqs. (3), $f_k(q_1,\ldots,q_{\overline{n}},t)$ is a known function of $q_1,\ldots,q_{\overline{n}}$ and t. The determination of R_k requires the removal of the constraint implied by Eq. (15). One way to accomplish the removal of this constraint is to add the component $\overline{u}_k \hat{\mathbf{a}}_k$ to the velocity of P_k, which, denoted now with an overbar, becomes

$$\overline{\mathbf{v}}^{P_k} = \mathbf{v}^{P_k} + \overline{u}_k \hat{\mathbf{a}}_k \quad (\overline{\mathbf{v}}_k^{P_k} = \hat{\mathbf{a}}_k). \tag{16}$$

$\mathbf{v}^{\overline{P}_k}$ is left intact, that is

$$\overline{\mathbf{v}}^{\overline{P}_k} = \mathbf{v}^{\overline{P}_k} \quad (\overline{\mathbf{v}}_k^{\overline{P}_k} = 0), \tag{17}$$

so that Eq. (15) becomes $(\overline{\mathbf{v}}^{P_k} - \overline{\mathbf{v}}^{\overline{P}_k}) \cdot \hat{\mathbf{a}}_k = f_k(q_1,\ldots,q_{\overline{n}},t)$ or

$$(\mathbf{v}^{P_k} + \overline{u}_k \hat{\mathbf{a}}_k - \mathbf{v}^{\overline{P}_k}) \cdot \hat{\mathbf{a}}_k = f_k(q_1,\ldots,q_{\overline{n}},t), \tag{18}$$

in agreement with Eqs. (16) and (17). Thus, the velocity of P_k relative to $\overline{P}_k$ in the $\hat{\mathbf{a}}_k$ direction (which is zero if $f_k=0$) increases by $\overline{u}_k \hat{\mathbf{a}}_k$. Expressions for the velocities of P_i $(i=1,\ldots,\nu)$ must be updated accordingly, i.e.,

$$\overline{\mathbf{v}}^{P_i} = \mathbf{v}^{P_i} + \overline{u}_k \mathbf{v}_k^{P_i} \quad (i=1,\ldots,\nu) \tag{19}$$

where $\mathbf{v}_k^{P_i}$ is a vector function of $q_1,\ldots,q_{\overline{n}}$ and t. Equations (15), (18), and (19) indicate that the motion of S is restored if the constraint equation

$$\overline{u}_k = 0 \tag{20}$$

is imposed on S, a step performed *after* an additional equation of motion is formulated (Kane and Levinson [16] Sec. 4.9 and Djerassi [24]), namely

$$F_k + R_k + F_k^* = 0. \tag{21}$$

F_k and F_k^* include contributions from all the active forces and inertia forces associated with S, whereas R_k is the contribution of $\mathbf{R}_k$ and $\overline{\mathbf{R}}_k$; for, by virtue of Eqs. (16) and (17) this contribution can be evaluated as follows:

$$(R_k \hat{\mathbf{a}}_k) \cdot \overline{\mathbf{v}}_k^{P_k} + (-R_k \hat{\mathbf{a}}_k) \cdot \overline{\mathbf{v}}_k^{\overline{P}_k} = (R_k \hat{\mathbf{a}}_k) \cdot \hat{\mathbf{a}}_k - 0 = R_k. \tag{22}$$

Also, $R_{p+1},\ldots,R_n$ contribute nothing to Eq. (21), since, by hypothesis, constraints of the first kind have been imposed; hence R_k is the only unknown in Eq. (21).

Equation (21) is valid for $k=n+1,\ldots,n+M$; for, the procedure leading to Eq. (21) can be used to determine each of $R_{n+1},\ldots,R_{n+M}$, one at a time. What remains to be shown, however, is that $\mathbf{R}_k$ and $\overline{\mathbf{R}}_k$ contribute only to Eq. (21), the kth equation. This can be done formally if the indicated procedure is applied to particles P_j and $\overline{P}_j$ $(j \neq k)$. Then equations corresponding to Eqs. (15)–(22) can be written with j replacing k. Specifically, equations corresponding to Eqs. (20) and (21) read

$$\overline{u}_j = 0 \tag{23}$$

$$F_j + R_j + F_j^* = 0. \tag{24}$$

Now, P_j and $\overline{P}_j$ are in contact with one another momentarily $(f_j \neq 0)$ or continually $(f_j = 0)$, therefore either both $\mathbf{v}^{P_j}$ and $\mathbf{v}^{\overline{P}_j}$, or none, are augmented by $\overline{u}_k \mathbf{g}_k$, where $\mathbf{g}_k$ is a vector function of $q_1,\ldots,q_{\overline{n}}$ and t. In both events $R_k \hat{\mathbf{a}}_k$ and $-R_k \hat{\mathbf{a}}_k$ contribute nothing to Eq. (24). By similarity, $R_j \hat{\mathbf{a}}_j$ and $-R_j \hat{\mathbf{a}}_j$ contribute nothing to Eq. (21). This conclusion, extended to any two pairs of particles $P_j - \overline{P}_j$ and $P_k - \overline{P}_k$ $(j,k=n+1,\ldots,n+M, j \neq k)$, validates Eqs. (13)–(14).

Note that $\overline{\mathbf{a}}^{P_i}$, defined as $\overline{\mathbf{a}}^{P_i} = {}^N d/dt(\overline{\mathbf{v}}^{P_i})$, reads $\overline{\mathbf{a}}^{P_i} = \mathbf{a}^{P_i} + \dot{\overline{u}}_k \mathbf{v}_k^{P_i} + \overline{u}_k \dot{\mathbf{v}}_k^{P_i}$ (where $\mathbf{a}^{P_i} = {}^N d/dt(\mathbf{v}^{P_i})$); hence the use of Eqs. (20) leads to $\overline{\mathbf{a}}^{P_i} = \mathbf{a}^{P_i}$ and $\overline{\mathbf{v}}^{P_i} = \mathbf{v}^{P_i}$. As a result, $\mathbf{F}^{P_i}$ and $\mathbf{F}^{*P_i}$, in general functions of $\mathbf{v}^{P_i}$ and $\mathbf{a}^{P_i}$, need not be reformulated following the introduction of $\overline{u}_k$ due to the subsequent use of Eqs. (20). It may thus be concluded that $\overline{u}_k$ in Eqs. (19) is introduced for the sole purpose of identifying $\mathbf{v}_k^{P_i}$ $(i=1,\ldots,\nu, k=n+1,\ldots,n+M)$. These are used to generate Eqs. (21) by substitution in equations similar to Eqs. (2), with $\mathbf{v}_k^{P_i}$ replacing $\mathbf{v}_r^{P_i}$, thus leaving intact Eqs. (2) (and, specifically, $\mathbf{F}^{P_i}$ and $\mathbf{F}^{*P_i}$ $(i=1,\ldots,\nu)$) and hence Eqs. (8).

Also note that the M pairs of particles can include duplicates of the same physical particles (e.g., P_k and $\overline{P}_k$, $k=n+1$, $n+2$, $n+3$), in which event the indicated conclusion holds if the associated unit vectors, i.e., $\hat{\mathbf{a}}_k$ $(k=n+1,n+2,n+3)$, are mutually perpendicular. Finally note that Eqs. (13) and (14) apply if subsets of particles comprise rigid bodies. Then $\overline{u}_k$ may be a measure number of the angular velocity of a rigid body, and R_k would be the associated measure number of a constraint torque of a couple (Djerassi [24], Sec. 3).

Next it is shown that auxiliary generalized speeds can be defined in connection with constraints of the first kind, which give rise to expressions each of which contain one of $R_{p+1},\ldots,R_n$ as an unknown.

Constraint Forces of the First Kind Revisited

Suppose that m auxiliary variables $\overline{u}_k$ $(k=p+1,\ldots,n)$ are defined as

$$\overline{u}_k \triangleq -(\mathbf{v}^{P_k} - \mathbf{v}^{\overline{P}_k}) \cdot \hat{\mathbf{a}}_k + f_k(q_1,\ldots,q_{\overline{n}},t) \quad (k=p+1,\ldots,n) \tag{25}$$

so that

$$(\mathbf{v}^{P_k} + \overline{u}_k \hat{\mathbf{a}}_k - \mathbf{v}^{\overline{P}_k}) \cdot \hat{\mathbf{a}}_k = f_k(q_1,\ldots,q_{\overline{n}},t) \quad (k=p+1,\ldots,n), \tag{26}$$

an expression resembling Eqs. (18). It follows, in view of Eqs. (6), that

$$\overline{u}_k + \sum_{r=1}^{n} A_{kr} u_r + B_k = 0 \quad (k=p+1,\ldots,n), \tag{27}$$

equations that can be solved for u_k $(k=p+1,\ldots,n)$, yielding

$$u_k=\sum_{r=1}^{p} C_{kr}u_r+D_k+\sum_{r=p+1}^{n} E_{kr}\overline{u}_r \quad (k=p+1,\ldots,n). \tag{28}$$

The manner in which C_{kr} and E_{kr} are related to A_{kr} can best be described by a matrix representation, as in Eqs. (*b*) of the Appendix. Djerassi and Kane [25] show that, if $u_1,\ldots,u_p,\overline{u}_{p+1},\ldots,\overline{u}_n$ replace $u_1,\ldots,u_p,u_{p+1},\ldots,u_n$ as independent variables, then the associated equations of motion are

$$F_r+F_r^*+\sum_{k=p+1}^{n} C_{kr}(F_k+F_k^*)=0 \quad (r=1,\ldots,p) \tag{29}$$

$$\sum_{k=p+1}^{n} E_{kr}(F_k+F_k^*)=0 \quad (r=p+1,\ldots,n). \tag{30}$$

Equations (30) are identical to equations obtained if the last *m* of Eqs. (11) are solved for $R_{p+1},\ldots,R_n$. Moreover, Eqs. (27) and (5) lead to the equations

$$\overline{u}_k=0 \quad (k=p+1,\ldots,n) \tag{31}$$

which play the role of constraint equations resembling Eqs. (13). When these constraints are imposed on Eqs. (29) and (30), then the following results are obtained. Equations (29) become a minimal set of dynamical equations in *n* unknowns $\dot{u}_1,\ldots,\dot{u}_n$ (as are Eqs. (8)), and Eqs. (30) are replaced with

$$\sum_{k=p+1}^{n} E_{kr}(F_k+F_k^*)+R_r=0 \quad (r=p+1,\ldots,n). \tag{32}$$

These equations are identical to those obtained if the *m* last of Eqs. (11) are solved for $R_{p+1},\ldots,R_n$. R_r, the only unknown in the *r*th of Eqs. (32) (an equation associated with $\overline{u}_r$), is the contribution of $\mathbf{R}_r$ and $\overline{\mathbf{R}}_r$ $(r=p+1,\ldots,n)$ defined in Eqs. (10), since, in connection with the *r*th of Eqs. (26),

$$(R_r\hat{\mathbf{a}}_r)\cdot\overline{\mathbf{v}}_r^{P_k}+(-R_r\hat{\mathbf{a}}_r)\cdot\overline{\overline{\mathbf{v}}}_r^{P_k}=(R_r\hat{\mathbf{a}}_r)\cdot\hat{\mathbf{a}}_r-0=R_r, \tag{33}$$

where $\overline{\mathbf{v}}_r^{P_k}$ and $\overline{\overline{\mathbf{v}}}_r^{P_i}$ are given by expressions similar to those in Eqs. (16) and (17) with *r* replacing *k* (an alternative proof of Eqs. (32) is given in the Appendix). Equations (32) can be obtained straightforwardly if the velocities of P_i $(i=1,\ldots,\nu)$ are expressed in terms of $u_1,\ldots,u_p,\overline{u}_{p+1},\ldots,\overline{u}_n$ with the aid of Eqs. (28), namely

$$\overline{\mathbf{v}}^{P_i}=\sum_{r=1}^{p}\mathbf{v}_r^{P_i}\mathbf{u}_r+\sum_{r=p+1}^{n}\overline{\mathbf{v}}_r^{P_i}\overline{\mathbf{u}}_r+\overline{\mathbf{v}}_t^{P_i} \quad (i=1,\ldots,\nu). \tag{34}$$

Then substitutions in the last *m* of Kane's equations (Kane and Levinson [16]) which, in the present context, read

$$\sum_{i=1}^{\nu}(\mathbf{F}^{P_i}+\mathbf{F}^{*P_i})\cdot\overline{\mathbf{v}}_r^{P_i}+\sum_{k=p+1}^{n}\mathbf{R}_k\cdot(\overline{\mathbf{v}}_r^{P_k}-\overline{\overline{\mathbf{v}}}_r^{P_k})=0$$

$$(r=p+1,\ldots,n)$$

lead to equations identical with Eqs. (32).

Each of Eqs. (32) and (14) has one measure number of a constraint force as the only unknown. Having obtained these equations, one is free to solve Eqs. (8) (or (29)) and (9) for $\dot{u}_1,\ldots,\dot{u}_n$, disregarding constraint forces, and invoking, in the case of singular configurations (see, e.g., Kamman and Houston [5], Blajer et al. [26], and Singh and Likins [7]) the orthogonal complement matrix technique. (The latter is related to Eqs. (29) and (32) in a manner presented in the Appendix.) With $\dot{u}_1,\ldots,\dot{u}_n$ in hand, one can evaluate $R_{p+1},\ldots,R_n$ in Eqs. (32) and $R_{n+1},\ldots,R_{n+M}$ in Eqs. (14) at a cost proportional to nM and nm^2, respectively. For, each of $F_r+F_r^*$ $(p+1,\ldots,n)$ in Eqs. (32) and of $F_r+F_r^*$ $(r=n+1,\ldots,n+M)$ in Eqs. (14) comprises a linear combination of $\dot{u}_1,\ldots,\dot{u}_n$, and can be evaluated at a cost proportional to *n*. One can thus generate and solve dynamical equations irrespective of reaction forces evaluation, and evaluate reaction forces at the minimum possible cost.

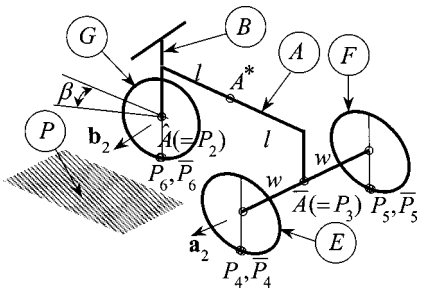


Fig. 1 Tricycle

The steps required to form Eqs. (32) and (14) can be summarized as follows:

1. For $k=p+1,\ldots,p+m(=n)$ (constraints of the first kind)

1.1 Consider $2m$ particles P_k and $\overline{P}_k$ with reference to Eqs. (3) and define $2m$ constraint forces in accordance with Eqs. (10);

1.2 Introduce $\overline{u}_k$ as in Eqs. (27) and solve Eqs. (27) for u_k as in Eqs. (28);

1.3 Use Eqs. (28) to replace u_k with $\overline{u}_k$ in expressions for $\mathbf{v}^{P_i}$ $(i=1,\ldots,\nu)$, and obtain $\overline{\mathbf{v}}^{P_i}$ $(i=1,\ldots,\nu)$ as in Eqs. (34);

1.4 Generate Eqs. (32) using $\overline{\mathbf{v}}^{P_i}$ $(i=1,\ldots,\nu)$, $\overline{\mathbf{v}}^{P_k}$ and $\overline{\overline{\mathbf{v}}}^{P_k}$;

1.5 Use Eqs. (31) to eliminate $\overline{u}_k$ and $\dot{\overline{u}}_k$ from Eqs. (32) and solve the latter for R_k.

2 For $k=n+1,\ldots,n+M$ (constraints of the second kind)

2.1 Consider $2M$ particles P_k and $\overline{P}_k$ with reference to Eqs. (15) and define $2M$ constraint forces in accordance with Eqs. (12);

2.2 Add $\overline{u}_k\mathbf{a}_k$ to $\mathbf{v}^{P_k}$, obtaining $\overline{\mathbf{v}}^{P_k}$ and $\overline{\overline{\mathbf{v}}}^{P_k}$, as in Eqs. (16) and (17);

2.3 Use Eqs. (16) and (17) to obtain corresponding expressions for $\overline{\mathbf{v}}^{P_i}$ $(i=1,\ldots,\nu)$, as in Eqs. (19);

2.4 Generate Eqs. (14) using $\overline{\mathbf{v}}^{P_i}$ $(i=1,\ldots,\nu)$, $\overline{\mathbf{v}}^{P_i}$ and $\overline{\overline{\mathbf{v}}}^{P_k}$;

2.5 Use Eqs. (13) to eliminate $\overline{u}_k$ and $\dot{\overline{u}}_k$ from Eqs. (14) and solve the latter for R_k.

The idea of auxiliary variables is not new (Hamel [27] and Shan [28]). However, the procedure just discussed becomes possible due to Kane's unique method of auxiliary generalized speed. It can be applied manually to simple problems; however its application to larger systems requires multibody packages that permit the introduction of auxiliary generalized speeds. One such package is Autolev (Kane and Levinson [29]) used to generate the following examples.

Examples

A Tricycle Moving on a Horizontal Plane. Figure 1 shows a tricycle comprising a central body *A*, two supporting wheels *E* and *F*, and a steering wheel *G* attached to *A* via fork *B*. The wheels rotate freely about their respective axes. P_i $(i=4,5,6)$ are points of *E*, *F*, and *G*, respectively, momentarily in contact with $\overline{P}_i$ $(i=4,5,6)$, points of a horizontal plane *P*. $\overline{A}$ is the midpoint of the centers E^* and F^* of *E* and *F*, $\hat{A}$ is the center of *G* and A^* is the mass center of *A*. $\mathbf{a}_i$, $\mathbf{b}_i$ and $\mathbf{n}_i$ $(i=1,2,3)$ are triads of dextral, mutually perpendicular unit vectors fixed in *A*, *B*, and *N*, respectively, such that $\mathbf{a}_3$, $\mathbf{b}_3$, and $\mathbf{n}_3$ are all perpendicular to *P*. $\mathbf{a}_2$ is aligned with the axes of *E* and *F* and $\mathbf{b}_2$ is aligned with the axis of *G*. *l* and *w* are lengths of segments shown in Fig. 1, and *h* is the height of A^* above *P*. Finally, M_A and I_{A3} are the mass of *A* and its central moment of inertia for $\mathbf{a}_3$; and $D\mathbf{a}_1$ is a driving force exerted on A^*.

Define u_r $(r=1,2,3)$ as

$$u_k \triangleq \mathbf{v}^{A^*}\cdot\mathbf{a}_r \quad (r=1,2), \quad u_3 \triangleq \boldsymbol{\omega}^A\cdot\mathbf{n}_3 \tag{35}$$

where $\boldsymbol{\omega}^A$ and $\mathbf{v}^{A^*}$ are the angular velocity of A in N, and the velocity of A^* in N. Then $\boldsymbol{\omega}^A$ and $\mathbf{v}^{A^*}$ related to motions of A defined as unconstrained are given by

$$\boldsymbol{\omega}^A \triangleq u_3\mathbf{n}_3 \quad \mathbf{v}^{A^*}=u_1\mathbf{a}_1+u_2\mathbf{a}_2. \tag{36}$$

Assuming that the fork and the wheels are massless, one can show, by substitutions in Eqs. (1) (with $n=3$), that the following equations govern unconstrained motions of A:

$$D-M_A(\dot{u}_1-u_2u_3)=0, \quad -M_A(\dot{u}_2+u_1u_3)=0, \quad -I_{A3}\dot{u}_3=0. \tag{37}$$

If E, F, and G roll without slip on P, then the motion of S proceeds such that $\mathbf{v}^{\bar{A}}\cdot\mathbf{a}_2=0$ and $\mathbf{v}^{\hat{A}}\cdot\mathbf{b}_2=0$, where $\mathbf{v}^{\bar{A}}$ and $\mathbf{v}^{\hat{A}}$ are the velocities of $\bar{A}$ and $\hat{A}$, respectively. Substitutions give rise to the following two constraint equations, structured as Eqs. (5) (with $m=2$), namely

$$u_2-lu_3=0, \quad -u_1s_\beta+(u_2+lu_3)c_\beta=0 \tag{38}$$

where $s_\beta \triangleq \mathrm{Sin}(\beta)$ and $c_\beta \triangleq \mathrm{Cos}(\beta)$, and β is the steering angle.

Next, let $\mathbf{R}_2$ and $\mathbf{R}_3$ be constraint forces exerted on $\hat{A}$ and $\bar{A}$ in the $\mathbf{b}_2$ and $\mathbf{a}_2$ directions, so that $\hat{A}$ and $\bar{A}$ play the roles of P_2 and P_3, respectively, and $\mathbf{R}_2=R_2\mathbf{b}_2$ and $\mathbf{R}_3=R_3\mathbf{a}_2$ (Step 1.1). Also, let the vertical components of forces exerted by $\bar{P}_i$ on P_i $(i=4,5,6)$ be $\mathbf{R}_i=R_i\mathbf{n}_3$ $(i=4,5,6)$; hence $\bar{\mathbf{R}}_i=-R_i\mathbf{n}_3$ $(i=4,5,6)$ and $M=3$ (Step 2.1). It is required to determine R_2, R_3, R_4, R_5, and R_6.

Equations (38) are satisfied if $\mathbf{R}_2$ and $\mathbf{R}_3$ are exerted on $\hat{A}$ and $\bar{A}$, respectively, in which event Eqs. (37) give way to the following equations:

$$\begin{gathered} D-R_2s_\beta-M_A(\dot{u}_1-u_2u_3)=0 \\ R_3+R_2c_\beta-M_A(\dot{u}_2+u_1u_3)=0 \\ -R_3l+R_2lc_\beta-I_{A3}\dot{u}_3=0, \end{gathered} \tag{39}$$

in accordance with Eqs. (11). To avoid inversion associated with the solutions of Eqs. (39) for R_2 and R_3, define $\bar{u}_2$ and $\bar{u}_3$ as $\bar{u}_2=-(u_2-lu_3)$ and $\bar{u}_3=-[-u_1s_\beta+(u_2+lu_3)c_\beta]$ (Step 1.2); then Steps 1.3–1.5 lead to

$$\begin{gathered} R_2-[(m_p\dot{u}_1+m_Au_1^2/l)t_\beta+m_p\dot{\beta}u_1/c_\beta^2]/(4c_\beta)=0 \\ R_3[(m_m\dot{u}_1+m_Au_1^2/l)t_{gb}+m_m\dot{\beta}u_1/c_\beta^2]/4=0 \end{gathered} \tag{40}$$

where $m_m=M_A-I_{A3}/l^2$ and $m_p=M_A+I_{A3}/l^2$. Now, the identification of R_4, R_5, and R_6 require a choice of $\bar{u}_4$, $\bar{u}_5$, and $\bar{u}_6$ leading, in accordance with Step 2.2, to

$$\bar{\mathbf{v}}^{P_r}=\mathbf{v}^{P_r}+\bar{u}_r\mathbf{n}_3 \quad (r=4,5,6). \tag{41}$$

This is the case if $\bar{\boldsymbol{\omega}}^A$ and $\bar{\mathbf{v}}^{A^*}$ are expressed as

$$\begin{aligned} \bar{\boldsymbol{\omega}}^A &= u_3\mathbf{a}_3+\bar{u}_4[1/(2w)\mathbf{a}_1+1/(4l)\mathbf{a}_2]+\bar{u}_5[-1/(2w)\mathbf{a}_1 \\ &\quad +1/(4l)\mathbf{a}_2]+\bar{u}_6[-1/(2l)\mathbf{a}_2] \\ \bar{\mathbf{v}}^{A^*} &= u_1\mathbf{a}_1+u_2\mathbf{a}_2+\bar{u}_4[h/(4l)\mathbf{a}_1-h/(2w)\mathbf{a}_2+1/4\mathbf{a}_3] \\ &\quad +\bar{u}_5[h/(4l)\mathbf{a}_1+h/(2w)\mathbf{a}_2+1/4\mathbf{a}_3]+\bar{u}_6[-h/(2l)\mathbf{a}_1 \\ &\quad +1/2\mathbf{a}_3] \end{aligned} \tag{42}$$

in accordance with Step 2.3. Then Steps 2.4–2.5 lead to

$$\begin{gathered} R_4+R/2+m_Ah/w[2u_1(\dot{\beta}/c_\beta^2+s)+2t_\beta\dot{u}_1)]/8=0 \\ R_5+R/2-m_Ah/w[2u_1(\dot{\beta}/c_\beta^2+s)+2t_\beta\dot{u}_1)]/8=0 \\ R_6-m_Ag-R=0 \end{gathered} \tag{43}$$

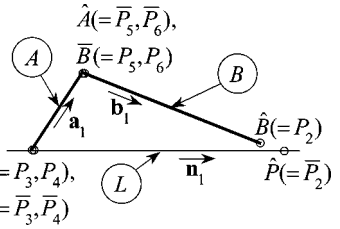


Fig. 2 Crank and slider mechanism

where $t_\beta=\tan(\beta)$, $s=t_\beta u_1/l$ and $R=hF/2l-m_Ag/2+m_Ah(s^2-4\dot{u}_1/l)/8$.

The following comment concerning the generation of Eqs. (42) is in order. First, it can be verified that Eqs. (42) give rise to Eqs. (41) if they are used in the following expressions:

$$\bar{\mathbf{v}}^{P_r}=\bar{\mathbf{v}}^{A^*}+\bar{\boldsymbol{\omega}}^A\times\mathbf{p}^{A^*/P_r} \quad (r=4,5,6), \tag{44}$$

where $\mathbf{p}^{A^*/P_r}$ is the position vector from A^* to P_r. Second, it follows from Eqs. (44) that $\bar{\boldsymbol{\omega}}_r^A$ and $\bar{\mathbf{v}}_r^{A^*}$, the coefficients of $\bar{u}_r$ $(r=4,5,6)$ in Eqs. (42), are related to each other as

$$\bar{\mathbf{v}}_r^{P_r}=\bar{\mathbf{v}}_r^{A^*}+\bar{\boldsymbol{\omega}}_r^A\times\mathbf{p}^{A^*/P_r} \quad (r=4,5,6) \tag{45}$$

where $\mathbf{p}^{A^*/P_r}$ $(r=4,5,6)$ (Fig. 1) and $\bar{\mathbf{v}}_r^{P_r}$ (Eq. (41)) are known, as are the *directions* of $\bar{\mathbf{v}}_r^{A^*}$ and $\bar{\boldsymbol{\omega}}_r^A$ $(r=4,5,6)$ (for $r=6$ $\mathbf{p}^{A^*/P_6}=l\mathbf{a}_1-h\mathbf{a}_3$, $\bar{\mathbf{v}}_6^{P_6}=\mathbf{a}_3$, $\bar{\boldsymbol{\omega}}_6^A\|\mathbf{a}_2$ and $\bar{\mathbf{v}}_6^{A^*}\|h\mathbf{a}_1-l\mathbf{a}_3$, etc.). Moreover, the vectors $\bar{\mathbf{v}}_r^{P_r}$, $\bar{\mathbf{v}}_r^{A^*}$ and $\bar{\boldsymbol{\omega}}_r^A\times\mathbf{p}^{A^*/P_r}$ appearing in Eqs. (45) are all perpendicular to $\bar{\boldsymbol{\omega}}_r^A$. Hence, each of Eqs. (45) gives rise to two scalar equations in two unknowns, namely, the magnitudes of $\bar{\mathbf{v}}_r^{A^*}$ and $\bar{\boldsymbol{\omega}}_r^A$. The latter can thus be determined and used to form Eqs. (42).

This example is relatively complex. Usually, things are simpler, as in the next example.

Crank and Slider Mechanism. Figure 2 shows bars A and B of lengths a and b, comprising a crank-and-slider mechanism. $\bar{A}$ and $\hat{A}$, $\bar{B}$ and $\hat{B}$, and $\bar{P}$ and $\hat{P}$, are the endpoints of A and B and points fixed in N, respectively. $\mathbf{a}_i$, $\mathbf{b}_i$ and $\mathbf{n}_i$ $(i=1,2,3)$ are triads of dextral, mutually perpendicular unit vectors fixed in A, B, and N, respectively, such that $\mathbf{a}_3$, $\mathbf{b}_3$, and $\mathbf{n}_3$ are all perpendicular to P, the plane of motion. $\mathbf{a}_1$ and $\mathbf{b}_1$ are aligned with $\bar{A}$ and B, and $\mathbf{n}_1$ is aligned with L, the line passing through $\bar{P}$ and $\hat{P}$. Finally, let self-explanatory notation stand for the inertial properties of A and B.

Define q_r and u_r $(r=1,2)$ as

$$q_1 \triangleq \mathrm{Cos}^{-1}(\mathbf{a}_1\cdot\mathbf{n}_1), \quad q_2 \triangleq \mathrm{Cos}^{-1}(\mathbf{b}_1\cdot\mathbf{n}_1), \quad u_r \triangleq \dot{q}_r(r=1,2). \tag{46}$$

Then $\boldsymbol{\omega}^A$ and $\boldsymbol{\omega}^B$, the angular velocities of A and B, can be expressed

$$\boldsymbol{\omega}^A=u_1\mathbf{n}_3, \quad \boldsymbol{\omega}^B=u_2\mathbf{n}_3. \tag{47}$$

Suppose that $\bar{A}$ coincides with $\bar{P}$ at all times and that, in this configuration, the motion of the mechanism is defined as unconstrained. Then the following equations

$$\begin{gathered} -[I_A+(M_A/4+M_B)a^2]\dot{u}_1-abM_Bc_{12}\dot{u}_2/2-abM_Bs_{12}u_2^2/2=0 \\ -0.5abM_Bc_{12}\dot{u}_1-[I_B+M_Bb^2/4]\dot{u}_2+abM_Bs_{12}u_1^2/2=0, \end{gathered} \tag{48}$$

where $s_{12} \triangleq \mathrm{Sin}(q_1-q_2)$ etc., govern unconstrained motions of the mechanism.

Constrained motions require that $\hat{B}$ slides along L, i.e., that $\mathbf{v}^{\hat{B}} \cdot \mathbf{n}_2 = 0$, or

$$ac_1u_1 + bc_2u_2 = 0 \tag{49}$$

where $c_1 \triangleq \mathrm{Cos}(q_1)$, etc.

Next, let $\mathbf{R}_2$, given by $\mathbf{R}_2 = R_2\mathbf{n}_2$, be a force exerted on $\hat{B}$, which plays the role of P_2 (Step 1.1). Let $\mathbf{R}_3$ and $\mathbf{R}_4$, given by $\mathbf{R}_3 = R_3\mathbf{n}_1$ and $\mathbf{R}_4 = R_4\mathbf{n}_2$, be forces exerted on $\overline{A}$, which plays the roles of P_3 and P_4, by $\overline{P}$, playing the roles of $\overline{P}_3$ and $\overline{P}_4$; and let $\mathbf{R}_5$ and $\mathbf{R}_6$, given by $\mathbf{R}_5 = R_5\mathbf{n}_1$ and $\mathbf{R}_6 = R_6\mathbf{n}_2$, be forces exerted by $\hat{A}$—which plays the roles of $\overline{P}_5$ and $\overline{P}_6$, on $\overline{B}$, playing the roles of P_5 and P_6 (Step 2.1). It is required to determine R_2, R_3, R_4, R_5, and R_6.

Defining $\overline{u}_2 = -(ac_1u_1 + bc_2u_2)$ (Step 1.2) and carrying out Steps 1.3–1.5, one has

$$R_2 - ma(-c_1\dot{u}_1 + as_1u_1^2 + t_2u_2^2) + M_Ba(c_{12}\dot{u}_1 + s_{12}u_1^2)/(2c_2) = 0, \tag{50}$$

where $m = (I_B/b^2 + M_B/4)/c_2^2$. Moreover, choosing $\overline{u}_3$, $\overline{u}_4$, $\overline{u}_5$, and $\overline{u}_6$ such that, in accordance with Step 2.2

$$\overline{\boldsymbol{\omega}}^A = \boldsymbol{\omega}^A, \quad \overline{\mathbf{v}}^A = \overline{u}_3\mathbf{n}_1 + \overline{u}_4\mathbf{n}_2,$$
$$\overline{\mathbf{v}}^B = \mathbf{v}^B + \overline{u}_3\mathbf{n}_1 + \overline{u}_4\mathbf{n}_2 + \overline{u}_5\mathbf{n}_1 + \overline{u}_6\mathbf{n}_2$$

one has, following Step 2.3, $\overline{\mathbf{v}}^{A^*} = \overline{\mathbf{v}}^A + \boldsymbol{\omega}^A \times a/2\mathbf{a}_1$, etc. Finally one obtains, taking Steps 2.4–2.5, four additional equations with R_3, R_4, R_5, and R_6 as unknowns. The sixth of these reads

$$R_6 - a(c_1m - s_1t_2M_B/2)\dot{u}_1 + pm + ac_1t_2M_Bu_1^2/2 = 0 \tag{51}$$

where $p = as_1u_1^2 + bs_2u_2^2$. Note that

$$\overline{\mathbf{v}}^P = \mathbf{v}^P(=0), \quad \overline{\mathbf{v}}^{\hat{A}} = \mathbf{v}^{\hat{A}} + \overline{u}_3\mathbf{n}_1 + \overline{u}_4\mathbf{n}_2 \tag{52}$$

hence $\overline{\mathbf{v}}^A$ and $\overline{\mathbf{v}}^P$ do not include $\overline{u}_5$ and $\overline{u}_6$, whereas $\overline{\mathbf{v}}^{\hat{A}}$ and $\overline{\mathbf{v}}^B$ both include $\overline{u}_3$ and $\overline{u}_4$, in agreement with the statement following Eq. (24).

Conclusions

A new methodology for the determination of noncontributing, constraint forces, based upon the idea of auxiliary generalized speeds, was presented. The methodology imposes uniformity in the formulation of constraint equations for both kinds of constraints (Eqs. (3) and (15)), and in the definition of constraint forces (Eqs. (10) and (12)) and auxiliary generalized speeds (Eqs. (26) and (18), and Eqs. (31) and (13)). It specifically requires that constraints be defined as relationships between velocities of particles, and that constraints and constraint forces be colinear vectors. Finally, the methodology de-couples the generation of the dynamical equations from that of reaction forces, enabling solutions of minimal-dimension sets of dynamical equations (in DAE/ODE formulations); and, avoiding matrix inversions, minimizes the computational effort required to identify reaction forces.

Appendix

On the Relationships of Eqs. (28) and (31) With Other Formulations. Let the following matrices be defined with the aid of quantities used earlier:

$$_{p\times 1}u \underset{(7)}{=} |u_1 \quad \dots \quad u_p|^T, \quad _{m\times 1}u \underset{(7)}{=} |u_{p+1} \quad \dots \quad u_n|^T,$$

$$_{m\times 1}\overline{u} \underset{(27)}{=} |\overline{u}_{p+1} \quad \dots \quad \overline{u}_n|^T$$

$$_{m\times p}A \underset{(5)}{=} \begin{vmatrix} A_{p+1,1} & \dots & A_{p+1,p} \\ \dots & & \\ A_{n,1} & \dots & A_{n,p} \end{vmatrix},$$

$$_{m\times m}A \underset{(5)}{=} \begin{vmatrix} A_{p+1,p+1} & \dots & A_{p+1,n} \\ \dots & & \\ A_{n,p+1} & \dots & A_{n,n} \end{vmatrix}$$

$$_{m\times n}A \underset{(5)}{=} |_{m\times p}A|_{m\times m}A|, \quad _{m\times 1}R \underset{(11)}{=} |R_{p+1} \quad \dots \quad R_n|^T,$$

$$_{m\times 1}B \underset{(27)}{=} |B_{p+1} \quad \dots \quad B_n|^T$$

$$_{p\times 1}F \underset{(11)}{=} |F_1 + F_1^* \quad \dots \quad F_p + F_p^*|^T,$$

$$_{m\times 1}F \underset{(11)}{=} |F_{p+1} + F_{p+1}^* \quad \dots \quad F_n + F_n^*|^T$$

$$_{m\times m}E \underset{(28)}{=} \begin{vmatrix} E_{p+1,p+1} & \dots & E_{p+1,n} \\ \dots & & \\ E_{n,p+1} & \dots & E_{n,n} \end{vmatrix},$$

$$_{m\times p}C \underset{(28)}{=} \begin{vmatrix} C_{p+1,1} & \dots & C_{p+1,p} \\ \dots & & \\ E_{n,1} & \dots & E_{n,p} \end{vmatrix},$$

where lower left indices play the double role of matrix identification and dimension indication, and where numbers under equal signs refer to equations numbered correspondingly. Then Eqs. (27) can be written as $_{m\times 1}\overline{u} + _{m\times p}A_{p\times 1}u + _{m\times m}A_{m\times 1}u + _{m\times 1}B = 0$ and, solved for $_{m\times 1}u$, they yield

$$_{m\times 1}u = -(_{m\times m}A)^{-1}(_{m\times p}A_{p\times 1}u + _{m\times 1}B) - (_{m\times m}A)^{-1}\,_{m\times 1}\overline{u}. \tag{a}$$

The coefficients of $_{p\times 1}u$ and $_{m\times 1}\overline{u}$ in Eqs. (28) are the entries of $_{m\times p}C$ and $_{m\times m}E$, respectively, hence by comparison with Eq. (*a*)

$$_{m\times p}C = -(_{m\times m}A)^{-1}\,_{m\times p}A, \quad _{m\times m}E = -(_{m\times m}A)^{-1}. \tag{b}$$

The m last of Eqs. (11) can be cast in a matrix form, reading

$$_{m\times 1}F + (_{m\times m}A)^T\,_{m\times 1}R = 0. \tag{c}$$

The left-hand side of Eq. (*c*) should replace $_{m\times 1}F$ in Eqs. (30) when written in a matrix form, so as to include contributions of constraint forces, that is

$$(_{m\times m}E)^T[_{m\times 1}F + (_{m\times m}A)^T\,_{m\times 1}R] = 0 \tag{d}$$

or, in view of Eq. (*b*),

$$(_{m\times m}A)^{-T}\,_{m\times 1}F + _{m\times 1}R = 0, \tag{e}$$

a matrix form of Eqs. (32). Also, the matrix form of Eqs. (29) is, in view of Eqs. (*b*),

$$_{p\times 1}F - (_{m\times p}A)^T(_{m\times m}A)^{-T}\,_{m\times 1}F = 0. \tag{f}$$

Defining the following matrices

$$_{n\times m}A' \triangleq |_{m\times p}O| - (_{m\times m}A)^{-T}|^T,$$
$$_{n\times p}A'' \triangleq |_{p\times p}I| - (_{m\times p}A)^T(_{m\times m}A)^{-T}|^T \tag{g}$$

where $_{m\times p}O$ and $_{p\times p}I$ are null and unit matrices, respectively, one can replace Eqs. (*e*) and (*f*) with

$$(_{n\times m}A')^T\,_{n\times 1}F - _{m\times 1}R = 0, \quad (_{n\times p}A'')^T\,_{n\times 1}F = 0. \tag{h}$$

These equations can equivalently be obtained by a premultiplication of the equation $_{n\times 1}F + (_{m\times n}A)^T\,_{m\times 1}R = 0$, a matrix form of Eqs. (11), with $(_{n\times m}A')^T$ and $(_{n\times p}A'')^T$, respectively. The columns of $_{n\times m}A'$ span the same space as do rows of $_{m\times n}A$, whereas $(_{n\times p}A'')^T$ is an orthogonal complement of $(_{m\times n}A)^T$

$[({}_{n\times p}A'')^T({}_{m\times n}A)^T = {}_{p\times m}O]$, hence its columns comprise a base for the null-space of ${}_{m\times n}A$. Starting with Maggi's equations, Kurdila et al. [30] arrive at identically structured equations, and use them as a basis for comparison of different ways, indicated in the Introduction, to obtain orthogonal complement matrices.

References

[1] Shabana, A., 1998, *Dynamics of Multibody Systems*, Cambridge University Press, New York.

[2] Wehage, R. A., and Haug, E. J., 1982, "Generalized Coordinates Partitioning for Dimension Reduction Analysis of Constrained Dynamical Systems," ASME J. Mech. Des., **104**, pp. 247–255.

[3] Nikravesh, P. E., and Haug, E. J., 1983, "Generalized Coordinate Partitioning for Analysis of Mechanical Systems With Nonholonomic Constraints," ASME J. Mech., Transm., Autom. Des., **105**, pp. 379–384.

[4] Kim, S. S., and Vanderploeg, M. J., 1986, "QR Decomposition for State Space Representation of Constrained Mechanical Systems," ASME J. Mech., Transm., Autom. Des., **108**, pp. 183–188.

[5] Kamman, J. W., and Huston, R. L., 1984, "Dynamics of Constrained Multibody Dynamics," ASME J. Appl. Mech., **51**, pp. 899–903.

[6] Loduha, T. A., and Ravani, B., 1995, "On First-Order Decoupling of Equations of Motion for Constrained Dynamical Systems," ASME J. Appl. Mech., **62**, pp. 216–222.

[7] Singh, R. P., and Likins, P. W., 1985, "Singular Value Decomposition for Constrained Dynamical Systems," ASME J. Appl. Mech., **52**, pp. 943–948.

[8] Amirouche, F. M. L., Jia, T., and Ider, S. K., 1988, "A Recursive Householder Transformation for Complex Dynamical Systems With Constraints," ASME J. Appl. Mech., **55**, pp. 729–734.

[9] Liang, C. D., and Lance, G. M., 1987, "A Differentiable Null-Space Method for Constrained Dynamic Analysis," ASME J. Mech., Transm., Autom. Des., **109**, pp. 405–411.

[10] Hemami, H., and Weimer, F. C., 1981, "Modelling of Nonholonomic Dynamic Systems With Applications," ASME J. Appl. Mech., **48**, pp. 177–182.

[11] Nikravesh, P. E., 1990, "Systematic Reduction of Multibody Equations of Motion to a Minimal Set," Int. J. Non-Linear Mech., **25**(2/3), pp. 143–151.

[12] Schiehlen, W., Ed., 1990, *Multibody Systems Handbook*, Springer-Verlag, Berlin.

[13] Papastavridis, J. G., 1988, "On the Nonlinear Appell's Equations and the Determination of Generalized Reaction Forces," Int. J. Eng. Sci., **26**(6), pp. 609–625.

[14] Papastavridis, J. G., 1990, "Maggi's Equations of Motion and the Determination of Constraint Reactions," J. Guid. Control Dyn., **13**(2), pp. 213–220.

[15] Udvadia, E. F., and Kalaba, E. R., 1996, *Analytical Dynamics: A New Approach*, Cambridge University Press, New York.

[16] Kane, T. R., and Levinson, D. A., 1985, *Dynamics: Theory and Applications*, McGraw-Hill, New York.

[17] Shan, D., 1973, "Equations of Motion of Systems With Second-Order Nonlinear Nonholonomic Constraints," PMM, **37**(2), pp. 349–354.

[18] Kitzka, F., 1986, "An Example for the Application of a Nonlinear Constraint of Second Order in Particles Mechanics," Z. Angew. Math. Mech., **66**(7), pp. 312–314.

[19] Blajer, W., 1992, "A Projection Method Approach to Constrained Dynamic Analysis," ASME J. Appl. Mech., **59**, pp. 643–649.

[20] Whittaker, E. T., 1944, *A Treatise on the Analytical Dynamics*, Dover, New York.

[21] Pars, L. A., 1979, *Treatise on Analytical Dynamics*, Ox Bow Press, Woodbridge, CT (reprint from the original 1965 edition).

[22] Gelfand, I. M., and Fomin, S. V., 1963, *Calculus of Variations*, Prentice-Hall, Englewood Cliffs, NJ.

[23] Lanczos, C., 1970, *The Variational Principles of Mechanics*, 4th Ed., Dover, New York.

[24] Djerassi, S., 1997, "Determination of Noncontributing Forces and Noncontributing Impulses in Three-Phase Motions," ASME J. Appl. Mech., **64**, pp. 582–589.

[25] Djerassi, S., and Kane, T. R., 1985, "Equations of Motion Governing the Deployment of a Flexible Linkage From a Spacecraft," J. Astronaut. Sci., **33**(4), pp. 417–428.

[26] Blajer, W., Bestle, D., and Schiehlen, W., 1994, "An Orthogonal Complement Matrix Formulation for Constrained Multibody Systems," ASME J. Mech. Des., **116**, pp. 423–428.

[27] Hamel, G., 1949, *Theoretische Mechanik*, Springer-Verlag, Berlin.

[28] Shan, D., 1975, "On the Determination of Forces of Constraints Reaction," PMM, **39**(6), pp. 1129–1134.

[29] Kane, T. R., and Levinson, D. A., 1996, *Dynamics Online: Theory and Implementation With Autolev*, Online Dynamics, Sunnyvale, CA.

[30] Kurdila, A., Papastavridis, J. G., and Kamat, M. P., 1990, "Role of Maggi's Equations in Computational Methods for Constrained Multibody Systems," J. Guid. Control Dyn., **13**(1), pp. 113–120.

A Dynamic Generalized Self-Consistent Model for Wave Propagation in Particulate Composites

R.-B. Yang
Department of Aeronautical Engineering,
Feng Chia University,
Taichung 40724, Taiwan
e-mail: rbyang@fcu.edu.tw

Wave propagations in an inhomogeneous medium (e.g., voids, particles, defects, inclusions) undergo multiple scattering which results in a frequency-dependent velocity and attenuation of coherent wave. The aim of this study is to analyses multiple scattering of plane compressional and shear waves in a composite containing randomly distributed spherical inclusions in a homogenous isotropic medium. To calculate effective wave numbers of ultrasonic waves propagating in the heterogeneous material, a generalized self-consistent multiple scattering model is used in this study. Numerical results for the effective phase velocity and attenuation of both P and SV waves are calculated for a wide range of frequencies and concentrations. The proposed dynamic generalized self-consistent model for particulate composites recovers both well-known static effective moduli in the static limit and the results at higher frequencies and concentrations agree well with published experimental data. [DOI: 10.1115/1.1576806]

1 Introduction

Ultrasonic waves are widely used in the field of nondestructive evaluation (NDE) of composite materials. Advances towards the development of new ultrasonic quantitative NDE techniques have been made possible by the study of propagation and diffraction of elastic waves in such materials. Waves propagating in an inhomogeneous medium undergo multiple scattering which results in a frequency-dependent velocity and attenuation of the coherent waves. The overall dynamic response of the medium may be conveniently modeled by means of the complex wave number, $\langle k \rangle$, describing the coherent wave propagating in an equivalent homogeneous material given by

$$\langle k \rangle = \frac{\omega}{V(\omega)} + i\alpha(\omega), \tag{1}$$

where $V(\omega)$ and $\alpha(\omega)$ denote the phase velocity and attenuation of the average waves, respectively. This effective wave number $\langle k \rangle$ is related to the overall elastodynamic constants which play an important role in ultrasonic nondestructive evaluation of composite materials. For example, in order to determine the response of the inhomogeneous medium that is subjected to transient loads or incident ultrasonic waves, the information of the overall dynamic properties are considered indispensable.

The problem of the propagation of multiple scattered waves in a random distribution of three-dimensional inclusions has been studied extensively in the literature. Foldy [1] and Waterman and Truell [2] studied the multiple scattering of waves by point scatterers and developed a statistical averaging procedure to estimate the phase velocity and attenuation of the coherent waves in terms of the microstructures and constituent material properties of the inhomogeneous mediums. Sayers and Smith [3], Datta et al. [4], Shindo et al. [5], and O'Neill et al. [6] have used the similar multiple scattering approach to predict the ultrasonic wave speeds and attenuation of particulate composites. In these methods, the multiple scattering formulas yield the effective wave numbers of the average wave in terms of the forward and backward scattering amplitudes of an isolated inclusion. The results are accurate provided the correlations and interactions between individual inclusions can be ignored. It seems to indicate that these results are valid at low frequencies and at low concentrations. However, an experimentally observed fact in the scattering of electromagnetic waves by a dense distribution of discrete scatterers is that the assumption of independent scattering leads to overestimation of scattering effects ([7]). As it would be shown in this paper, the zero-frequency limits of these estimates do not reduce to their well-known static limits. Hence, it is not clear whether these predicted results are very accurate at low frequencies.

Mal and Bose [8] have studied analytically the scattering of plane waves by spherical elastic inclusions which are arbitrarily distributed in an infinite matrix medium. The propagation characteristics of the average wave were obtained by a statistical approach through the introduction of a pair-correlation function and the use of the quasicrystlline approximation. A similar approach was used by Datta [9], Willis [10,11], and Varadan et al. [12]. These multiple scattering models reduce to the same form of the long wavelength limits which are identical to the lower bound of Hashin and Shtrikman [13]. However, at higher concentration, the correct pair-correlation function is difficult to obtain and the effective dynamic constants may be quite sensitive to the choice of the pair-correlation function. In general, multiple scattering theories can accurately predict the overall properties at low frequencies and concentrations.

To account for multiple scattering at high concentrations, a dynamic self-consistent scheme was used by Sabina and Willis [14] and Kim et al. [15] for the approximate analysis of elastic waves in random particulate composites. The frequency-dependent effective phase velocity and coherent attenuation can be obtained by an iterative method. As indicated by Sabina and Willis [14], the self-consistent expression of dynamic effective properties is valid when the relevant elastic waves have wavelengths at least as great as $4a$, where a is the radius of the spherical inclusions. Both aforementioned works based on the self-consistent scheme reduced to the static properties given by Hill [16] and Budiansky [17] in the static limit.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, June 5, 2002; final revision, Dec. 17, 2002. Associate Editor: A. K. Mal. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Chair, Department of Mechanics 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.

There are several micromechanical models (e.g., differential method, composite spheres model, self-consistent method, generalized self-consistent model, Mori-Tanaka method) that can be used to estimate the overall static elastic moduli of the composite materials. Although all models recover dilute behavior adequately, Christensen [18] have concluded that only the generalized self-consistent model gives physically reasonable results at high concentrations and covers the full range of volume fractions, $1 \leqslant c \leqslant 1$. Theoretical estimates of the effective dynamic moduli based on homogenization and other methods have been found to be unsatisfactory at higher frequencies and particle concentrations. To account for multiple scattering at high concentration, Yang and Mal [19,20] have developed a generalized self-consistent multiple scattering model, which combines the generalized self-consistent model (GSCM) together with the Waterman and Truell's statistical approach to calculate the effective dynamic moduli of a fiber-reinforced composite. The comparison of theoretical prediction of this model with measured wave velocity data by Huang et al. [21] for a longitudinal wave in the SiC/Ti fiber/matrix composite showed excellent agreement.

The aim of the present study is to analyses multiple scattering of plane longitudinal or shear waves due to a random distribution of elastic spherical inclusions in a homogenous elastic medium. We consider a particulate composite, which consists of a homogeneous, isotropic matrix medium containing randomly distributed spherical particles of identical properties. The generalized self-consistent multiple scattering model is studied for the spherical inclusion case. Unlike wave propagation in the fiber-reinforced composites can be decoupled into antiplane (SH) and in-plane (P and SV) cases, the study of multiple scattering of elastic waves in a particulate composite becomes relatively complicated by the inherent coupling of plane longitudinal and shear waves in the dynamic GSCM. In this study, a similar mathematical treatment of wave diffraction by a single spherical inclusion developed by Shindo et al. [5] was adopted and modified to the dynamic GSCM. Numerical results are presented for silicon carbide/aluminum and lead/epoxy particulate composites. The effective phase velocity and attenuation of the coherent waves are then calculated for a wide range of frequencies and concentrations. The theoretical results of dispersion and attenuation are compared with the experimental results of Kinra et al. [22,23] as well as the theoretical results by Waterman and Truell [2] and Sabina and Willis [14]. In the Raleigh limit, the proposed method recovers both well-known static effective moduli of particulate composites by Hashin's composite sphere model ([24]) and Christensen and Lo's GSCM ([25]), which have not been obtained by other multiple scattering methods and formulations.

2 Scattering of P and S Waves in the Generalized Self-Consistent Model (GSCM)

We consider a random distribution of identical spherical inclusions of radius a embedded in an isotropic and infinite matrix. Let λ_1, μ_1, ρ_1 be the Lamé constants and mass density of the inclusion and λ_2, μ_2, ρ_2 those of the matrix. The geometry is depicted in Fig. 1 where (x,y,z) is the Cartesian coordinate system and (r,θ,φ) is the spherical coordinate system. Let the components of the displacement vector $\mathbf{u}$ in the r, θ, φ directions be labeled by u_r, u_θ, and u_ϕ. The displacement equation of motion is

$$(\lambda+2\mu)\nabla\nabla\cdot\mathbf{u}-\mu\nabla\times\nabla\times\mathbf{u}=\rho\frac{\partial^2\mathbf{u}}{\partial t^2}. \tag{2}$$

The generalized self-consistent multiple scattering model is also shown in Fig. 1. The spherical inclusion of radius a is embedded in a shell of the matrix material of outer radius b, which in turn is embedded in an infinite medium possessing the unknown effective Lamé constants $\langle\lambda\rangle$, $\langle\mu\rangle$ and effective density $\langle\rho\rangle$. The material properties and the field variables in the inclusion and matrix will be identified by the indices 1 and 2. The corresponding

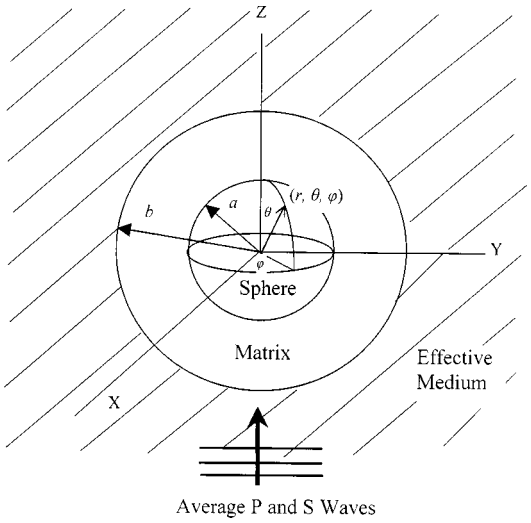


Fig. 1 The generalized self-consistent multiple scattering model

quantities in the effective medium will be identified by an angular bracket $\langle\ \rangle$. The radius of the matrix b is related to the volume fraction c of inclusions by

$$c=\frac{a^3}{b^3}. \tag{3}$$

Although the effective density varied with frequency for composites with a strong density contrast in the constituents, it is appropriate to assume that the effective density is the mean density for small density contrast and can be given by

$$\langle\rho\rangle=c\rho_1+(1-c)\rho_2. \tag{4}$$

Suppose that a time-harmonic plane longitudinal (P) wave propagating in the positive z-direction and a plane shear (S) wave polarized in the x-direction and propagating in the positive z-direction. Thus,

$$\mathbf{u}^i=w_0e^{i\langle k\rangle z-i\omega t}\mathbf{e}_z+u_0e^{i\langle K\rangle z-i\omega t}\mathbf{e}_x, \tag{5}$$

where a superscript i stands for the incident waves, ω is the circular frequency and w_0, u_0 are the amplitudes of the longitudinal and shear waves. $\langle k\rangle$ and $\langle K\rangle$ are the average wave numbers of the P and S waves in the effective medium, respectively. Then, suppressing the time factor $e^{-i\omega t}$ and dropping the angular brackets in the effective properties, i.e., replacing $\langle k\rangle$ by k and $\langle K\rangle$ by K, the total displacement fields in the effective medium, matrix and the spherical inclusion may be expressed in the forms:

$$\mathbf{u}=\mathbf{u}^i+\sum_{n=0}^{\infty}\sum_{m=-n}^{n}[A_{mn}\mathbf{L}_{mn}^{(3)}(kr)+B_{mn}\mathbf{M}_{mn}^{(3)}(Kr)+C_{mn}\mathbf{N}_{mn}^{(3)}(Kr)],\quad r>b \tag{6}$$

$$\mathbf{u}=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}[A'_{mn}\mathbf{L}_{mn}^{(3)}(k_2r)+B'_{mn}\mathbf{M}_{mn}^{(3)}(K_2r)+C'_{mn}\mathbf{N}_{mn}^{(3)}(K_2r)+D_{mn}\mathbf{L}_{mn}^{(1)}(k_2r)+E_{mn}\mathbf{M}_{mn}^{(1)}(K_2r)+F_{mn}\mathbf{N}_{mn}^{(1)}(K_2r)],\quad a<r<b \tag{7}$$

$$\mathbf{u}=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}[A_{mn}^0\mathbf{L}_{mn}^{(1)}(k_1r)+B_{mn}^0\mathbf{M}_{mn}^{(1)}(K_1r)+C_{mn}^0\mathbf{N}_{mn}^{(1)}(K_1r)],\quad r<a, \tag{8}$$

where $A_{mn,}$, B_{mn}, C_{mn}, A'_{mn}, B'_{mn}, C'_{mn}, D_{mn}, E_{mn}, F_{mn}, A^0_{mn}, B^0_{mn}, and C^0_{mn} are in general complex and must determined from the interface conditions. Spherical vector wave functions $\mathbf{L}^{(3)}_{mn}$, $\mathbf{M}^{(3)}_{mn}$, and $\mathbf{N}^{(3)}_{mn}$, are given by

$$\mathbf{L}^{(3)}_{mn}(kr)=\left[\mathbf{e}_r\frac{\partial}{\partial r}h_n(kr)P^m_n(\cos\theta)+\mathbf{e}_\theta h_n(kr)\frac{1}{r}\frac{\partial}{\partial\theta}P^m_n(\cos\theta)+\mathbf{e}_\varphi\frac{im}{r\sin\theta}h_n(kr)P^m_n(\cos\theta)\right]e^{im\phi} \tag{9}$$

$$\mathbf{M}^{(3)}_{mn}(Kr)=\left[\mathbf{e}_\theta\frac{im}{\sin\theta}h_n(Kr)P^m_n(\cos\theta)-\mathbf{e}_\varphi h_n(Kr)\frac{\partial}{\partial\theta}P^m_n(\cos\theta)\right]e^{im\phi} \tag{10}$$

$$N^{(3)}_{mn}(Kr)=\left\{\mathbf{e}_r\frac{n(n+1)}{r}h_n(Kr)P^m_n(\cos\theta)+\mathbf{e}_\theta\frac{1}{r}\frac{\partial}{\partial r}[rh_n(Kr)]\frac{\partial}{\partial\theta}P^m_n(\cos\theta)+\mathbf{e}_\varphi\frac{im}{r\sin\theta}\frac{\partial}{\partial r}[rh_n(Kr)]P^m_n(\cos\theta)\right\}e^{im\phi}, \tag{11}$$

where h_n is the nth-order spherical Hankel function of the first kind and P^m_n is the associated Legendre function of the first kind. $\mathbf{L}^{(1)}_{mn}$, $\mathbf{M}^{(1)}_{mn}$, and $\mathbf{N}^{(1)}_{mn}$ are obtained from Eqs. (9)–(11) by replacing h_n with the nth-order spherical Bessel function of the first kind j_n. The boundary conditions require the displacement and stress vector to vary continuously across the interfaces. The stress components may be found from the corresponding displacements through the stress-displacement relations

$$\frac{\sigma_{rr}}{\mu}=\frac{\lambda}{\mu}\frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial r}(r^2u_r\sin\theta)+\frac{\partial}{\partial\theta}(ru_\theta\sin\theta)+\frac{\partial}{\partial\phi}(ru_\phi)\right]+2\frac{\partial u_r}{\partial r} \tag{12}$$

$$\frac{\sigma_{r\theta}}{\mu}=\frac{1}{r}\frac{\partial u_r}{\partial\theta}+\frac{\partial u_\theta}{\partial r}-\frac{u_\theta}{r} \tag{13}$$

$$\frac{\sigma_{r\phi}}{\mu}=\frac{1}{r\sin\theta}\frac{\partial u_r}{\partial\phi}+\frac{\partial u_\phi}{\partial r}-\frac{u_\phi}{r}. \tag{14}$$

From the continuity conditions at $r=a$ and $r=b$, the relationships among $A_{mn,}$, C_{mn}, A'_{mn}, C'_{mn}, D_{mn}, F_{mn}, A^0_{mn}, and C^0_{mn} are found to be

$$\mathbf{P}\begin{Bmatrix}A_{mn}\\C_{mn}\end{Bmatrix}+\tilde{\mathbf{P}}\begin{Bmatrix}\Phi_{mn}\\X_{mn}\end{Bmatrix}=\mathbf{P}^{(2)}\begin{Bmatrix}A'_{mn}\\C'_{mn}\end{Bmatrix}+\tilde{\mathbf{P}}^{(2)}\begin{Bmatrix}D_{mn}\\F_{mn}\end{Bmatrix},\quad \text{at } r=b \tag{15}$$

$$\mu\mathbf{Q}\begin{Bmatrix}A_{mn}\\C_{mn}\end{Bmatrix}+\mu\tilde{\mathbf{Q}}\begin{Bmatrix}\Phi_{mn}\\X_{mn}\end{Bmatrix}=\mu_2\mathbf{Q}^{(2)}\begin{Bmatrix}A'_{mn}\\C'_{mn}\end{Bmatrix}+\mu_2\tilde{\mathbf{Q}}^{(2)}\begin{Bmatrix}D_{mn}\\F_{mn}\end{Bmatrix},\quad \text{at } r=b \tag{16}$$

$$\mathbf{P}^{(1)}\begin{Bmatrix}A'_{mn}\\C'_{mn}\end{Bmatrix}+\tilde{\mathbf{P}}^{(1)}\begin{Bmatrix}D_{mn}\\F_{mn}\end{Bmatrix}=\mathbf{P}^0\begin{Bmatrix}A^0_{mn}\\C^0_{mn}\end{Bmatrix},\quad \text{at } r=a \tag{17}$$

$$\mu_2\mathbf{Q}^{(1)}\begin{Bmatrix}A'_{mn}\\C'_{mn}\end{Bmatrix}+\mu_2\tilde{\mathbf{Q}}^{(1)}\begin{Bmatrix}D_{mn}\\F_{mn}\end{Bmatrix}=\mu_1\mathbf{Q}^0\begin{Bmatrix}A^0_{mn}\\C^0_{mn}\end{Bmatrix},\quad \text{at } r=a. \tag{18}$$

Here, we define the following matrices:

Table 1 Substitution table

	Substitute	For
$\mathbf{P}^{(2)}$, $\mathbf{Q}^{(2)}$	k, K	k_2, K_2
$\tilde{\mathbf{P}}^{(2)}$, $\tilde{\mathbf{Q}}^{(2)}$	h_n	j_n
	k, K	k_2, K_2
$\mathbf{P}^{(1)}$, $\mathbf{Q}^{(1)}$	kb, Kb	k_1a, K_1a
$\tilde{\mathbf{P}}^{(1)}$, $\tilde{\mathbf{Q}}^{(1)}$	h_n	j_n
	kb, Kb	k_1a, K_1a
$\mathbf{P}^0$, $\mathbf{Q}^0$	kb, Kb	k_2a, K_2a

$$\mathbf{P}=\begin{bmatrix}P_{11}&P_{12}\\P_{21}&P_{22}\end{bmatrix},\quad \mathbf{Q}=\begin{bmatrix}Q_{11}&Q_{12}\\Q_{21}&Q_{22}\end{bmatrix}, \tag{19}$$

where

$$P_{11}=nh_n(kb)-kbh_{n+1}(kb) \tag{20}$$

$$P_{12}=n(n+1)h_n(Kb),\quad P_{21}=h_n(kb) \tag{21}$$

$$P_{22}=(n+1)h_n(Kb)-Kbh_{n+1}(Kb) \tag{22}$$

$$Q_{11}=(n^2-n-\tfrac{1}{2}K^2b^2)h_n(kb)+2kbh_{n+1}(kb) \tag{23}$$

$$Q_{12}=n(n+1)[(n-1)h_n(Kb)-Kbh_{n+1}(Kb)] \tag{24}$$

$$Q_{21}=(n-1)h_n(kb)-kbh_{n+1}(kb) \tag{25}$$

$$Q_{22}=(n^2-1-\tfrac{1}{2}K^2b^2)h_n(Kb)+Kbh_{n+1}(Kb). \tag{26}$$

The matrices $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{Q}}$ are obtained from $\mathbf{P}$ and $\mathbf{Q}$ by replacing h_n with j_n. The other matrices in Eqs. (15)–(18) may be obtained by the replacement as shown in Table 1. The 8×8 matrix equation for the undetermined coefficients of the normal expansion solutions for this problem is given by

$$\begin{bmatrix}\mathbf{P}&-\mathbf{P}^{(2)}&-\tilde{\mathbf{P}}^{(2)}&\mathbf{0}\\ \alpha\mathbf{Q}&-\mathbf{Q}^{(2)}&-\tilde{\mathbf{Q}}^{(2)}&\mathbf{0}\\ \mathbf{0}&-\mathbf{P}^{(1)}&-\mathbf{P}^{(1)}&-\mathbf{P}^0\\ \mathbf{0}&\mathbf{Q}^{(1)}&\tilde{\mathbf{Q}}^{(1)}&-\beta\mathbf{Q}^0\end{bmatrix}\begin{Bmatrix}A_{mn}\\C_{mn}\\A'_{mn}\\C'_{mn}\\D_{mn}\\F_{mn}\\A^0_{mn}\\C^0_{mn}\end{Bmatrix}=-\begin{bmatrix}\tilde{\mathbf{P}}\\ \alpha\tilde{\mathbf{Q}}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}\begin{Bmatrix}\Phi_{mn}\\X_{mn}\end{Bmatrix}, \tag{27}$$

where $\alpha=\mu/\mu_2$, $\beta=\mu_1/\mu_2$. Also, the relationship among B_{mn}, B'_{mn}, E_{mn}, and B^0_{mn} are

$$h_n(Kb)B_{mn}+j_n(Kb)Y_{mn}=h_n(K_2b)B'_{mn}+j_n(K_2b)E_{mn} \tag{28}$$

$$\mu[(n-1)h_n(Kb)-Kbh_{n+1}(Kb)]B_{mn}+\mu[(n-1)j_n(Kb)-Kbj_{n+1}(Kb)]Y_{mn}=\mu_2[(n-1)h_n(K_2b)-K_2bh_{n+1}(K_2b)]B'_{mn}+\mu_2[(n-1)j_n(K_2b)-K_2bj_{n+1}(K_2b)]E_{mn} \tag{29}$$

$$h_n(K_2a)B'_{mn}+j_n(K_2a)E_{mn}=j_n(K_1a)B^0_{mn} \tag{30}$$

$$\mu_2[(n-1)h_n(K_2a)-K_2ah_{n+1}(K_2a)]B'_{mn}+\mu_2[(n-1)j_n(K_2a)-K_2aj_{n+1}(K_2a)]E_{mn}=\mu_1[(n-1)j_n(K_1a)-K_1aj_{n+1}(K_1a)]B'_{mn}. \tag{31}$$

In Eqs. (15)–(16) and (28)–(29), Φ_{mn}, X_{mn}, and Y_{mn} are expressed by

$$\Phi_{mn}=\frac{i^{n-1}}{k}(2n+1)\delta_{m0} \tag{32}$$

$$X_{mn}=\frac{i^{n-1}}{2K}\frac{2n+1}{n(n+1)}[\delta_{m1}-n(n+1)\delta_{m,-1}] \tag{33}$$

$$Y_{mn}=\frac{i^{n-1}}{2}\frac{2n+1}{n(n+1)}[\delta_{m1}+n(n+1)\delta_{m,-1}], \tag{34}$$

where δ_{m0}, δ_{m1}, and $\delta_{m,-1}$ are the Kronecker delta. The 4×4 matrix equation for the undetermined coefficients of the normal expansion solutions is

$$\begin{bmatrix} \xi_{11} & -\xi_{12} & -\xi_{13} & 0 \\ \alpha\xi_{21} & -\xi_{22} & -\xi_{23} & 0 \\ 0 & \xi_{32} & \xi_{33} & -\xi_{34} \\ 0 & \xi_{42} & \xi_{43} & -\beta\xi_{44} \end{bmatrix} \begin{Bmatrix} B_{mn} \\ B'_{mn} \\ E_{mn} \\ B^0_{mn} \end{Bmatrix} = -Y_{mn} \begin{bmatrix} \zeta_1 \\ \alpha\zeta_2 \\ 0 \\ 0 \end{bmatrix}, \tag{35}$$

$$\xi_{11}=h_n(Kb), \quad \xi_{12}=h_n(K_2b), \quad \xi_{13}=j_n(K_2b) \tag{36}$$

$$\xi_{21}=(n-1)h_n(Kb)-Kbh_{n+1}(Kb) \tag{37}$$

$$\xi_{22}=(n-1)h_n(K_2b)-K_2bh_{n+1}(K_2b) \tag{38}$$

$$\xi_{23}=(n-1)j_n(K_2b)-K_2bj_{n+1}(K_2b) \tag{39}$$

$$\xi_{32}=h_n(K_2a), \quad \xi_{33}=j_n(K_2a), \quad \xi_{34}=j_n(K_1a) \tag{40}$$

$$\xi_{42}=(n-1)h_n(K_2a)-K_2ah_{n+1}(K_2a) \tag{41}$$

$$\xi_{43}=(n-1)j_n(K_2a)-K_2aj_{n+1}(K_2a) \tag{42}$$

$$\xi_{44}=(n-1)j_n(K_1a)-K_1aj_{n+1}(K_1a) \tag{43}$$

$$\zeta_1=j_n(Kb), \quad \zeta_2=(n-1)j_n(Kb)-Kbj_{n+1}(Kb). \tag{44}$$

3 Multiple Scattering Formulas

The effective complex wave number $\langle k\rangle$ of the multiple scattered waves can be obtained in terms of the frequency of the waves and the microstructure of the composite. The averaging technique developed by Waterman and Truell involves a configurational averaging technique using the joint probability distribution for the occurrence of a given configuration of scatterers to average the resulting wave over all configurations. For multiple scattering by a random distribution of spherical scatters, Waterman and Truell derived an expression for effective wave numbers in terms of the number of scatters per unit volume n_0 and the far-field amplitude for a single scatter.

For longitudinal waves, the effective wave number is

$$\left[\frac{\langle k\rangle}{k_2}\right]^2=\left[1+\frac{2\pi n_0 f(0)}{k_2^2}\right]^2-\left[\frac{2\pi n_0 f(\pi)}{k_2^2}\right]^2. \tag{44}$$

For shear waves, the effective wave number is

$$\left[\frac{\langle K\rangle}{K_2}\right]^2=\left[1+\frac{2\pi n_0 g(0)}{K_2^2}\right]^2-\left[\frac{2\pi n_0 g(\pi)}{K_2^2}\right]^2. \tag{45}$$

where k_2, K_2 denotes the wave numbers of the P waves and S waves in the matrix material, respectively. The number of scatters per unit volume n_0 is related to concentration c by

$$n_0=\frac{3c}{4\pi a^3}. \tag{46}$$

The scattered field $\mathbf{u}^s$ at a large distance from an isolated scatter is given by

$$\mathbf{u}^s\approx|w_0|f(\theta)\frac{e^{ikr}}{r}\mathbf{e}_r+|u_0|g_1(\theta,\phi)\frac{e^{iKr}}{r}\mathbf{e}_\theta+|u_0|g_2(\theta,\phi)\frac{e^{iKr}}{r}\mathbf{e}_\phi, \tag{47}$$

where $|w_0|$ and $|u_0|$ are the incident displacement amplitudes. The function $f(\theta)$ denotes the scattering amplitude of the longitudinal waves in the far-field, and the functions $g_1(\theta,\phi)$ and $g_2(\theta,\phi)$ are far-field scattering amplitudes of the shear waves in the directions of $\mathbf{e}_\theta$ and $\mathbf{e}_\phi$, respectively. The definitions of these equations are given in Appendix A. The forward and backward scattering amplitudes of longitudinal waves are represented by $f(0)$ and $f(\pi)$ which can be calculated from an isolated scatter contained in the infinite matrix. Also, the forward and backward scattering amplitudes of shear waves are termed by $g(0)$ and $g(\pi)$ which are given by

$$g(0)=\cos\phi g_1(0,\phi)-\sin\phi g_2(0,\phi) \tag{48}$$

$$g(\pi)=\cos\phi g_1(\pi,\phi)-\sin\phi g_2(\pi,\phi). \tag{49}$$

It can be further shown that $g(0)$ and $g(\pi)$ are independent of ϕ. The scattering theory described above yields propagation constants $\langle k\rangle$ and $\langle K\rangle$ in terms of the far-field amplitude of the single scatter and is valid provided the effect of correlation in position of the scatterers can be neglected. The approximation is valid only at relatively low concentrations. If the inclusions are sufficiently dense and closely spaced and the wavelength is comparable to the size of spheres, the spherical inclusions cannot be considered as independent scatters and a modified multiple scattering theory must be used. In addition, the solution of Eqs. (44) and (45) exhibits low-frequency velocity limits different from those predicted by the well-known static solution (for example, the generalized self-consistent model (GSCM)). To satisfy the low-frequency limit and nondilute concentration, the Waterman-Truell model is modified by the implementation of GSCM.

3.1 Generalized Self-Consistent Multiple Scattering Model. It is assumed that each inclusion is surrounded by a shell of matrix and the composite is embedded in an effective medium. Then, in the scattering formulas, k_2 must be replaced by $\langle k\rangle$ and K_2 must be replaced by $\langle K\rangle$ leading to the equations

$$1=\left[1+\frac{2\pi n_0 f(0)}{\langle k\rangle^2}\right]^2-\left[\frac{2\pi n_0 f(\pi)}{\langle k\rangle^2}\right]^2 \tag{50}$$

$$1=\left[1+\frac{2\pi n_0 g(0)}{\langle K\rangle^2}\right]^2-\left[\frac{2\pi n_0 g(\pi)}{\langle K\rangle^2}\right]^2, \tag{51}$$

where

$$f(0)=\sum_{n=0}^{\infty}(-i)^nA_{0n}, \quad f(\pi)=\sum_{n=0}^{\infty}(i)^nA_{0n} \tag{52}$$

$$g(0)=\sum_{n=1}^{\infty}\frac{(-i)^n}{2}\left[n(n+1)C_{1n}-C_{-1n}+\frac{n(n+1)}{K}B_{1n}+\frac{1}{K}B_{-1n}\right] \tag{53}$$

$$g(\pi)=\sum_{n=1}^{\infty}\frac{(i)^n}{2}\left[n(n+1)C_{1n}-C_{-1n}+\frac{n(n+1)}{K}B_{1n}+\frac{1}{K}B_{-1n}\right]. \tag{54}$$

The scattering coefficients A_{mn}, B_{mn}, C_{mn} given by Eqs. (27) and (35) are functions of the unknown effective wave numbers. It should be noted that since A_{mn}, B_{mn}, C_{mn} are transcendental functions of the unknowns $\langle k\rangle$ and $\langle K\rangle$, explicit solutions cannot be obtained from Eqs. (50)–(51). We use an iterative procedure to solve them for $\langle k\rangle$ and $\langle K\rangle$. The iteration is started by taking effective properties k, K, and ρ in the dynamic GSCM model equal to matrix values k_2, K_2, and ρ_2, respectively. Then, the forward and backward scattered amplitudes are calculated and the homogenization is carried out by using Eqs. (44) and (45). Next, we substitute the corrected effective wave numbers $\langle k\rangle$ and $\langle K\rangle$

Table 2 Properties of constituents

	Density (g/cm^3)	λ (GPa)	μ (GPa)
SiC	3.181	98	188.1
Al	2.706	57.5	26.5
Epoxy	1.202	4.916	1.731
Lead	11.3	38.46	8.357

for k and K and these procedures are repeated until convergence is obtained. Numerical results for specific type of particulate composites will be presented by the effective wave speed and attenuation in a later section.

3.2 The Static Limit. The static limits of the effective dynamic constants obtained from Eqs. (50) and (51) can be evaluated by using the asymptotic expansions of the spherical Bessel and Hankel functions for $ka \to 0$. However, due to the complexity of the system of equations, the closed-form expressions for the static limits are difficult to be obtained. Using the present theory, the numerical results in the low-frequency limit show that the effective moduli of a particulate composite converge to the well-known static formulas, namely the effective bulk modulus obtained by Hashin [24] and effective shear modulus derived by Christensen and Lo [25]. The expressions for the static effective elastic moduli of particulate composites in terms of their constituent properties are given in Appendix B. To the author's knowledge, this seems to be the first time in the literature to be able to recover both well-known and rigorous static effective moduli which could cover a full range of volume fractions ($0 \leqslant c \leqslant 1$) of the particulate composites through an effective plane wave propagation approach.

4 Numerical Results

Numerical calculations are performed for two random particulate composite: SiC-Al (metal-matrix composites with ceramic reinforcement) and lead-epoxy (polymer matrix). We assume that both matrix and inclusion are elastic and isotropic. The properties of the all the constituents taken from Shindo et al. [5] and Kinra et al. [21] are given in Table 2.

For SiC-Al particulate composites, the density ratio contrast (ρ_1/ρ_2) is not too high, it is appropriate to assume that the effective density is the mean density. Figures 2 and 3 show the normalized phase velocity and attenuation of P and S waves calculated from the present theory and from the Waterman-Truell method for concentrations $c=0.05$ and $c=0.15$ in a SiC-Al particulate composite. For low concentration ($c=0.05$), the phase velocities obtained from both methods are close. As the concentration increases, the so-called condition of weak scattering density is not satisfied and the discrepancies become significant. It should be noted that only present theory converges the effective bulk modulus obtained by Hashin [24] and effective shear modulus derived by Christensen and Lo [25] at the zero-frequency limit. The attenuation predicted by the Waterman-Truell formula appears to be significantly higher at lower frequency than that obtained from dynamic GSCM. Phase velocity and attenuation of P and S waves calculated from the present theory for a concentration $c=0.3$ in a SiC-Al particulate composite are shown in Fig. 4. Unlike the former case where phase velocities for P and S waves decrease monotonously in the frequency range from $k_2a=0$ to $k_2a=2.0$, the phase velocity of P waves for $c=0.3$ attains its local minimum at $k_2a=1.4$. Figures 5 and 6 show the phase velocity and attenuation versus concentration in SiC-Al composites for $k_2a=1.0$. It is seen that, at any frequency, attenuation increases initially with concentration, attains a maximum and then declines to zero as concentration increases to 1. The maximum attenuation for normalized frequencies $k_2a=1.0$ occurs at $c=0.17$ for P waves and at $c=0.20$ for S waves.

For composites with high-density contrast ratio, the influence of density variations on the dynamic properties of composites is significant and the effective density becomes complex and frequency-dependent ([26]). Thus, in the case of lead-epoxy composites, the dynamic effective density derived by Sabina and Willis [14] as shown in Appendix C is used instead of mean density. The longitudinal wave speed obtained from the present theory using the dynamic GSCM is compared with the theoretical results by Waterman and Truell, Sabina and Willis as well as with the experimental results by Kinra et al. [22,23] as shown in Figs. 7–9. Although these models gives reasonable agreement with the data for 5% and 15% volume concentrations of lead spheres in epoxy, it shows only dynamic GSCM predicts the overall trend of the effective wave speeds for high volume fraction $c=52\%$. Furthermore, it can be seen in Fig. 9 that the discrepancy among the zero-frequency limits of phase velocities predicted by different methods is significant at high volume fraction. This is due to the mutual interactions among inclusions become predominant for dense packing of scatterers. However, it can be shown that only the proposed theory converges to the well-known static solutions ([25]), which gives physically reasonable results at high concentrations and covers the full range of volume fractions ([18]).

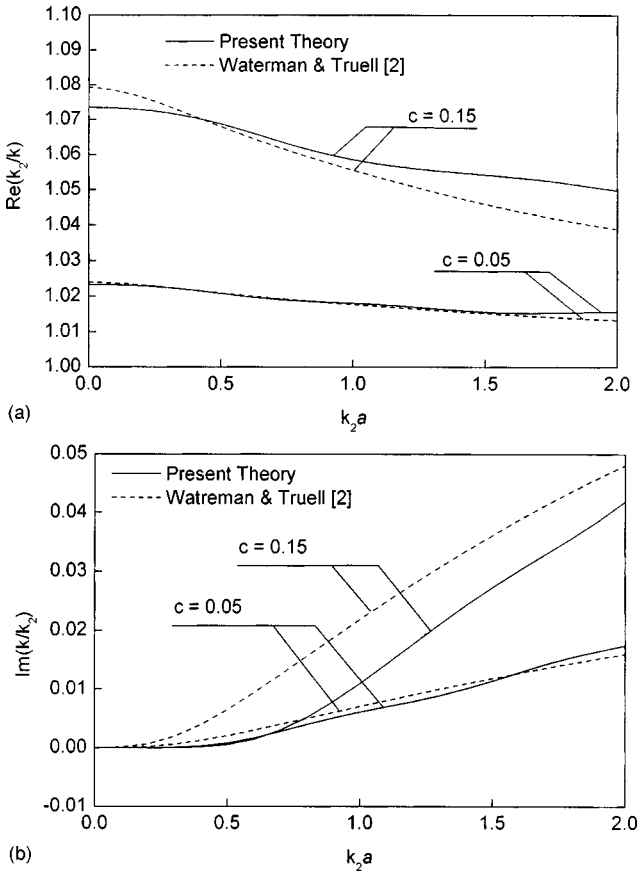


Fig. 2 Plots of phase velocity (*a*) and attenuation (*b*) of P waves with normalized frequency, calculated from the present theory and from the Waterman-Truell method for concentrations $c=0.05$ and $c=0.15$ in a SiC-Al particulate composite

5 Conclusions

We have modeled the coherent wave field propagating in an isotropic medium with randomly distributed spherical inclusions by the dynamic generalized self-consistent model. The phase velocity and attenuation of compressional and shear waves are strongly dependent on the normalized frequency, k_2a, i.e., they are dependent both on frequency and inclusion size. In the past, theoretical predictions of the overall elastodynamic moduli have been shown to be in good agreement with experimental results at

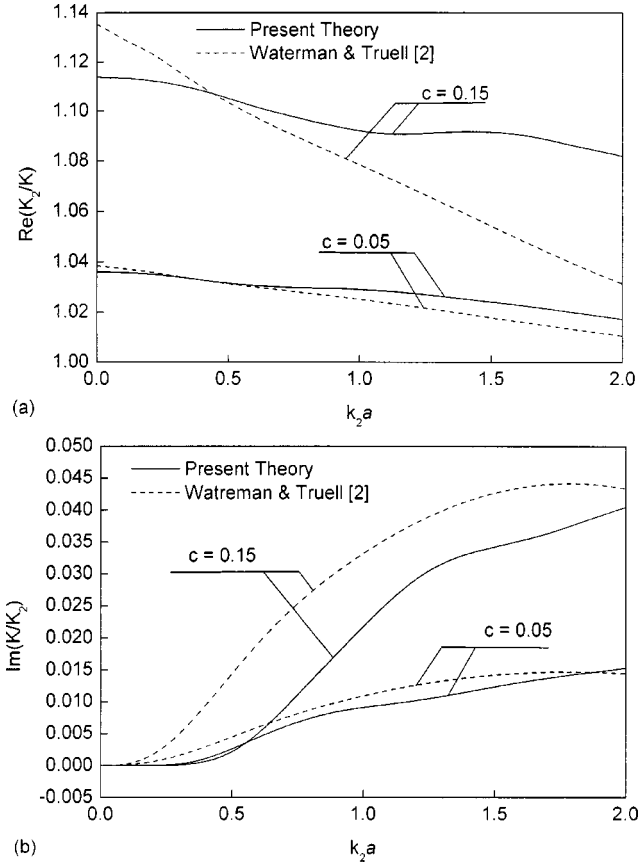


Fig. 3 Plots of phase velocity (*a*) and attenuation (*b*) of S waves with normalized frequency, calculated from the present theory and from the Waterman-Truell method for concentrations $c=0.05$ and $c=0.15$ in a SiC-Al particulate composite

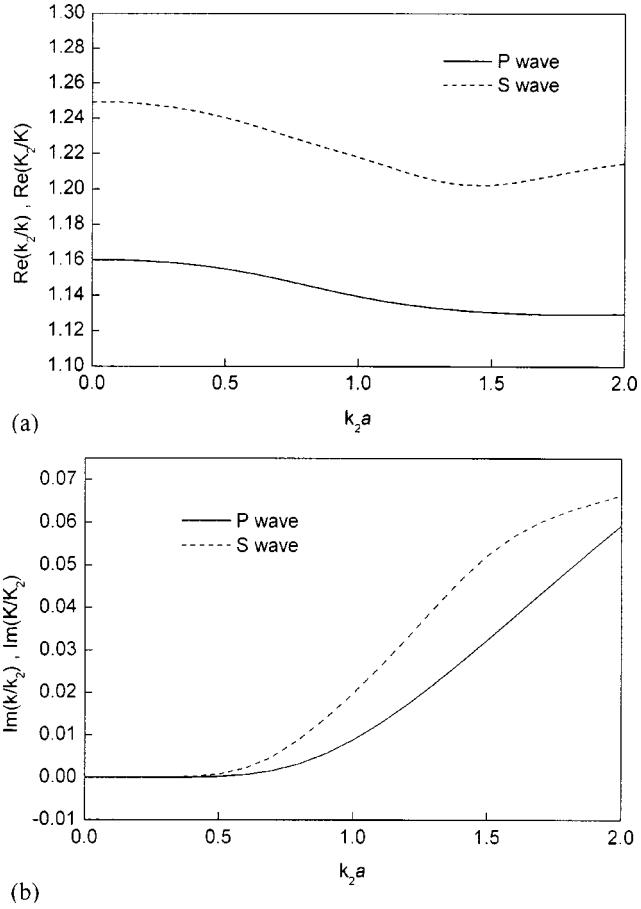


Fig. 4 Phase velocity (*a*) and attenuation (*b*) of P and S waves calculated from the present theory for concentrations $c=0.3$ in a SiC-Al particulate composite

low volume concentration of the spherical inclusions. At the high volume concentrations, the effect of multiple scattering by inclusions becomes significant and the theoretical estimates are not satisfactory. In addition, the overall dynamic elastic moduli calculated by Waterman and Truell's formulations do not reduce to their independently estimated static values in the limit of zero frequency. The proposed dynamic generalized self-consistent model which covers a full range of volume fractions ($1 \leqslant c \leqslant 1$) can not only reproduce both well-known effective moduli of the particulate composites in the static but also appear to yield reasonable results at higher concentrations and frequencies.

Acknowledgment

This work was supported by the National Science Council of Taiwan under Grant No. NSC 89-2212-E-035-029.

Appendix A

The far-field scattering amplitude functions for both P and S waves are given by

$$f(\theta,\phi)=\sum_{n=0}^{\infty}(-i)^n A_{0n} P_n^0(\cos\theta) \tag{A1}$$

$$g_1(\theta,\phi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}(-i)^n\left[C_{mn}\frac{\partial}{\partial\theta}P_n^m(\cos\theta)+\frac{mB_{mn}}{K_2\sin\theta}P_n^m(\cos\theta)\right]e^{im\phi} \tag{A2}$$

$$g_2(\theta,\phi)=-\sum_{n=0}^{\infty}\sum_{m=-n}^{n}(-i)^{n+1}\left[\frac{mC_{mn}}{\sin\theta}P_n^m(\cos\theta)+\frac{B_{mn}}{K_2}\frac{\partial}{\partial\theta}P_n^m(\cos\theta)\right]e^{im\phi}. \tag{A3}$$

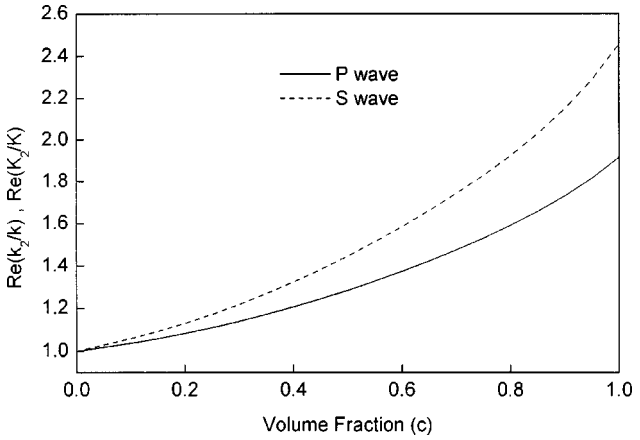


Fig. 5 Phase velocity versus concentration in SiC-Al composites for normalized frequency $k_2a=1.0$

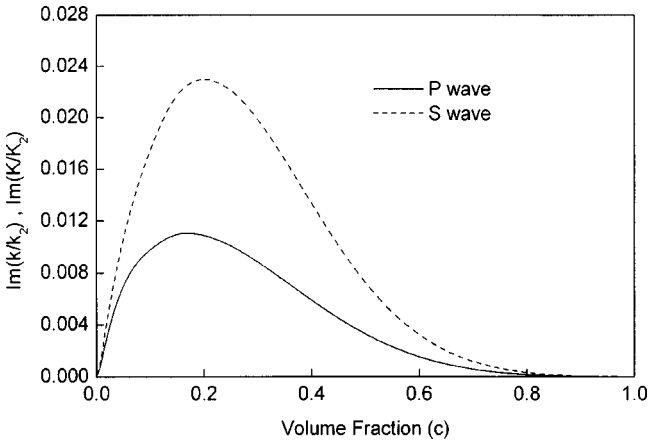


Fig. 6 Attenuation versus concentration in SiC-Al composites for normalized frequency $k_2a=1.0$

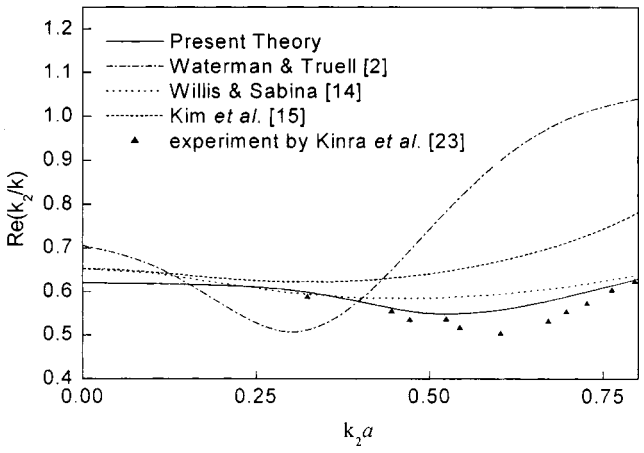


Fig. 9 P-wave phase velocity of lead/epoxy composites at a concentration of $c=0.52$

Appendix B

Christensen and Lo [18] used the GSCM to obtain the static effective shear and bulk modulus of a particulate composite. The effective shear modulus is given by the solution of the quadratic equation

$$A(\mu/\mu_2)^2+2B(\mu/\mu_2)+C=0, \tag{B1}$$

where

$$A=8M(4-5\nu_2)\eta_1 c^{10/3}-2[63M\eta_2+2\eta_1\eta_3]c^{7/3}+252M\eta_2 c^{5/3}$$
$$-50M(7-12\nu_2+8\nu_2^2)\eta_2 c+4(7-10\nu_2)\eta_2\eta_3, \tag{B2}$$

$$B=-2M(1-5\nu_2)\eta_1 c^{10/3}+2[63M\eta_2+2\eta_1\eta_3]c^{7/3}$$
$$-252M\eta_2 c^{5/3}+75M(3-\nu_2)\eta_2\nu_2 c+\frac{3}{2}(15\nu_2-7)\eta_2\eta_3, \tag{B3}$$

$$C=4M(5\nu_2-7)\eta_1 c^{10/3}-2[63M\eta_2+2\eta_1\eta_3]c^{7/3}+252M\eta_2 c^{5/3}$$
$$+25M(\nu_2^2-7)\eta_2 c-(7+5\nu_2)\eta_2\eta_3, \tag{B4}$$

with

$$M=\frac{\mu_1}{\mu_2}-1, \tag{B5}$$

$$\eta_1=M(7-10\nu_2)(7+5\nu_1)+105(\nu_1-\nu_2), \tag{B6}$$

$$\eta_2=M(7+5\nu_1)+35(1-\nu_1), \tag{B7}$$

$$\eta_3=M(8-10\nu_2)+15(1-\nu_2), \tag{B8}$$

where ν_1 and ν_2 are the Poisson's ratios of the inclusion and matrix materials, respectively. The bulk modulus of the composite was also obtained as

$$\kappa=\kappa_2+\frac{c(\kappa_1-\kappa_2)}{1+(1-c)\dfrac{(\kappa_1-\kappa_2)}{\left(\kappa_2+\dfrac{4}{3}\mu_2\right)}}. \tag{B9}$$

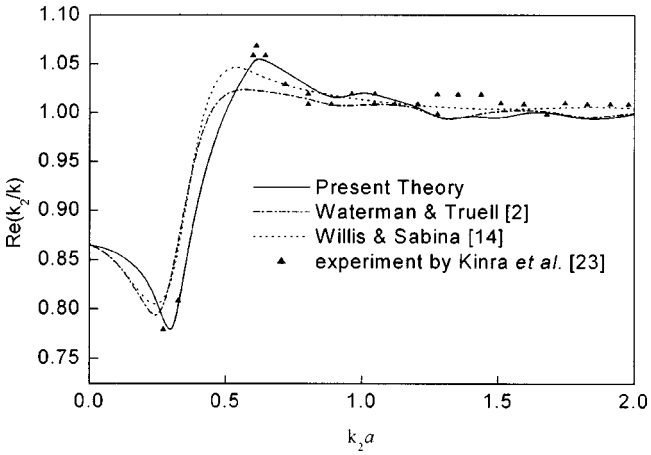


Fig. 7 P-wave phase velocity of lead/epoxy composites at a concentration of $c=0.05$

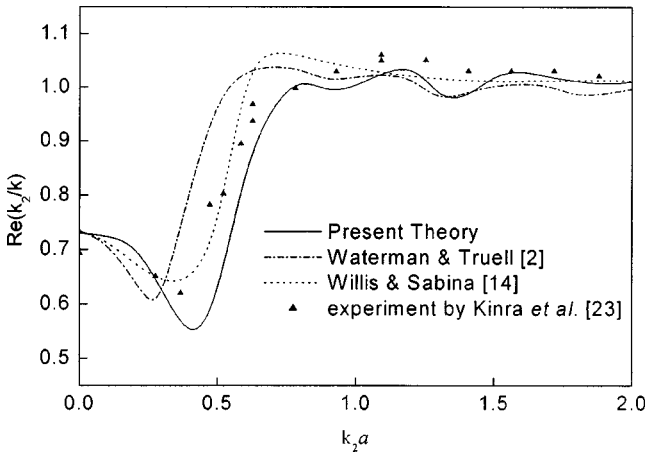


Fig. 8 P-wave phase velocity of lead/epoxy composites at a concentration of $c=0.15$

Appendix C

The effective properties derived by Sabina and Willis [14] can be expressed as

$$\kappa=\kappa_2+\frac{cg(h)g(-h)(\kappa_1-\kappa_2)}{1+3(\kappa_1-\kappa)\varepsilon_p/(3\kappa+4\mu)} \tag{C1}$$

$$\mu=\mu_2+\frac{cg(h)g(-h)(\mu_1-\mu_2)}{1+2(\mu_1-\mu)[2\mu\varepsilon_p+(3\kappa+4\mu)\varepsilon_s]/[5\mu(3\kappa+4\mu)]} \tag{C2}$$

$$\rho=\rho_2+\frac{cg(h)g(-h)(\rho_1-\rho_2)}{1+(\rho_1-\rho)(3-\varepsilon_p-2\varepsilon_s)/(3\rho)}. \tag{C3}$$

The function $g(h)$ for the spherical inclusion is

$$g(h)=3[\sin(ha)-ha\cos(ha)]/(ha)^3, \qquad (C4)$$

where the wave number h is replaced with k for P-wave incidence and with K for S-wave incidence. The terms ε_p and ε_s are given by

$$\varepsilon_p=3(1-ika)[\sin(ka)-ka\cos(ka)](e^{ika})/(ka)^3 \qquad (C5)$$

$$\varepsilon_s=3(1-iKa)[\sin(Ka)-Ka\cos(Ka)](e^{iKa})/(Ka)^3. \qquad (C6)$$

Equations $(C1)$–$(C6)$ can be solved by iteration.

References

[1] Foldy, L. L., 1945, ''The Multiple Scattering of Waves,'' Phys. Rev., **67**, pp. 107–119.

[2] Waterman, P. C., and Truell, R., 1961, ''Multiple Scattering of Elastic Waves,'' J. Math. Phys., **2**, pp. 512–537.

[3] Sayers, C. M., and Smith, R. L., 1983, ''Ultrasonic Velocity and Attenuation in an Epoxy Matrix Containing Lead Inclusion,'' J. Phys. D: Appl. Phys., **16**, pp. 1189–1194.

[4] Datta, S. K., Ledbetter, H. M., Shindo, Y., and Shah, A. H., 1988, ''Phase Velocity and Attenuation of Plane Elastic Waves in a Particulate-Reinforced Composite Medium,'' Wave Motion, **10**, pp. 171–182.

[5] Shindo, Y., Nozaki, H., and Datta, S. K., 1995, ''Effect of Interface Layers on Elastic Wave Propagation in a Metal Matrix Composite Reinforced by Particles,'' ASME J. Appl. Mech., **62**, pp. 178–185.

[6] O'Neill, T. J., Tebbutt, J. S., and Challis, R. E., 2001, ''Convergence Criteria for Scattering Models of Ultrasonic Wave Propagation in Suspensions of Particles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control, **48**, pp. 419–424.

[7] Tsang, L., and Kong, J. A., 1980, ''Multiple Scattering of Electro-Magnetic Waves by Random Distributions of Discrete Scatterers With Coherent Potential and Quantum Mechanical Formalism,'' J. Appl. Phys., **51**, pp. 3465–3485.

[8] Mal, A. K., and Bose, S. K., 1974, ''Dynamic Moduli of a Suspension of Imperfectly Bonded Spheres,'' Proc. Cambridge Philos. Soc., **76**, pp. 578–600.

[9] Datta, S. K., 1977, ''A Self-Consistent Approach to Multiple Scattering by Elastic Ellipsoidal Inclusions,'' ASME J. Appl. Mech., **44**, pp. 657–662.

[10] Willis, J. R., 1980, ''A Polarization Approach to the Scattering of Elastic Waves—I. Scattering by a Single Inclusion,'' J. Mech. Phys. Solids, **28**, pp. 287–305.

[11] Willis, J. R., 1980, ''A Polarization Approach to the Scattering of Elastic Waves—II. Multiple Scattering From Inclusions,'' J. Mech. Phys. Solids, **28**, pp. 307–327.

[12] Varadan, V. K., Ma, Y., and Varadan, V. V., 1985, ''A Multiple Scattering Theory for the Elastic Wave Propagation in Discrete Random Media,'' J. Acoust. Soc. Am., **77**, pp. 375–385.

[13] Hashin, Z., and Shtrikman, S., 1962, ''A Variational Approach to the Theory of the Elastic Behavior of Multiphase Materials,'' J. Mech. Phys. Solids, **2**, pp. 127–140.

[14] Sabina, F. J., and Willis, J. R., 1988, ''A Simple Self-Consistent Analysis of Wave Propagation in Particulate Composite,'' Wave Motion, **10**, pp. 127–142.

[15] Kim, J.-Y., Ih, J.-G., and Lee, B.-H., 1995, ''Dispersion of Elastic Waves in Random Particulate Composites,'' J. Acoust. Soc. Am., **97**, pp. 1380–1388.

[16] Hill, R., 1965, ''A Self Consistent Mechanics of Composite Materials,'' J. Mech. Phys. Solids, **13**, pp. 213–222.

[17] Budiansky, B., 1965, ''On the Elastic Moduli of Some Heterogeneous Materials,'' J. Mech. Phys. Solids, **13**, pp. 223–227.

[18] Christensen, R. M., 1990, ''A Critical Evaluation for a Class of Micro-Mechanics Model,'' J. Mech. Phys. Solids, **38**, pp. 379–404.

[19] Yang, R.-B., and Mal, A. K., 1994, ''Multiple Scattering of Elastic Waves in a Fiber-Reinforced Composite,'' J. Mech. Phys. Solids, **42**, pp. 1945–1968.

[20] Yang, R.-B., and Mal, A. K., 1996, ''Elastic Waves in a Composite Containing Inhomogeneous Fibers,'' Int. J. Eng. Sci., **34**, pp. 67–79.

[21] Huang, W., and Rokhlin, S. I., 1995, ''Frequency Dependencies on Ultrasonic Wave Velocity and Attenuation in Fiber Composites: Theory and Experiments,'' Rev. Prog. Quant. Nondestr. Eval., **14**, pp. 1233–1238.

[22] Kinra, V. K., and Anand, A., 1982, ''Wave Propagation in a Random Particulate Composite at Long and Short Wave Lengths,'' Int. J. Solids Struct., **18**, pp. 367–380.

[23] Kinra, V. K., and Rousseau, C., 1987, ''Acoustical and Optical Branches of Wave Propagation,'' J. Wave Mat. Interact., **2**, pp. 141–152.

[24] Hashin, Z., 1962, ''The Elastic Moduli of the Heterogeneous Materials,'' ASME J. Appl. Mech., **29**, pp. 143–150.

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

[26] Willis, J. R., 1985, ''The Nonlocal Influence of Density Variations in a Composite,'' Int. J. Solids Struct., **21**, pp. 805–817.

The Analysis of Constrained Impulsive Motion

L.-S. Wang

W.-T. Chou

Institute of Applied Mechanics,
National Taiwan University,
Taipei, Taiwan R.O.C.

Impulsive problems for mechanical systems subject to kinematic constraints are discussed in this paper. In addition to the applied impulses, there may exist suddenly changed constraints, or termed impulsive constraints. To describe the states of the system during the impulsive motion, three different phases, i.e., prior motion, virtual motion, and posterior motion, are defined which are subject to different sets of constraints, and thus have different degrees-of-freedom. A fundamental principle, i.e., the principle of velocity variation, for the constrained impulsive motion is enunciated as a foundation to derive the privileged impulse-momentum equations. It is shown that for a system with no applied impulse, a conservation law can be stated as the conservation of the virtual-privileged momenta. The proposed methodology provides a systematic scheme to deal with various types of impulsive constraints, which is illustrated in the paper by solving the constrained impulsive problems for the motion of a sleigh. [DOI: 10.1115/1.1577599]

1 Introduction

A mechanical system undergoes an impulsive motion if the motion, particularly the velocities, of the particles in the system are changed so rapidly that the duration of the process may be regarded to be instantaneous. Such phenomenon occurs when the system is subject to applied impulses or the constraints on the system are suddenly changed, and the associated problems are termed the *constrained impulsive problems*. In contrast to those continuously imposed constraints, termed the *regular constraints*, the suddenly varied constraints, called the *impulsive constraints*, may significantly affect the analysis of the impulsive motion, in which even the degree-of-freedom may be changed. How to systematically perform the analysis of the constrained impulsive motion subject to various types of impulsive constraints is the main theme of this paper.

Impulsive problems occur frequently in the motion of a mechanical system, and have been treated in many classical textbooks, such as [1–6], or in some recent literatures [7–12], among others. They also appear in the study of the motion subject to one-sided constraints, cf. [13–15] and the references therein, in which the continuous motion and jumps of kinematic variables exist hybridly. Analysis of such hybrid system, [16,17], requires an effective tool to deal with both the regular constraints and the impulsive constraints. According to their durations, the latter is further divided in this paper into four types, i.e. live, inert, elastic, and released. Similar classifications can be found in [1], but in which the live impulsive constraints are not mentioned and the released constraints have different characteristics. On the other hand, while the live constraints, as well as the inert ones, were discussed in [3,8,18,19], the effects of released and elastic impulsive constraints have not been extensively analyzed. Although the mathematical formulation for impulsive constraints has been attempted in [12,15,20,21], a synthesized approach to treat the above-mentioned four categories of impulsive constraints is still in demand. To attain this goal, a fundamental principle for impulsive motion needs to be invoked.

In 1903, Appell [1] enunciated a basic principle for percussion or impulsive motion analogous to D'Alembert's principle in dynamics, which leads to a variational equation in terms of virtual displacements. However, due to the special character of the impulsive motion that the position of each particle is held fixed, the variation on the position such as the virtual displacement seems to be a little unnatural. On the other hand, in 1909, Jourdain [22] introduced the notion of variation on the velocity while keeping the time and the position fixed. This concept leads to the so-called Jourdain's variational equation, [22], or the principle of virtual power [23–26] for dynamical problems, and the corresponding variational equation on the (finite or infinitesimal) velocity variations is termed the second form of fundamental equations in [3]. Since the position is not changed for such variations, it is deemed that the fundamental equation in velocity variations is more appropriate to be used for dealing with constrained impulsive problems, cf. [3,11]. However, in the application of the variational equation, it is required to clearly identify the associated conditions on the velocity variations. For the aforementioned different types of impulsive constraints, we thus have to indicate their relations with the compatibility conditions on the velocity variations in the basic principle.

In this paper, we shall separate the impulsive motion of the system into three phases. The prior motion and the posterior motion refer to the motion immediately before and after the instant at which the impulsive motion occurs, respectively. In addition to the regular constraints, the system is subject to live and released constraints in the prior motion, while it is restricted by live and inert constraints in the posterior motion. To accommodate the elastic constraints, which are imposed on the system during the impulsive motion, it is postulated that the system is in the state of the third phase, i.e., the so-called *virtual motion*, in which live, inert, and elastic constraints are active. We note that the compatibility conditions on the velocity variations induced from the kinematic conditions in the virtual motion are those should be included in the fundamental variational equation. Accordingly, analogous to the principle of virtual power for finite-force motion, we propose the principle of velocity variations for the constrained impulsive motion of finite degree-of-freedom system as enunciated in Section 4. This fundamental principle essentially says that the applied impulses can be divided into the effective impulses which generate the jumps of velocities, and the net applied impulses, which cannot change the jumps. Hence the sum total of the product of the latter and the velocity variations compatible with the constraints in the virtual motion must vanish.

From the fundamental principle, the basic variational equation for impulsive motion can be immediately derived. The phase of virtual motion may be described by generalized velocities, which are obtained from the regular geometric constraints, and quasi-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, May 23, 2001; final revision, Dec. 3, 2002. Associate Editor: N. C. Perkins. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

velocities, which may arise naturally for a given problem. The holonomic or nonholonomic relationships between these velocities are solved to find a set of independent velocities, termed the virtual-privileged velocities, with total number being equal to the degree-of-freedom of the virtual motion. The variation of the velocity of each particle in the system can be then expressed in terms of the variations of the virtual-privileged velocities. Substituting the expressions into the basic variational equation, the so-called *privileged impulse-momentum equations for impulsive motion* are deduced, from which it is observed that the differences between the prior virtual-privileged momenta and the posterior virtual-privileged momenta are balanced by the applied virtual-privileged impulses. The privileged impulse-momentum equations, formulated in the virtual motion, can be then used to find the relation between the prior motion and the posterior motion. In particular, the conservation law of the virtual-privileged momenta can be stated for the constrained impulsive problem if there is no applied impulse. However, if elastic constraints appear, the privileged impulse-momentum equations are insufficient to determine the states in the posterior motion from the prior motion, and additional criteria, such as the law of impact, should be invoked.

While the privileged impulse-momentum equations for impulsive motion subject to *regular* constraints derived here are similar to those obtained by using Kane's approach, [7,27], the impulsive constraints and the notion of virtual motion were not discussed there. The basic variational equation introduced here provides the more suitable scenario to deal with the impulsive constraints, and paves the way for further synthesis with other methodologies for either continuous motion or for body with infinite degree-of-freedom. The privileged velocities mentioned before, called the *generalized speeds* in [7], are systematically obtained here by using the active kinematic constraints and the expressions of the quasi-velocities. For some problems associated with complex mechanical systems, Kane's approach may be superior to the others in deriving the equations of motion for finite-force problems, cf. [28]. The methodology proposed in this paper thus also have those advantages due to the similarity.

On the other hand, from the geometric point of view, the privileged equations may be viewed as the balance of the projections of the underlying quantities to some appropriate subspaces determined by the constraints, cf. [20,29–31]. In modern geometric mechanics, [32,33], the virtual displacement is regarded as the tangent vectors to the configuration space and the geometric formulation of Lagrangian mechanics, [34,35], is mainly based on the D'Alembert principle. Since the velocity variations are essentially the tangent vectors to the velocity space, it may be interesting to develop the geometric formulation for the principle of virtual power or the principle of velocity variations. In particular, the treatments of the force as a 1-form, [35], a horizontal 1-form, [34], or a semibasic 1-form, [12], in geometric Lagrangian mechanics may not be appropriate in formulating the Jourdain variational equation in modern geometric terminologies.

The rest of this paper is organized as follows. The classification of impulsive constraints and the separation of different phases of motion are discussed in Section 2. The kinematic constraints for various phases of motion and the notion of the privileged velocities are described in Section 3. The fundamental principle for impulsive motion is then stated in Section 4, along with the derivation of the privileged impulse-momentum equations for impulsive motion. The application of the equations to solve the prior-value problems is discussed in Section 5. Section 6 describes the application of the fundamental principle and the privileged equations to a system including rigid bodies. A physical example, i.e., the sleigh under impulsive motion, is then given in Section 7 to illustrate the proposed methodology, in which Newton's method and Lagrange's method are also used to solve the same problem to manifest their differences. Some concluding remarks are finally given in Section 8.

2 Classification of Impulsive Motions

The impulsive motion of a mechanical system refers to a sudden change of motion due to either applied impulses or sudden changes of constraints. The limiting case of a large force acting for a short interval of time may be viewed as an *impulsive force*, and an *impulse* is defined to be the limit of the integration of an impulsive force over the short interval. If sudden changes of constraints occur, each particle in the system may be treated as being imposed by the impulsive constraint forces, or constraint impulses. Similar to the notion of constraint force for a nonimpulsive motion, or termed *finite-force motion*, the constraint impulses only appear when the constraints exist. During the impulsive motion, each particle in the system may be impressed by external impulses or the interactive impulses from the other particles. The total impulses may be further grouped into the applied impulses and the constraint impulses. The problem associated with the impulsive motion of a constrained mechanical system exerted by applied impulses is termed the *constrained impulsive problem*. The kinematic condition that restricts the motion in the whole process of the impulsive motion is called a *regular constraint*. On the other hand, the constraint that appears or vanishes during the impulsive motion is named the *impulsive constraint*. Either regular constraints or impulsive constraints may exert constraint impulses during the impulsive motion.

To illustrate these notions, we consider a ball rolling on a rough surface. The ball rolls without sliding and thus is subject to some nonholonomic constraints. If there is an external impulse acting on the ball while the condition of pure rolling is not affected, these nonholonomic constraints are regular. If the ball hits the wall such that there are new constraints appearing, such constraints are the impulsive constraints. According to the duration of their effectiveness, the impulsive constraints may be further classified into the following categories:

(a) Inert Impulsive Constraints

For an inelastic ball falling on the floor, the constraint occurs during the impulsive motion, and holds after the motion. Such impulsive constraints that appear during the impulsive motion and are satisfied afterwards are called the *inert impulsive constraints*.

(b) Elastic Impulsive Constraints

If the ball is elastic and is dropped to the floor, the sudden constraints imposed by the floor during the impulsive motion disappear afterwards. Such impulsive constraints are termed the *elastic impulsive constraints*.

(c) Live Impulsive Constraints

If a ball rolls on a rough surface, and the surface is suddenly set to move, the form of the original constraints still holds, but some terms may be changed due to the sudden motion of the surface. As an example, consider a vertical disk with radius a rolling on a plane moving with velocity $(\dot{x}_p, \dot{y}_p)$. The rolling-without-sliding constraints can be expressed as

$$\dot{x}_c - a\dot{\phi}\cos\theta = \dot{x}_p, \tag{1}$$

$$\dot{y}_c - a\dot{\phi}\sin\theta = \dot{y}_p, \tag{2}$$

where $(\dot{x}_c, \dot{y}_c)$ denotes the velocity of the center of the disk, and $\dot{\phi}$, θ represent the spin rate, the heading angle of the disk, respectively. If the plane is moved suddenly so that $\dot{x}_p$ and $\dot{y}_p$ have abrupt changes, the set of constraints (1), (2) are varied accordingly. Such impulsive constraints are grouped as the *live impulsive constraints*.

(d) Released Impulsive Constraints

In contrast to the inert impulsive constraints, there may be some constraints vanishing during the impulsive motion and afterwards. Such impulsive constraints are termed the *released impulsive constraints*.

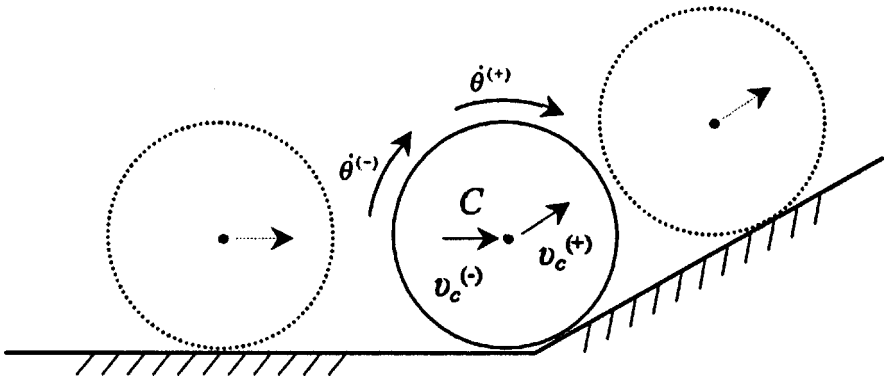


Fig. 1 A ball rolls across the boundary between two surfaces

Although a live impulsive constraint may be divided into a combination of a released constraint and an inert constraint, with certain specific relationship, it is treated specially in one category due to its frequent appearance and the distinct structure, [3].

A similar scheme for the classification of the constraints occurring in the impulsive motion was given by Appell [1], in which among the four types there the first one is essentially the aforementioned regular constraints, the second type is the inert constraints, the fourth type corresponds to the elastic constraints, while the live constraints defined above is not mentioned. In contrast to the released constraint, Appell's third type of constraints is regarded active during the impulsive motion. However, from some observations, the notion of released constraints introduced here may be more appropriate to be adopted. Consider a ball being grasped in hand, and suddenly released at $\bar{t}$. It is obvious that the velocity of the ball will not change during the short interval of the impulsive motion and immediately after the instant $\bar{t}$. The constraint of grasping is suddenly released and is thus not active during the impulsive motion.

Although the system undergoing released constraints will not experience a sudden change of velocity, the degree-of-freedom of its motion varies significantly. In fact, released constraints often occur simultaneously along with other kinds of impulsive constraints or the imposition of applied impulses to result in impulsive motion. Consider a ball rolling on a rough horizontal surface, and suddenly bumping into an inclined rough surface such that it may start to ascend, cf. Fig. 1. At the instant of impact, the constraint from the horizontal surface is released, which should not provide any action (constraint impulse) on the ball except that the degree-of-freedom of the system is increased. The suddenly appeared constraint from the inclined surface, the inert constraint, can then be imposed on the motion.

The difference between the released constraints and Appell's third type of constraints affects the analysis of impulsive motion. In fact, in [8], the constrained impulsive problems associated with the Appell's first and the second types of constraints, termed persistent constraints, are described to be determinant, while for the third and the fourth type, the problems are claimed to be indeterminant, for which additional laws are required. From the observations made above, it is seen that the problems associated the released constraint are determinant, and only problems with elastic constraint need more special treatment. However, it is noted that in the above-mentioned examples, the normal relative velocity between bodies before the impulsive motion subject to the released constraints vanishes. If it is nonzero, the corresponding impulsive constraint should be treated as a combination of a released constraint and an elastic one.

From the above discussions, the analysis of impulsive motion should take into account the occurrences of the impulsive constraints. Different classes of constraints may be active in different stages during the impulsive motion. This observation matches with the discussion in [12], where the analysis of motion subject to impulsive constraints based on modern geometric mechanics is performed individually for different types, i.e., permanent nonholonomic (regular), permanent impulsive (inert), elastic, etc.

To specifically identify the action of various impulsive constraints, we shall divide the impulsive motion into three phases. The motion immediately before the instant $\bar{t}$ at which the impulsive motion occurs is called the *prior motion*, during which live and released constraints are active. Immediately after $\bar{t}$, the system is in the *posterior motion* and is subject to inert and live constraints. In between the prior motion and the posterior motion, the system is influenced by live, inert, and elastic constraints, and may be postulated in the so-called *virtual motion*. The classification of the impulsive constraints and the separation of three phases during the impulsive motion pave the way to systematically treat the constrained impulsive problems in the following sections.

3 Kinematics of Impulsive Motions

Consider a system $\mathcal{S}$ of particles, whose number may be finite or infinite, indexed by an index set I. Each particle P_i in $\mathcal{S}$, $i \in I$, with mass m_i, is located at the position $\mathbf{r}_i \in \mathbb{R}^3$. The *configuration* of $\mathcal{S}$ is specified by the positions of all particles in $\mathcal{S}$. The system is assumed to be of *finite degree-of-freedom*, in the sense that there is a set of N particles in $\mathcal{S}$ such that the positions of all particles can be determined from those of the N particles, i.e.,

$$\mathbf{r}_i = \mathbf{r}_i(x_1, \cdots, x_{3N}, t), \quad i \in I,$$

where $(x_{3j-2}, x_{3j-1}, x_{3j})$ are the three components of $\mathbf{r}_j$, $j = 1, \cdots, N$. A system consisting of particles of finite number or rigid continua is in such a category. Although the methodology developed in this paper may be extended formally to systems with infinite degree-of-freedom, the finite degree-of-freedom is assumed here to avoid some technical difficulties and to enhance the clarity of later discussions.

If $\mathcal{S}$ is exerted by applied impulses or undergoes sudden change of constraints at instant $\bar{t}$, the system may experience a sudden change of motion between $(\bar{t}-, \bar{t}+)$, i.e., the interval immediately before and after the impulsive motion. Although it is more physically realistic to regard the duration of the motion in a very short interval $[\bar{t}-\tau, \bar{t}+\tau]$, the analysis of motion of $\mathcal{S}$ is usually performed by taking the limit as $\tau \to 0$. For each particle P_i in $\mathcal{S}$, its velocity, which is finite during the motion, may have a discontinuity at $\bar{t}$, with the corresponding *jump of velocity* $\Delta \mathbf{v}_i = \mathbf{v}_i^{(+)} - \mathbf{v}_i^{(-)}$, where $\mathbf{v}_i^{(-)}$ (prior velocity) and $\mathbf{v}_i^{(+)}$ (posterior velocity) denote the velocities of P_i at $\bar{t}-$ and $\bar{t}+$, respectively. The integration of the velocity $\dot{\mathbf{r}}_i(t)$ with respect to time over the interval $[\bar{t}-\tau, \bar{t}+\tau]$ is seen to be zero as $\tau \to 0$, due to the fact that the velocity is finite in the interval. Accordingly, the position of each particle in $\mathcal{S}$ is unchanged during the impulsive motion, denoted

by $\overline{\mathbf{r}}_i$, $i \in I$, and only the jump of velocity may occur. The above reasoning can be formalized by using the Dirac theory of distributions as in [14] to show that the position is fixed during the impulsive motion, which is essentially an a priori assumption in [9].

Let the system be subject to independent regular constraints including K geometric constraints and L linear kinematic constraints. Due to the assumption of finite degree-of-freedom, the constraints on every particles in $\mathcal{S}$ can be transformed into the constraints on the selected N particles. The regular constraints may be then expressed as

$$f_s(x_1,\cdots,x_{3N},t)=0, \quad s=1,\cdots,K, \tag{3}$$

$$\sum_{i=1}^{3N} A_{si}(x_1,\cdots,x_{3N},t)\dot{x}_i + a_s(x_1,\cdots,x_{3N},t)=0, \quad s=1,\cdots,L, \tag{4}$$

respectively. By solving the K geometric constraints, the position of each particle can be expressed parametrically by introducing $n(=3N-K)$ number of generalized coordinates $q_1,\cdots,q_n$, as

$$\mathbf{r}_i=\hat{\mathbf{r}}_i(\mathbf{q},t), \quad i\in I,$$

where $\mathbf{q}$ denotes the ensemble $(q_1,\cdots,q_n)$. Differentiating $\mathbf{r}_i$ with respect to time, the velocities can be then written as

$$\dot{\mathbf{r}}_i=\sum_{j=1}^{n}\frac{\partial\hat{\mathbf{r}}_i}{\partial q_j}\dot{q}_j+\frac{\partial\hat{\mathbf{r}}_i}{\partial t}, \quad i\in I, \tag{5}$$

where $\dot{q}_j=dq_j/dt$, $j=1,\cdots,n$, are the *generalized velocities* of the system. Substituting the components of the velocities of the N particles in Eq. (5) into the linear kinematic constraints (4), the constraints on the generalized velocities can be expressed as

$$\sum_{j=1}^{n} B^R_{sj}(\mathbf{q},t)\dot{q}_j+b^R_s(\mathbf{q},t)=0, \quad s=1,\cdots,L. \tag{6}$$

For a constrained mechanical system, it is sometimes convenient to describe the motion in terms of p *quasi-velocities* $\dot{\pi}_k$, defined by

$$\dot{\pi}_k \triangleq \sum_{j=1}^{n} C_{kj}(\mathbf{q},t)\dot{q}_j+c_k(\mathbf{q},t), \quad k=1,\cdots,p, \tag{7}$$

in which the right-hand side is nonintegrable. Since the regular constraints are imposed during the whole process, the motion of the system in all three phases can be then characterized by the combined set of n generalized velocities and p quasi-velocities. The configuration at $\overline{t}$ is specified by the generalized coordinates $\overline{q}_1,\cdots,\overline{q}_n$, and the prior motion, the posterior motion are described by the combined sets of velocities $(\dot{q}_1^{(-)},\cdots,\dot{q}_n^{(-)},\dot{\pi}_1^{(-)},\cdots,\dot{\pi}_p^{(-)})$, and $(\dot{q}_1^{(+)},\cdots,\dot{q}_n^{(+)},\dot{\pi}_1^{(+)},\cdots,\dot{\pi}_p^{(+)})$, respectively.

Other than the regular constraints, additional constraints may be imposed on or released from $\mathcal{S}$ during the impulsive motion, which may include geometric ones. For example, consider the motion of two rigid balls rolling on a plane and colliding with each other. At the instant of collision, there is an impulsive geometric constraint that the distance between two centers equals the sum of their radii. However, unlike the regular geometric constraints, the impulsive geometric constraints cannot be used to reduce the number of generalized coordinates, since they are only satisfied at the instant of impulsive motion. A configuration for the constrained impulsive motion is said to be *possible* if these impulsive geometric constraints are satisfied at the instant $\overline{t}$, which may be described by $(\overline{q}_1,\cdots,\overline{q}_n)$ that satisfies the impulsive geometric constraints. After determining the possible configuration at the instant of impulsive motion, these impulsive geometric constraints should be transformed into linear kinematic forms. Combining with the others, it is assumed that there are totally L_I impulsive linear kinematic constraints in the form of

$$\sum_{j=1}^{n} B^I_{sj}(\mathbf{q},t)\dot{q}_j+b^I_s(\mathbf{q},t)=0, \quad s=1,\cdots,L_I. \tag{8}$$

According to the classification in the previous section, these impulsive constraints are further divided into four groups, i.e., live, inert, elastic, and released, with numbers of $l_{\mathcal{L}}$, $l_{\mathcal{I}}$, $l_{\mathcal{E}}$, $l_{\mathcal{R}}$, respectively. For each group, the corresponding terms in the form of Eq. (8) will be denoted by $(B^{\mathcal{L}}_{sj},b^{\mathcal{L}}_s)$, $(B^{\mathcal{I}}_{sj},b^{\mathcal{I}}_s)$, $(B^{\mathcal{E}}_{sj},b^{\mathcal{E}}_s)$, and $(B^{\mathcal{R}}_{sj},b^{\mathcal{R}}_s)$, respectively, in which the inhomogeneous terms of the live constraints, $b^{\mathcal{L}}_s$, $s=1,\cdots,l_{\mathcal{L}}$, are different before and after $\overline{t}$. As discussed before, the prior motion satisfies live and released constraints, the posterior motion satisfies live and inert constraints, while live, inert, and elastic constraints are active in the virtual motion. For each phase of the impulsive motion, the impulsive constraints and the regular constraints (6) are combined into a set of active kinematic conditions on the motion. The corresponding degrees-of-freedom of the system are thus $m^b=n-L-l_{\mathcal{L}}-l_{\mathcal{R}}$, $m^a=n-L-l_{\mathcal{L}}-l_{\mathcal{I}}$, and $m'=n-L-l_{\mathcal{L}}-l_{\mathcal{I}}-l_{\mathcal{E}}$ for prior motion, posterior motion, and virtual motion, respectively. Here, and in what follows, the superscripts "b," "a," and "$'$" are used to indicate the validation of the corresponding terms for $t<\overline{t}-$, $t>\overline{t}+$, and $t\in(\overline{t}-,\overline{t}+)$, respectively.

Recall that the motion is described by the aggregate of n generalized velocities and p quasi-velocities, which are related by different sets of constraints for the three phases. In particular, for the virtual motion, the generalized velocities and the quasi-velocities satisfy the $L+l_{\mathcal{L}}+l_{\mathcal{I}}+l_{\mathcal{E}}$ equations of kinematic constraint and p equations of quasi-velocities (7). By solving these equations, which are linear in velocities, we may select m' number of independent velocities, $\dot{\phi}'_1,\cdots,\dot{\phi}'_{m'}$, called the *virtual-privileged velocities*, such that the generalized velocities and the quasi-velocities compatible with the kinematic conditions can be expressed as

$$\dot{q}_j=\sum_{\sigma=1}^{m'} D'_{j\sigma}(\mathbf{q},t)\dot{\phi}'_\sigma+d'_j(\mathbf{q},t), \quad j=1,\cdots,n, \tag{9}$$

$$\dot{\pi}_k=\sum_{\sigma=1}^{m'} G'_{k\sigma}(\mathbf{q},t)\dot{\phi}'_\sigma+g'_j(\mathbf{q},t), \quad k=1,\cdots,p. \tag{10}$$

Furthermore, by substituting Eq. (9) into Eq. (5), one may express the possible velocity of each particle of $\mathcal{S}$ in virtual motion as

$$\dot{\mathbf{r}}_i=\sum_{\sigma=1}^{m'}\beta'_{i\sigma}(\mathbf{q},t)\dot{\phi}'_\sigma+\gamma'_i(\mathbf{q},t), \quad i\in I, \tag{11}$$

where

$$\beta'_{i\sigma}=\sum_{j=1}^{n}\frac{\partial\hat{\mathbf{r}}_i}{\partial q_j}D'_{j\sigma}, \quad \gamma'_i=\sum_{j=1}^{n}\frac{\partial\hat{\mathbf{r}}_i}{\partial q_j}d'_j+\frac{\partial\hat{\mathbf{r}}_i}{\partial t},$$

are continuously differentiable functions of $(\mathbf{q},t)$, but not of $\dot{\mathbf{q}}$, since the kinematic constraints and the quasi-velocity equations are all linear in velocity.

Analogous to the process of selecting the virtual-privileged velocities, the set of *prior-privileged velocities* $\{\dot{\phi}^b_\sigma\}$, and that of *posterior-privileged velocities* $\{\dot{\phi}^a_\sigma\}$ can be obtained for the prior motion and the posterior motion, respectively. Similar forms of Eqs. (9), (10), (11) are found for each phase to describe the motion, with the corresponding notations summarized in Table 1. It is noted that the selection of privileged velocities basically depends on the coefficients of the velocity terms, i.e., B^I_{sj}, B^R_{sj}, and C_{kj}, and hence the discontinuity of the inhomogeneous terms arising from the live constraints does not affect the selection process.

Table 1 Summary of the transformations

	Prior Motion	Virtual Motion	Posterior Motion
Active Impulsive Constraints	Released, Live	Live, Elastic, Inert	Live, Inert
Privileged Coordinates	$\phi^b_\sigma,\ \sigma=1,\ldots,m^b$	$\phi'_\sigma,\ \sigma=1,\ldots,m'$	$\phi^a_\sigma,\ \sigma=1,\ldots,m^a$
$\dot{\phi}_\sigma$ to $\dot{q}_j$ (9)	$(D^b_{j\sigma},d^b_j)$	$(D'_{j\sigma},d'_j)$	$(D^a_{j\sigma},d^a_j)$
$\dot{\phi}_\sigma$ to $\dot{\pi}_k$ (10)	$(G^b_{k\sigma},g^b_k)$	$(G'_{k\sigma},g'_k)$	$(G^a_{k\sigma},g^a_k)$
$\dot{\phi}_\sigma$ to $\dot{\mathbf{r}}_i$ (11)	$(\beta^b_{i\sigma},\gamma^b_i)$	$(\beta'_{i\sigma},\gamma'_i)$	$(\beta^a_{i\sigma},\gamma^a_i)$

If all the kinematic constraints are regular, i.e., there is no impulsive constraint, the kinematic conditions for the three phases of motion are the same, with the same degree-of-freedom $m'=m^b=m^a=m=n-L$. The same set of privileged velocities may be selected as well to describe the motion for three phases, i.e., $\dot{\phi}'_\sigma=\dot{\phi}^b_\sigma=\dot{\phi}^a_\sigma=\dot{\phi}_\sigma,\ \sigma=1,\cdots,m$. The problem is then simplified to the classical problem of impulsive motion, and the method presented in this paper is similar to the one used in [7].

An impulsive problem that the prior motion is given and the posterior motion is to be determined is called a *prior-value problem* (analogous to the initial-value problem for differential equations). For such problems, the posterior velocity of each particle P_i in $\mathcal{S}$ can be expressed in terms of the posterior-privileged velocities evaluated at $\overline{t+}$ as

$$\mathbf{v}_i^{(+)}=\sum_{\sigma=1}^{m^a}\beta^a_{i\sigma}(\overline{\mathbf{q}},\overline{t})\,\dot{\phi}^{a(+)}_\sigma+\gamma^a_i(\overline{t+}),\quad i\in I. \tag{12}$$

From the previous equations, once the posterior-privileged velocities are determined, the posterior velocity of each particle in $\mathcal{S}$ can be obtained subsequently.

To solve the prior-value problem, fundamental principles in mechanics needs to be invoked. To deal with the systems subject to constraints for finite-force problems, many variational principles have been developed, in which the concept of variations is introduced to ''test'' the limitations from the constraints. These variations are essentially infinitesimal quantities imposed on the system variables. As reviewed in [25], there are basically three types of variations. In the principle of virtual work or D'Alembert's principle in dynamics, the virtual displacement is an infinitesimal and instantaneous displacement imposed on the configuration of the system. In the principle of least constraint or the Gauss principle, [36], the variation is imposed on the acceleration of each particle while keeping the time, the position, and the velocity unchanged. On the other hand, in Jourdain's principle, [22,23], or the principle of virtual power, [24,26], the variation is on the velocity, denoted by $\delta_1\mathbf{v}_i$, which is instantaneous ($\delta_1 t=0$) and stationary ($\delta_1\mathbf{r}_i=0$).

In the literature, the term of virtual displacement and that of virtual velocity are sometimes used interchangeably. This may be due to the reason that in the original work of Bernoulli [37], the principle of virtual work is called the principle of virtual velocity. To tell the difference between the virtual displacement and the variation of velocity defined above, the latter shall be termed the velocity variation in this paper. In modern geometric mechanics, the virtual displacement is viewed as a tangent vector to the configuration space, [38], due to its infinitesimal character. Analogously, the velocity variation may be treated as a tangent vector to the velocity space, and is thus intrinsically different from the notion of virtual displacement.

The velocity variation plays a central role in the fundamental principle for impulsive motion presented in the next section, which requires that they must be compatible with the kinematic conditions of constraints in the virtual motion. From the above discussions, any possible velocity of each particle in virtual motion can be represented in terms of the virtual-privileged velocities, Eq. (11). By taking δ_1-variation on (11), the velocity variation $\delta_1\dot{\mathbf{r}}_i$ can be then expressed in terms of the variations of the virtual-privileged velocities $\delta_1\dot{\phi}'_\sigma$ as

$$\delta_1\dot{\mathbf{r}}_i=\sum_{\sigma=1}^{m'}\beta'_{i\sigma}(\mathbf{q},t)\,\delta_1\dot{\phi}'_\sigma,\quad i\in I. \tag{13}$$

Since the virtual-privileged velocities $\dot{\phi}'_\sigma$ are independent, there is no constraint on the corresponding variations $\delta_1\dot{\phi}'_\sigma$. This fact shall be used in the next section to establish the required equations of jumps from the variational equation.

4 Kinetic Equations for Impulsive Motions

The dynamics of impulsive motion may be thought of as the limiting case for a finite-force problem. In addition to the D'Alembert-Lagrange equation and the Gauss-Gibbs equation, Jourdain in 1909, [22], established a variational equation for systems subject to linear kinematic constraints in terms of the infinitesimal variation of velocities. Later, Pars [3] extended the equations to accommodate the finite variations of velocities, and obtained the so-called second form of the fundamental equations as

$$\sum_{i\in I}(m_i\ddot{\mathbf{r}}_i-\mathbf{F}^{(A)}_i)\cdot\Delta_1\dot{\mathbf{r}}_i=0, \tag{14}$$

where $\ddot{\mathbf{r}}_i$ is the actual acceleration of particle P_i, $\mathbf{F}^{(A)}_i$ denotes the resultant applied force acting on P_i, and $\Delta_1\dot{\mathbf{r}}_i$ is an arbitrary finite or infinitesimal variation of possible velocity compatible with the constraints.

To extend the previous variational equation to deal with the constrained impulsive problem, due to the appearance of the impulsive constraints, the set of conditions from the constraints with which the velocity variations must be compatible should be identified. It may be envisioned that during the impulsive motion, the system undergoes the stage of virtual motion, during which the system satisfies the kinematic conditions from the regular constraints and the impulsive constraints including the types of inert, elastic, and live. As a result, the velocity variations for virtual motion must be compatible with the conditions induced from these kinematic conditions. Substituting Eq. (13) into Eq. (14), and noting the independency of $\{\delta_1\dot{\phi}'_\sigma\}_{\sigma=1,\cdots,m'}$, we obtain

$$\sum_{i\in I}(m_i\ddot{\mathbf{r}}_i-\mathbf{F}^{(A)}_i)\cdot\beta'_{i\sigma}=0,\quad \sigma=1,\cdots,m',$$

where the acceleration of each particle P_i may be very large during the impulsive motion, and leads to the jump of the corresponding velocity.

To obtain the relation between the prior motion and the posterior motion for the system $\mathcal{S}$, the previous equations are further expressed as

$$\frac{d}{dt}\left(\sum_{i\in I} m_i\dot{\mathbf{r}}_i\cdot\beta'_{i\sigma}\right)-\sum_{i\in I} m_i\dot{\mathbf{r}}_i\cdot\left(\sum_{j=1}^{n}\frac{\partial\beta'_{i\sigma}}{\partial q_j}\dot{q}_j+\frac{\partial\beta'_{i\sigma}}{\partial t}\right)$$

$$-\sum_{i\in I}\mathbf{F}_i^{(A)}\cdot\beta'_{i\sigma}=0,\quad \sigma=1,\cdots,m'. \tag{15}$$

Since $\beta'_{i\sigma}$ is continuous and the velocities, $\dot{\mathbf{r}}_i$, $i\in I$, and $\dot{q}_j$, $j=1,\cdots,n$ are finite, the second term in the previous equation is finite for each $\sigma=1,\cdots,m'$. Let the σth *virtual-privileged momentum* of $\mathcal{S}$ and the σth *virtual-privileged applied force* acting on $\mathcal{S}$ be defined as

$$\mathsf{L}_\sigma\triangleq\sum_{i\in I} m_i\dot{\mathbf{r}}_i\cdot\beta'_{i\sigma}(\mathbf{q},t), \tag{16}$$

$$\mathsf{F}_\sigma^{(A)}\triangleq\sum_{i\in I}\mathbf{F}_i^{(A)}\cdot\beta'_{i\sigma}(\mathbf{q},t), \tag{17}$$

respectively. Integrating Eq. (15) from $\overline{t}-\tau$ to $\overline{t}+\tau$, and letting $\tau\to 0$, the second term vanishes, and the *privileged impulse-momentum equations* can be established:

$$\mathsf{L}_\sigma^{(+)}-\mathsf{L}_\sigma^{(-)}=\mathsf{P}_\sigma^{(A)},\quad \sigma=1,\cdots,m', \tag{18}$$

where

$$\mathsf{L}_\sigma^{(-)}\triangleq\sum_{i\in I} m_i\mathbf{v}_i^{(-)}\cdot\beta'_{i\sigma}(\overline{\mathbf{q}},\overline{t}), \tag{19}$$

$$\mathsf{L}_\sigma^{(+)}\triangleq\sum_{i\in I} m_i\mathbf{v}_i^{(+)}\cdot\beta'_{i\sigma}(\overline{\mathbf{q}},\overline{t}), \tag{20}$$

denote the σth *prior and posterior virtual-privileged momenta*, respectively, and

$$\mathsf{P}_\sigma^{(A)}\triangleq\int_{\overline{t}-}^{\overline{t}+}\mathsf{F}_\sigma^{(A)}dt=\int_{\overline{t}-}^{\overline{t}+}\sum_{i\in I}\mathbf{F}_i^{(A)}\cdot\beta'_{i\sigma}(\mathbf{q},t)dt \tag{21}$$

is the σth *virtual-privileged applied impulse*.

Again, by the continuity of $\beta'_{i\sigma}$ and the negligible change of configuration and interval of time during the impulsive motion, the virtual-privileged applied impulses can be further expressed as

$$\mathsf{P}_\sigma^{(A)}=\sum_{i\in I}\left(\int_{\overline{t}-}^{\overline{t}+}\mathbf{F}_i^{(A)}dt\right)\cdot\beta'_{i\sigma}(\overline{\mathbf{q}},\overline{t})=\sum_{i\in I}\mathbf{P}_i^{(A)}\cdot\overline{\beta}'_{i\sigma},$$
$$\sigma=1,\cdots,m', \tag{22}$$

where $\mathbf{P}_i^{(A)}\triangleq\int_{\overline{t}-}^{\overline{t}+}\mathbf{F}_i^{(A)}dt$ is the *applied impulse* acting on particle P_i, $i\in I$. Substituting the previous equation into (18), the privileged impulse-momentum equations are rewritten as

$$\sum_{i\in I}(m_i\mathbf{v}_i^{(+)}-m_i\mathbf{v}_i^{(-)}-\mathbf{P}_i^{(A)})\cdot\beta'_{i\sigma}(\overline{\mathbf{q}},\overline{t})=0,\quad \sigma=1,\cdots,m'. \tag{23}$$

From Eq. (22), given applied impulses acting on $\mathcal{S}$, the virtual-privileged applied impulses can be determined, and the jumps of the m' virtual-privileged momenta defined as

$$\Delta\mathsf{L}_\sigma\triangleq\mathsf{L}_\sigma^{(+)}-\mathsf{L}_\sigma^{(-)},\quad \sigma=1,\cdots,m', \tag{24}$$

are then obtained from Eq. (23).

The above derivation of the privileged impulse-momentum equations is based on the framework for finite-force problems. To gain more direct insight on the impulsive motion subject to constraints, and to avoid the technicalities in the transition from finite-force problems to impulsive problems, one may follow the principle of velocity variations for constrained impulsive motion as stated below:

Consider a system of particles P_i, $i\in I$, connected with one another in any way. The system may be subjected to applied impulses $\mathbf{P}_i^{(A)}$ on particle P_i. These applied impulses would impart to the free particles certain determinate jumps on their velocities. However, due to the constraints, regular or impulsive, the actual jumps are different from those on the free motions. Conceive that the applied impulse $\mathbf{P}_i^{(A)}$ be resolved into the effective impulse $\mathbf{P}_i^{(E)}=m_i\mathbf{v}_i^{(+)}-m_i\mathbf{v}_i^{(-)}$, and another component termed the net applied impulse. Owing to the constraints, only the effective impulses generate the actual jumps. The net applied impulses are incapable to change the jumps during the impulsive motion, and the sum total of the product of the net applied impulses and the velocity variations compatible with the constraints during the virtual motion must vanish.

This principle is an analog of the principle of virtual power for finite-force motion stated in [26], and the statements are similar to the exposition of D'Alembert's principle by Mach [37]. It is noted that Appell in [1] enunciated a fundamental principle for impulsive motion analogous to D'Alembert's principle, and, based on which, established the general variational equation in terms of virtual displacements. However, as discussed before, a virtual displacement refers to a variation of the position of the particle, which is in fact not allowed during the impulsive motion. While the variational equation corresponds to the previous principle has similar characteristics as that derived by Appell, the notion of velocity variation while keeping position fixed is more acceptable for the impulsive motion.

Based on this fundamental principle, the corresponding variational equation for impulsive motion can be immediately expressed as

$$\sum_{i\in I}(m_i\mathbf{v}_i^{(+)}-m_i\mathbf{v}_i^{(-)}-\mathbf{P}_i^{(A)})\cdot\delta_1\dot{\mathbf{r}}_i=0, \tag{25}$$

for all $\delta_1\dot{\mathbf{r}}_i$ compatible with the constraints specified during the virtual motion at $(\overline{\mathbf{q}},\overline{t})$. Substituting (13) into the previous variational equation and noting the independency of the variations of the virtual-privileged velocities, the privileged impulse-momentum Eqs. (23) immediately follow. This process may be viewed as the projection from the space of $\delta_1\dot{\mathbf{r}}_i$ to that of $\delta_1\dot{\phi}'_\sigma$ through Eq. (13), and is essentially the basic idea for the so-called projection method, [29,30]. Similar techniques were also used in [12,31] to derive jump conditions from a variational equation similar to (25) in virtual displacement without applied impulse. There the projection is from the tangent bundle to the distribution that annihilates the constraint submanifold. Since the velocity variations should reside in the second tangent bundle, or the jet space of order 2, the projection through $\beta'_{i\sigma}$ should be regarded as different from those in the tangent bundle.

Moreover, it is noted that the projection in Eq. (13) does not depend on the velocity. This is due to the fact that only linear kinematic constraints are treated in this paper. If the system is subject to nonlinear ones in the form of $\psi(x_1,\cdots,x_{3N},\dot{x}_1,\cdots,\dot{x}_{3N},t)=0$, the projection or the transformation may depend on the velocities. For impulsive problems, the discontinuity of the velocity then renders the transformation indeterminant. In [12], the impulsive constraints are associated with Chetaev bundle, [39], of reaction forces from the constraints, linear or nonlinear. However, if the kinematic constraint is nonlinear, it is not clear how to compute the corresponding impulse of these Chetaev forces, which depends on the velocity in general. As a result, more understandings on the mechanism for the impulsive motion subject to nonlinear kinematic constraints are desired.

5 Prior-Value Problems

We are now ready to apply the privileged impulse-momentum equations to study constrained impulsive problems. In the absence of applied impulses, from (18), we immediately have the following *conservation law for constrained impulsive motion*: the virtual-privileged momenta of the system are conserved if there is no applied impulse.

For the prior-value problem, the prior virtual-privileged momentum can be computed from the state of the prior motion through Eq. (19). The posterior virtual-privileged momentum is then determined from the privileged impulse-momentum Eq. (23). To find the posterior motion, the relation between the m' posterior virtual-privileged momenta and the m^a posterior-privileged velocities is constructed by substituting (12) into (20),

$$\mathsf{L}^{(+)}_{\sigma'}=\sum_{\sigma=1}^{m^a}\left(\sum_{i\in I} m_i\overline{\beta}'_{i\sigma'}\cdot\overline{\beta}^a_{i\sigma}\right)\dot{\phi}^a_\sigma(\overline{t}+)+\sum_{i\in I} m_i\overline{\beta}'_{i\sigma'}\cdot\gamma^a_i(\overline{t}+)$$

$$=\sum_{\sigma=1}^{m^a}\mathsf{M}_{\sigma'\sigma}\dot{\phi}^a_\sigma(\overline{t}+)+\mathsf{L}^{(+)}_{\sigma'0},\quad \sigma'=1,\cdots,m', \tag{26}$$

where

$$\mathsf{M}_{\sigma'\sigma}\triangleq\sum_{i\in I} m_i\overline{\beta}'_{i\sigma'}\cdot\overline{\beta}^a_{i\sigma},\quad \mathsf{L}^{(+)}_{\sigma'0}\triangleq\sum_{i\in I} m_i\overline{\beta}'_{i\sigma'}\cdot\gamma^a_i(\overline{t}+).$$

Here, and in what follows, the overbar of a continuous quantity denotes its value at $(\overline{\mathbf{q}},\overline{t})$, e.g., $\overline{\beta}^a_{i\sigma}=\beta^a_{i\sigma}(\overline{\mathbf{q}},\overline{t})$.

The previous set of m' Eqs. (26) is to be solved for the m^a $(=n-L-l_{\mathcal{L}}-l_{\mathcal{I}})$ unknowns, $\dot{\phi}^a_\sigma(t+)$, $\sigma=1,\cdots,m^a$. The difference between the number of equations and that of unknowns is $l_{\mathcal{E}}$, which is the number of elastic constraints. As a result, it is not possible to determine the posterior motion without additional $l_{\mathcal{E}}$ conditions if elastic constraints exist. Such conditions may be obtained by invoking the law of impact, cf. [2,19], or [27,40] for more recent developments. For example, to determine the posterior motion of a ball being dropped to the floor, the coefficient of restitution κ is needed to find the extend of rebound. The elastic constraint is $\dot{\mathbf{r}}\cdot\mathbf{n}=0$, where $\mathbf{n}$ is the normal vector corresponding the floor, and the posterior velocity is related to the prior velocity as

$$\mathbf{v}^{(+)}\cdot\mathbf{n}=-\kappa\mathbf{v}^{(-)}\cdot\mathbf{n}.$$

From the previous example, it may be conceived that for the elastic constraints, $\Sigma_{j=1}^n B^{\mathcal{E}}_{sj}\dot{q}_j+b^{\mathcal{E}}_s=0$, $s=1,\cdots,l_{\mathcal{E}}$, there are $l_{\mathcal{E}}$ *generalized coefficients of restitution* κ_s to bear the relationship between the "approaching" constraint functions and the "leaving" constraint functions as

$$\sum_{j=1}^n \overline{B}^{\mathcal{E}}_{sj}\dot{q}^{(+)}_j+\overline{b}^{\mathcal{E}}_s=-\kappa_s\left(\sum_{j=1}^n \overline{B}^{\mathcal{E}}_{sj}\dot{q}^{(-)}_j+\overline{b}^{\mathcal{E}}_s\right),\quad s=1,\cdots,l_{\mathcal{E}},$$

where the coefficients and the inhomogeneous terms are assumed to be continuous. From the formula of the posterior generalized velocities, with coefficients listed in Table 1, the previous equations are further expressed as

$$\sum_{\sigma=1}^{m^a}\left(\sum_{j=1}^n \overline{B}^{\mathcal{E}}_{sj}\overline{D}^a_{j\sigma}\right)\dot{\phi}^a_\sigma(\overline{t}+)$$

$$=-\sum_{j=1}^n \overline{B}^{\mathcal{E}}_{sj}(d^a_j(\overline{t}+)+\kappa_s\dot{q}^{(-)}_j)-(1+\kappa_s)\overline{b}^{\mathcal{E}}_s,\quad s=1,\cdots,l_{\mathcal{E}}. \tag{27}$$

Equations (26), (27) can be then combined to obtain the m^a posterior-privileged velocities, as illustrated by the following simple example.

Example 1. Consider a particle of mass m being tossed to the floor $(z=0)$ with prior velocity $v_x^{(-)}\mathbf{e}_x+v_z^{(-)}\mathbf{e}_z$. Assume that the surface of the floor is ideally smooth, and the coefficient of restitution of the contact is κ. It is desired to determine the posterior velocity of the particle.

The impulsive constraint provided by the floor is $\dot{z}=0$, which is an elastic constraint. Hence, the virtual motion can be expressed as $\dot{\mathbf{r}}=\dot{x}\mathbf{e}_x$, where $\dot{x}$ can be selected as the privileged velocity for the virtual motion, and the coefficient corresponding to $\dot{x}$ is $\overline{\beta}'_1=\mathbf{e}_x$. The prior and posterior virtual-privileged momentum can be found to be $\mathsf{L}_1^{(-)}=(mv_x^{(-)}\mathbf{e}_x+mv_z^{(-)}\mathbf{e}_z)\cdot\mathbf{e}_x=mv_x^{(-)}$, $\mathsf{L}_1^{(+)}=(mv_x^{(+)}\mathbf{e}_x+mv_z^{(+)}\mathbf{e}_z)\cdot\mathbf{e}_x=mv_x^{(+)}$, respectively. The conservation of the virtual-privileged momentum leads to $v_x^{(-)}=v_x^{(+)}$. However, this equation is not sufficient to determine all the posterior velocities. The additional condition from the law of impact, i.e., $v_z^{(+)}=-\kappa v_z^{(-)}$, is needed to obtain the posterior velocity $v_x^{(-)}\mathbf{e}_x-\kappa v_z^{(-)}\mathbf{e}_z$. □

In contrast to the previous case for an impulsive motion without elastic constraint, we have the same number of unknowns and equations in (26), and the coefficients $\overline{\beta}^a_{i\sigma}=\overline{\beta}'_{i\sigma}$ if the same set of virtual-privileged velocities and posterior-privileged velocities is chosen. Accordingly, the matrix M formed by the components

$$\mathsf{M}_{\sigma'\sigma}=\sum_{i\in I} m_i\overline{\beta}'_{i\sigma'}\cdot\overline{\beta}'_{i\sigma},\quad \sigma,\sigma'=1,\cdots,m', \tag{28}$$

is positive definite and shall be termed the *privileged mass matrix*. Under this framework, the posterior-privileged velocities can be obtained from the following equations:

$$\dot{\phi}^{a(+)}_\sigma=\sum_{\sigma'=1}^{m'}(\mathsf{M}^{-1})_{\sigma\sigma'}(\mathsf{L}^{(+)}_{\sigma'}-\mathsf{L}^{(+)}_{\sigma'0}),\quad \sigma=1,\cdots,m', \tag{29}$$

from which the posterior velocities can be determined from (12).

For the treatment of released constraints, the above deduction of determinacy is apparently different from that claimed in [8], which regards the third type of impulsive problem in [1] as indeterminant. To illustrate the difference, we consider again the problem of releasing a ball from hand grasping. As described in Section 2, the constraint of grasping is treated as a released constraint, which is effective before but ineffective during and after the impulsive motion. Since there is no kinematic constraint imposed in the virtual motion, the virtual-privileged velocities of the ball can be selected to be the three components of the linear velocity of the ball, and the corresponding virtual-privileged momenta are nothing but the three components of the linear momentum $M\mathbf{v}$. If hand does not provide additional applied impulse, from the conservation law of virtual-privileged momenta, we have $M\mathbf{v}^{(-)}=M\mathbf{v}^{(+)}=0$ if the ball is released from rest. The posterior motion is then determined, and the motion of the ball bears no sudden change during the impulsive motion, only experiences sudden increase of the degree-of-freedom. From this simple example, it is seen that the notion of released constraint introduced in this paper is more appropriate to be adopted to treat the released constraints from grasping.

Now we consider the special case that there is neither inert constraints nor released constraints existing, while some live constraints appear. For such impulsive motion, as discussed in the previous section, the same set of privileged velocities may be chosen for the prior motion and the posterior motion, $\dot{\phi}^b_\sigma=\dot{\phi}^a_\sigma$, where $\sigma=1,\cdots,m^a(=m^b=n-L-l_{\mathcal{L}})$, and we have

$$\mathbf{v}_i^{(-)}=\sum_{\sigma=1}^{m^a}\overline{\beta}^a_{i\sigma}\dot{\phi}^{a(-)}_\sigma+\gamma^a_i(\overline{t}-), \tag{30}$$

$$\mathbf{v}_i^{(+)}=\sum_{\sigma=1}^{m^a}\overline{\beta}^a_{i\sigma}\dot{\phi}^{a(+)}_\sigma+\gamma^a_i(\overline{t}+),\quad i\in I. \tag{31}$$

The privileged impulse-momentum Eqs. (23) then implies that

$$\sum_{\sigma=1}^{m^a}\mathsf{M}_{\sigma'\sigma}\Delta\dot{\phi}_\sigma+\Delta\mathsf{L}_{\sigma'0}-\mathsf{P}^{(A)}_{\sigma'}=0,\quad \sigma'=1,\cdots,m', \tag{32}$$

where $\Delta\dot{\phi}_\sigma = \dot{\phi}_\sigma^{a(+)} - \dot{\phi}_\sigma^{a(-)}$, and $\Delta \mathsf{L}_{\sigma' 0} = \mathsf{L}_{\sigma' 0}^{(+)} - \mathsf{L}_{\sigma' 0}^{(-)}$. Moreover, if there is no elastic constraint, we have $m' = m^a = m^b = n - L - l_{\mathcal{L}}$, and the jumps on the privileged velocities can be found from

$$\Delta\dot{\phi}_\sigma = \sum_{\sigma'=1}^{m'} (\mathsf{M}^{-1})_{\sigma\sigma'}(-\Delta \mathsf{L}_{\sigma' 0} + \mathsf{P}_{\sigma'}^{(A)}), \quad \sigma = 1, \cdots, m'. \tag{33}$$

With the prior quantities $\dot{\phi}_\sigma^{a(-)}$ being determined from the prior motion, the posterior quantities $\dot{\phi}_\sigma^{a(+)}$ are then computed from the previous equation, which, in turn, yield the posterior velocities of the system through Eq. (12).

After systematically discussing the impulsive motion associated with various impulsive constraints, the special case that the system is subject to applied impulses with only regular constraints being active is considered next. Since the kinematic conditions are the same for the prior, the virtual, and the posterior motion, the same set of privileged velocities can be chosen, denote by $\dot{\phi}_1, \cdots, \dot{\phi}_m$, with the same degree-of-freedom $m = n - L$. The corresponding coefficients for the velocity of each particle in three phases are thus the same, denoted by $\{\beta_{i\sigma}, \gamma_i\}_{i \in I, \sigma = 1, \cdots, m}$. Similar to the special treatment on the live constraints, the privileged impulse-momentum equations become, cf. Eq. (32),

$$\sum_{\sigma=1}^{m} \mathsf{M}_{\sigma'\sigma} \Delta\dot{\phi}_\sigma = \mathsf{P}_{\sigma'}^{(A)}, \quad \sigma' = 1, \cdots, m, \tag{34}$$

since $\Delta \mathsf{L}_{\sigma' 0} = 0$ due to the nonexistence of live constraints. By computing the privileged applied impulses from the formula

$$\mathsf{P}_{\sigma'}^{(A)} = \sum_{i \in I} \mathbf{P}_i^{(A)} \cdot \overline{\beta}_{i\sigma'}, \quad \sigma' = 1, \cdots, m, \tag{35}$$

and the privileged mass matrix from

$$\mathsf{M}_{\sigma\sigma'} = \sum_{i \in I} m_i \overline{\beta}_{i\sigma} \cdot \overline{\beta}_{i\sigma'}, \quad \sigma, \sigma' = 1, \cdots, m, \tag{36}$$

the jumps on the privileged velocities can be determined from Eq. (34), and the posterior velocities of the system can be found subsequently from Eq. (12).

6 Impulsive Motion for a System Containing Rigid Bodies

A finite degree-of-freedom mechanical system $\mathcal{S}$ may be composed of many particles and rigid bodies. The physical quantities such as the virtual-privileged momenta and the virtual-privileged applied impulses can be defined for each subsystem. Due to the superposition property of the Eq. (23) with respect to the subsystems, a physical quantity for the system $\mathcal{S}$ can be obtained by taking the summation of the corresponding ones of all subsystems.

The motion of a rigid body $\mathcal{B}$ in $\mathcal{S}$ may be described by the translational motion of some reference point Q in $\mathcal{B}$ and a rotation about Q, represented by the rotation dyadic Φ. The rotation dyadic Φ changes the attitude of the rigid body from a reference (initial) configuration to the current configuration, and satisfies the property

$$\Phi \cdot \Phi^c = \mathbb{1}, \quad \det \Phi = 1,$$

where Φ^c denotes the conjugate of Φ, and $\mathbb{1}$ is the identity dyadic. It can be shown from the previous defining property that

$$\dot{\Phi} = \boldsymbol{\omega} \times \Phi, \tag{37}$$

in dyadic notation, [41], where $\boldsymbol{\omega}$ is the angular velocity of the rigid body. The position vector of any point $P \in \mathcal{B}$ can be then expressed as

$$\mathbf{r}_P(t) = \mathbf{r}_Q(t) + \Phi(t) \cdot \mathbf{R}_{PQ},$$

where $\mathbf{R}_{PQ}$ denotes the vector from P to Q at the reference configuration. Taking the time-derivative of the previous equation, we obtain

$$\dot{\mathbf{r}}_P = \dot{\mathbf{r}}_Q + \boldsymbol{\omega} \times \Phi \cdot \mathbf{R}_{PQ} = \dot{\mathbf{r}}_Q + \boldsymbol{\omega} \times (\mathbf{r}_P - \mathbf{r}_Q). \tag{38}$$

Now let $(q_1, \cdots, q_n)$ be a set of generalized coordinates of $\mathcal{S}$. In general, the rotation dyadic Φ may be written in terms of the generalized coordinates as $\Phi = \Phi(q_1, \cdots, q_n, t)$, and its time-derivative is

$$\dot{\Phi} = \sum_{i=1}^{n} \frac{\partial \Phi}{\partial q_i} \dot{q}_i + \frac{\partial \Phi}{\partial t}.$$

From the previous equation and Eq. (37), the expression of the angular velocity $\boldsymbol{\omega}$ in terms of the generalized velocities can be derived from the following identity:

$$\mathbb{1} \times \boldsymbol{\omega} = \dot{\Phi} \cdot \Phi^C = \sum_{i=1}^{n} \left(\frac{\partial \Phi}{\partial q_i} \cdot \Phi^C \right) \dot{q}_i + \frac{\partial \Phi}{\partial t} \cdot \Phi^C,$$

as

$$\boldsymbol{\omega} = \sum_{i=1}^{n} \hat{\boldsymbol{\omega}}_i(\mathbf{q}, t) \dot{q}_i + \hat{\boldsymbol{\omega}}_0(\mathbf{q}, t). \tag{39}$$

In general, the right-hand side in the previous equation is not integrable, and the components of $\boldsymbol{\omega}$ with respect to certain frame can be treated as quasi-velocities of the system, cf. Eq. (7).

Consider the virtual motion of $\mathcal{S}$ during a constrained impulsive motion. With the appropriate selection of virtual-privileged velocities, $\dot{\phi}'_\sigma$, the velocities of the reference point Q and each particle P in $\mathcal{B}$ can be expressed in the form of (11) with coefficients (γ'_Q, β'_{Q0}), $(\beta'_{P\sigma}, \gamma'_P)$, $\sigma = 1, \cdots, m'$, respectively, and the angular velocity of $\mathcal{B}$ is further written as

$$\boldsymbol{\omega} = \sum_{\sigma=1}^{m'} \boldsymbol{\varpi}'_\sigma \dot{\phi}'_\sigma + \boldsymbol{\varpi}'_0. \tag{40}$$

Substituting these expressions into Eq. (38), it follows that

$$\sum_{\sigma=1}^{m'} (\beta'_{P\sigma} - \beta'_{Q\sigma}) \dot{\phi}'_\sigma + (\gamma'_P - \gamma'_Q) = \sum_{\sigma=1}^{m'} \boldsymbol{\varpi}'_\sigma \times (\mathbf{r}_P - \mathbf{r}_Q) \dot{\phi}'_\sigma + \boldsymbol{\varpi}'_0 \times (\mathbf{r}_P - \mathbf{r}_Q),$$

which leads to

$$\beta'_{P\sigma} - \beta'_{Q\sigma} = \boldsymbol{\varpi}'_\sigma \times (\mathbf{r}_P - \mathbf{r}_Q), \quad \gamma'_P - \gamma'_Q = \boldsymbol{\varpi}'_0 \times (\mathbf{r}_P - \mathbf{r}_Q), \tag{41}$$

for $\sigma = 1, \cdots, m'$, by the independency and the arbitrariness of the virtual-privileged velocities.

These relations, which based on the kinematic properties of a rigid body, are now applied to derive the prior and posterior virtual-privileged momenta, and virtual-privileged applied impulses for the rigid body $\mathcal{B}$ as follows. From Eqs. (19), (20), with the index set I being over $P \in \mathcal{B}$, we have

$$\begin{aligned} \mathsf{L}_\sigma^{\mathcal{B}(\pm)} &\triangleq \int_{\mathcal{B}} \mathbf{v}_P^{(\pm)} \cdot \overline{\beta}'_{P\sigma} dm(P) \\ &= \int_{\mathcal{B}} \mathbf{v}_P^{(\pm)} \cdot [\overline{\beta}'_{Q\sigma} + \overline{\boldsymbol{\varpi}}'_\sigma \times (\mathbf{r}_P - \mathbf{r}_Q)] dm(P) \\ &= \mathbf{L}^{\mathcal{B}(\pm)} \cdot \overline{\beta}'_{Q\sigma} + \mathbf{H}_Q^{\mathcal{B}(\pm)} \cdot \overline{\boldsymbol{\varpi}}'_\sigma, \quad \sigma = 1, \cdots, m', \end{aligned} \tag{42}$$

where

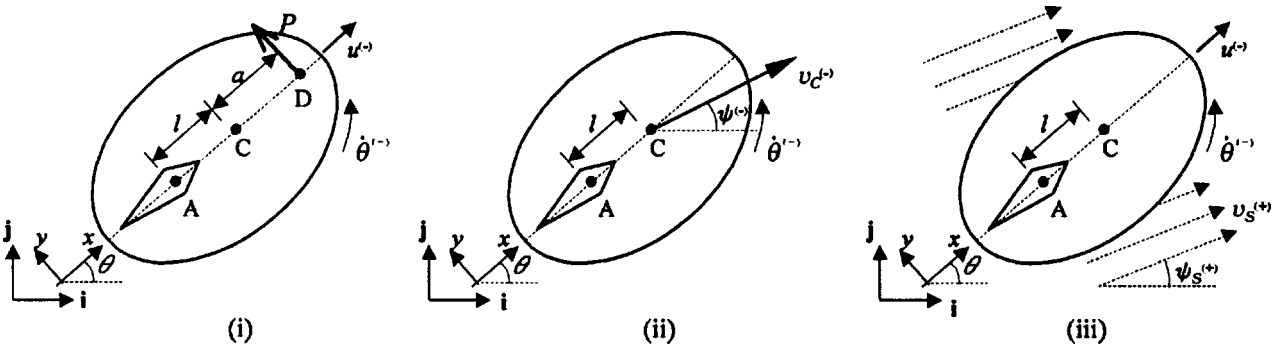


Fig. 2 The impulsive motion of a sleigh with a knifeblade

$$\mathbf{L}^{\mathcal{B}(\pm)}=\int_{\mathcal{B}}\mathbf{v}_P^{(\pm)}dm(P) \tag{43}$$

are the prior and posterior linear momenta of $\mathcal{B}$, respectively, and

$$\mathbf{H}_Q^{\mathcal{B}(\pm)}=\int_{\mathcal{B}}(\mathbf{r}_P-\mathbf{r}_Q)\times\mathbf{v}_P^{(\pm)}dm(P) \tag{44}$$

are the prior and posterior angular momenta of $\mathcal{B}$ about the reference point Q, respectively. Let C denote the center of mass of the rigid body $\mathcal{B}$, and M, I_Q be the total mass and the moment of inertia dyadic of the rigid body about Q, respectively. The linear momenta and the angular momenta defined in (43), (44) can be further expressed as

$$\mathbf{L}^{\mathcal{B}(\pm)}=M\mathbf{v}_C^{(\pm)}, \tag{45}$$

$$\mathbf{H}_Q^{\mathcal{B}(\pm)}=I_Q\cdot\boldsymbol{\omega}^{(\pm)}+(\mathbf{r}_C-\mathbf{r}_Q)\times(M\mathbf{v}_Q^{(\pm)}), \tag{46}$$

respectively.

On the other hand, the virtual-privileged applied impulses corresponding to $\mathcal{B}$ can be found to be, using Eq. (41),

$$\mathsf{P}_\sigma^{\mathcal{B}_{(A)}}\triangleq\sum_{P\in\mathcal{B}}\mathbf{P}_P^{(A)}\cdot\overline{\boldsymbol{\beta}}'_{P\sigma}=\sum_{P\in\mathcal{B}}\mathbf{P}_P^{(A)}\cdot[\overline{\boldsymbol{\beta}}'_{Q\sigma}+\overline{\boldsymbol{\varpi}}'_\sigma\times(\mathbf{r}_P-\mathbf{r}_Q)]$$

$$=\mathbf{P}^{\mathcal{B}_{(A)}}\cdot\overline{\boldsymbol{\beta}}'_{Q\sigma}+\mathbf{J}_Q^{\mathcal{B}_{(A)}}\cdot\overline{\boldsymbol{\varpi}}'_\sigma,\quad \sigma=1,\cdots,m', \tag{47}$$

where $\mathbf{P}^{\mathcal{B}_{(A)}}$, $\mathbf{J}_Q^{\mathcal{B}_{(A)}}$ are the resultant applied linear impulse, and the applied angular impulse acting on $\mathcal{B}$ about Q, defined as

$$\mathbf{P}^{\mathcal{B}_{(A)}}=\sum_{P\in\mathcal{B}}\mathbf{P}_P^{(A)}, \tag{48}$$

$$\mathbf{J}_Q^{\mathcal{B}_{(A)}}=\sum_{P\in\mathcal{B}}(\mathbf{r}_P-\mathbf{r}_Q)\times\mathbf{P}_P^{(A)}, \tag{49}$$

respectively.

After obtaining the formula for a single rigid body, the virtual-privileged momenta and applied impulses for the system consisting many bodies can be constructed. Let the system be composed of N_B rigid bodies, which may be expressed as

$$\mathcal{S}=\bigcup_{i=1}^{N_B}\mathcal{B}_i.$$

For each body $i, i=1,\ldots N_B$, we may choose a reference point Q_i, and denote the attitude dyadic by Φ_i. With the selection of privileged velocities of $\mathcal{S}$ in the virtual motion, the coefficients $(\boldsymbol{\varpi}'_\sigma)^{\mathcal{B}_i}$ and $\boldsymbol{\beta}'_{Q_i\sigma}$ for each body can be determined. The virtual-privileged momenta and the applied impulses of the system are then found from the following summation:

$$\mathsf{L}_\sigma^{(\pm)}=\sum_{i=1}^{N_B}\mathsf{L}_\sigma^{\mathcal{B}_i(\pm)},\quad \mathsf{P}_\sigma^{(A)}=\sum_{i=1}^{N_B}\mathsf{P}_\sigma^{\mathcal{B}_{i(A)}}, \tag{50}$$

respectively. With these formulas, the privileged impulse-momentum equations developed in Section 4 can be invoked to solve the constrained impulsive problems associated with interconnected rigid bodies. An example is given in the next section to illustrate the process.

7 The Impulsive Motion of a Sleigh With a Knifeblade

The proposed methodology discussed in the previous sections is now applied to study the impulsive motion of a sleigh with a knifeblade attached. Three possibilities of impulsive motions due to applied impulse, inert constraint, and live constraint, respectively, are discussed. Two alternative methods, namely, Newton's method and Lagrange's method, are used to solve the same problem for the case of inert constraint to describe the essential differences.

Consider a sleigh $\mathcal{B}$ with a knifeblade along its principle axis, as depicted in Fig. 2. Let its mass and moment of inertia about its center of mass C be denoted by M and I_C, respectively. When the sleigh moves on a horizontal ice surface S, it is assumed that the knifeblade makes contact with the surface at point A, and it is possible for the sleigh to move freely in the direction along the blade, and rotate freely about A. However, the motion of the sleigh in the direction perpendicular to the blade is prohibited. Let $\{\mathbf{i},\mathbf{j}\}$ be a fixed coordinate system, and $\{\mathbf{e}_x,\mathbf{e}_y\}$ be a coordinate system moving with $\mathcal{B}$, where $\mathbf{e}_x$ is parallel to the direction of the blade, which makes an angle θ with respect to the axis $\mathbf{i}$. Due to the physical condition of constraints, the velocity of A satisfies

$$\mathbf{v}_A\cdot\mathbf{e}_y=0, \tag{51}$$

and may be expressed as

$$\mathbf{v}_A=u\mathbf{e}_x, \tag{52}$$

where u is termed the *longitudinal velocity* of $\mathcal{B}$. The sleigh may rotate about the vertical axis $\mathbf{e}_z(=\mathbf{e}_x\times\mathbf{e}_y)$ with angular velocity $\dot\theta\mathbf{e}_z$. The velocity of C may be then expressed as

$$\mathbf{v}_C=u\mathbf{e}_x+l\dot\theta\mathbf{e}_y, \tag{53}$$

where l denotes the distance between C and A. For this problem, we may choose the coordinates of C, i.e., (x_C,y_C), and the angle θ as the generalized coordinates. The constraints (51) then becomes

$$-\sin\theta\dot x_C+\cos\theta\dot y_C-l\dot\theta=0, \tag{54}$$

which is nonholonomic. The longitudinal velocity u can be expressed as

$$u=\cos\theta\dot x_C+\sin\theta\dot y_C,$$

in which the right-hand side is not integrable, and thus should be treated as a quasi-velocity.

We are now ready to consider three cases of impulsive motions for this system, which occur at the instant $\bar{t}$ and at the configuration of $\theta=\theta_0$.

Case (i). As shown in Fig. 2(*i*), the sleigh is suddenly exerted by an impulse P acting at the point D at the instant $\bar{t}$. The longitudinal velocity of $\mathcal{B}$ and the angular velocity of $\mathcal{B}$ about the vertical axis at the instant $\bar{t}-$ are $u^{(-)}$ and $\dot{\theta}^{(-)}$, respectively. It is desired to determine these two quantities at the instant $\bar{t}+$.

For this case, the impulsive motion is due to the imposition of the applied impulse P, while the rigid body $\mathcal{B}$ is subject to the regular nonholonomic constraint, Eq. (54). One may choose the longitudinal velocity u and the angular velocity $\dot{\theta}$ as the privileged velocities to represent the possible motion of $\mathcal{B}$. In fact, the velocity of A and C can be written in terms of the privileged velocities as in Eqs. (52) and (53), respectively. Since the prior and the posterior motions satisfy the same nonholonomic constraint, the prior and posterior velocities of A and C, $\mathbf{v}_A^{(-)}$, $\mathbf{v}_C^{(-)}$, $\mathbf{v}_A^{(+)}$ and $\mathbf{v}_C^{(+)}$, can be also expressed in terms of the prior and posterior-privileged velocities, $u^{(-)}$, $\dot{\theta}^{(-)}$, $u^{(+)}$, and $\dot{\theta}^{(+)}$, with the same form, respectively.

Let C be selected as the reference point for $\mathcal{B}$. The coefficients of the velocity of C and the angular velocity of $\mathcal{B}$ corresponding to the privileged velocities can be obtained as

$$\bar{\beta}_{C1}=\mathbf{e}_x, \quad \bar{\beta}_{C2}=l\mathbf{e}_y, \qquad \bar{\varpi}_1=0, \quad \bar{\varpi}_2=\mathbf{e}_z, \tag{55}$$

respectively. The linear momenta, and the angular momenta about C given in (43), (44), are then

$$\mathbf{L}^{\mathcal{B}(\pm)}=Mu^{(\pm)}\mathbf{e}_x+Ml\dot{\theta}^{(\pm)}\mathbf{e}_y,$$

$$\mathbf{H}_C^{\mathcal{B}(\pm)}=I_C\dot{\theta}^{(\pm)}\mathbf{e}_z,$$

respectively. On the other hand, the applied linear impulse and the applied angular impulse are, respectively,

$$\mathbf{P}^{\mathcal{B}_{(A)}}=P\mathbf{e}_y, \quad \mathbf{J}_C^{\mathcal{B}_{(A)}}=Pa\mathbf{e}_z.$$

Substituting these terms into the formula of the privileged momenta and privileged applied impulses, the privileged impulse-momentum Eq. (18) can be applied to derive, in matrix form,

$$[\mathsf{M}]\left(\begin{bmatrix}u^{(+)}\\ \dot{\theta}^{(+)}\end{bmatrix}-\begin{bmatrix}u^{(-)}\\ \dot{\theta}^{(-)}\end{bmatrix}\right)=\begin{bmatrix}0\\ P(a+l)\end{bmatrix},$$

where the privileged mass matrix $[\mathsf{M}]$ is given by

$$[\mathsf{M}]=\begin{bmatrix}M & 0\\ 0 & (Ml^2+I_C)\end{bmatrix}. \tag{56}$$

Hence, the posterior longitudinal velocity and angular velocity of the sleigh $\mathcal{B}$ can be determined as

$$\begin{bmatrix}u^{(+)}\\ \dot{\theta}^{(+)}\end{bmatrix}=[\mathsf{M}]^{-1}\begin{bmatrix}0\\ P(a+l)\end{bmatrix}+\begin{bmatrix}u^{(-)}\\ \dot{\theta}^{(-)}\end{bmatrix}=\begin{bmatrix}u^{(-)}\\ \dfrac{a+l}{Ml^2+I_C}P+\dot{\theta}^{(-)}\end{bmatrix}.$$

□

Case (ii). Consider the landing of the sleigh on the ice surface at the instant $\bar{t}$, with prior velocity of the center of mass being $v_C^{(-)}$ in the direction of angle $\psi^{(-)}$ with respect to the axis **i**, and the prior angular velocity about $\mathbf{e}_z$ being $\dot{\theta}^{(-)}$, cf. Fig. 2(*ii*). The velocity of the landing point A may have lateral component before $\bar{t}$, but it immediately satisfies the constraint (51) after contact. This impulsive constraint is thus characterized as an inert constraint. The virtual motion and the posterior motion have the same characteristics as the motion in Case (i). Similar to the discussion there, the set of velocities $(u,\dot{\theta})$ may be chosen as the virtual-privileged velocities, $\dot{\phi}_1'=u$, $\dot{\phi}_2'=\dot{\theta}$. However, they are not the privileged velocities for the prior motion. Again, the point C is chosen as the reference point for $\mathcal{B}$, and the coefficients for the velocity of C and the angular velocity of $\mathcal{B}$ corresponding to the virtual-privileged velocities are the same as in (55), with $\bar{\beta}$, $\bar{\varpi}$ being replaced by $\bar{\beta}'$, $\bar{\varpi}'$ respectively. The required quantities can be found to be

$$\mathbf{L}^{\mathcal{B}_{(-)}}=Mv_C^{(-)}\cos(\psi^{(-)}-\theta_0)\mathbf{e}_x+Mv_C^{(-)}\sin(\psi^{(-)}-\theta_0)\mathbf{e}_y,$$

$$\mathbf{L}^{\mathcal{B}_{(+)}}=Mu^{(+)}\mathbf{e}_x+Ml\dot{\theta}^{(+)}\mathbf{e}_y,$$

$$\mathbf{H}_C^{\mathcal{B}_{(\pm)}}=I_C\dot{\theta}^{(\pm)}\mathbf{e}_z.$$

The prior and the posterior virtual-privileged momenta corresponding respectively to ϕ_1', ϕ_2' are then

$$\mathsf{L}_1^{(-)}=Mv_C^{(-)}\cos(\psi^{(-)}-\theta_0),$$

$$\mathsf{L}_1^{(+)}=Mu^{(+)},$$

$$\mathsf{L}_2^{(-)}=Mlv_C^{(-)}\sin(\psi^{(-)}-\theta_0)+I_C\dot{\theta}^{(-)},$$

$$\mathsf{L}_2^{(+)}=(Ml^2+I_C)\dot{\theta}^{(+)}.$$

Since there is no applied impulse, the conservation of virtual-privileged momenta can be then invoked to yield the posterior velocities as

$$u^{(+)}=v_C^{(-)}\cos(\psi^{(-)}-\theta_0),$$

$$\dot{\theta}^{(+)}=\frac{1}{Ml^2+I_C}(Mlv_C^{(-)}\sin(\psi^{(-)}-\theta_0)+I_C\dot{\theta}^{(-)}). \tag{57}$$

□

Case (iii). If the ice surface exhibits a sudden motion at $\bar{t}$ caused by external agents, such as an earthquake, the sleigh moving on it then experiences an impulsive constraint. Let the velocity of the surface be denoted by $\mathbf{v}_S$. Instead of (51), the constraint becomes now

$$(\mathbf{v}_A-\mathbf{v}_S)\cdot\mathbf{e}_y=0, \tag{58}$$

or

$$\mathbf{v}_A\cdot\mathbf{e}_y-\mathbf{v}_S\cdot\mathbf{e}_y=0, \tag{59}$$

in which the inhomogeneous term makes it become a live impulsive constraint. In particular, let the surface suddenly move with velocity $v_S^{(+)}$ in the direction of angle $\psi_S^{(+)}$ with respect to the axis **i**. The velocity of A and C during the impulsive motion can be then expressed as

$$\mathbf{v}_A=u\mathbf{e}_x+v_S(\cos\psi_S\mathbf{i}+\sin\psi_S\mathbf{j}),$$

$$\mathbf{v}_C=u\mathbf{e}_x+l\dot{\theta}\mathbf{e}_y+v_S(\cos\psi_S\mathbf{i}+\sin\psi_S\mathbf{j}),$$

respectively, in which u is now the longitudinal velocity of the sleigh relative to the surface, and v_S, ψ_S are the magnitude and the direction of $\mathbf{v}_S$, respectively. From the constraint, we have

$$v_S(\bar{t}-)=0, \quad v_S(\bar{t}+)=v_S^{(+)},$$

$$\psi_S(\bar{t}-)=0, \quad \psi_S(\bar{t}+)=\psi_S^{(+)}.$$

For this case, the prior motion, the virtual motion, and the posterior motion have the same degree-of-freedom, and the same set of privileged velocities $\dot{\phi}_1'=u$, $\dot{\phi}_2'=\dot{\theta}$ may be chosen. Again, Eq. (55), after replacing $\bar{\beta}$, $\bar{\varpi}$ by $\bar{\beta}'$, $\bar{\varpi}'$, is applicable, and the quantities of momenta for the body are found as

$$\mathbf{L}^{\mathcal{B}_{(-)}}=M(u^{(-)}\mathbf{e}_x+l\dot{\theta}^{(-)}\mathbf{e}_y),$$

$$\mathbf{L}^{\mathcal{B}_{(+)}}=M[u^{(+)}\mathbf{e}_x+l\dot{\theta}^{(+)}\mathbf{e}_y+v_S^{(+)}(\cos\psi_S^{(+)}\mathbf{i}+\sin\psi_S^{(+)}\mathbf{j})],$$

$$\mathbf{H}_C^{\mathcal{B}(\pm)} = I_C \dot{\theta}^{(\pm)} \mathbf{e}_z .$$

These formulas in turn give rise to the prior and the posterior virtual-privileged momenta corresponding to $\dot{\phi}_1'$, $\dot{\phi}_2'$ as

$$\mathsf{L}_1^{(-)} = M u^{(-)},$$

$$\mathsf{L}_1^{(+)} = M u^{(+)} + M v_s^{(+)} \cos(\psi_s^{(+)} - \theta_0),$$

$$\mathsf{L}_2^{(-)} = (M l^2 + I_C)\, \dot{\theta}^{(-)},$$

$$\mathsf{L}_2^{(+)} = (M l^2 + I_C)\, \dot{\theta}^{(+)} + M l v_s^{(+)} \sin(\psi_s^{(+)} - \theta_0),$$

respectively. The conservation of the virtual-privileged momenta is then invoked to find the posterior longitudinal velocity relative to the surface and the posterior angular velocity as

$$u^{(+)} = u^{(-)} - v_s^{(+)} \cos(\psi_s^{(+)} - \theta_0),$$

$$\dot{\theta}^{(+)} = \dot{\theta}^{(-)} - \frac{Ml}{Ml^2 + I_C} v_s^{(+)} \sin(\psi_s^{(+)} - \theta_0).$$

□

To show the distinct features of the proposed scheme, Newton's method and Lagrange's method for impulsive motion are applied next to solve the same problem as in Case (ii).

As discussed in [2], the integration of Newton-Euler's equations gives rise to the law of impulse for impulsive motion. For the problem of Case (ii), the only impulsive force acting on $\mathcal{B}$ is the one acting on A exerted by the ice surface, which has no component in the direction perpendicular to the blade by the nature of the constraint. Accordingly, we have the conservation of linear momentum in that direction,

$$\mathbf{L}^{(+)} \cdot \mathbf{e}_x = \mathbf{L}^{(-)} \cdot \mathbf{e}_x ,$$

or

$$M u^{(+)} = M v_C^{(-)} \cos(\psi^{(-)} - \theta_0).$$

On the other hand, there is no impulsive torque about A either, and the conservation of angular momentum is then

$$I_C \dot{\theta}^{(+)} \mathbf{e}_z + (\mathbf{r}_C - \mathbf{r}_A) \times (M \mathbf{v}_C^{(+)}) = I_C \dot{\theta}^{(-)} \mathbf{e}_z + (\mathbf{r}_C - \mathbf{r}_A) \times (M \mathbf{v}_C^{(-)}).$$

From the relation $\mathbf{v}_C^{(+)} = u^{(+)} \mathbf{e}_x + l \dot{\theta}^{(+)} \mathbf{e}_y$, the posterior longitudinal velocity and posterior angular velocity can be obtained exactly the same as in Eq. (57).

Although Newton's method seems to be more intuitive for this simple case, it is not straightforward to apply the method to more sophisticated problems. For example, to handle the impulsive problems for constrained multibody systems, it is rather complicated to solve the problems by inspection.

On the other hand, in Lagrange's method, cf. [3,5,18], the equations of impulsive motion are written in terms of the kinetic energy T of the system. The problem studied in Case (i) has been solved by Lagrange's method in [18]. Here, the same method is applied to attack the impulsive problem for Case (ii), and is summarized as follows. Consider the system with kinetic energy T, and subject to $l_{\mathcal{I}}$ inert impulsive constraints at $\bar{t}$. From the impulsive constraints, $n - l_{\mathcal{I}}$ generalized velocities, $\dot{q}_1, \cdots, \dot{q}_{n-l_{\mathcal{I}}}$, can be selected to represent the other ones such that

$$\dot{q}_j = \sum_{\sigma=1}^{n-l_I} D_{j\sigma} \dot{q}_\sigma + d_j, \quad j = n - l_{\mathcal{I}} + 1, \ldots, n. \tag{60}$$

Then the prior motion and the posterior motion satisfy the Lagrange's equations for impulsive motion, i.e., for $\sigma = 1, \cdots, n - l_{\mathcal{I}}$,

$$\left(\frac{\partial T}{\partial \dot{q}_\sigma} + \sum_{j=n-l_x+1}^{n} \frac{\partial T}{\partial \dot{q}_j} D_{j\sigma} \right) \Bigg|_{\bar{t}^+} = \left(\frac{\partial T}{\partial \dot{q}_\sigma} + \sum_{j=n-l_x+1}^{n} \frac{\partial T}{\partial \dot{q}_j} D_{j\sigma} \right) \Bigg|_{\bar{t}^-} . \tag{61}$$

For the previous example, with the generalized coordinates (x_c, y_c, θ), the kinetic energy of the sleigh is simply

$$T = \frac{1}{2} M (\dot{x}_C^2 + \dot{y}_C^2) + \frac{1}{2} I_C \dot{\theta}^2.$$

The inert constraint (54) may be expressed as

$$\dot{\theta} = -\frac{\sin\theta}{l} \dot{x}_C + \frac{\cos\theta}{l} \dot{y}_C . \tag{62}$$

Corresponding to the generalized coordinates x_c, y_c, the equations of the impulsive motion can be derived from (61) as

$$\begin{aligned} M \dot{x}_C^{(+)} - \frac{I_C}{l} \dot{\theta}^{(+)} \sin\theta_0 &= M \dot{x}_C^{(-)} - \frac{I_C}{l} \dot{\theta}^{(-)} \sin\theta_0 , \\ M \dot{y}_C^{(+)} + \frac{I_C}{l} \dot{\theta}^{(+)} \cos\theta_0 &= M \dot{y}_C^{(-)} + \frac{I_C}{l} \dot{\theta}^{(-)} \cos\theta_0 . \end{aligned} \tag{63}$$

The prior velocity of C is $(v_C^{(-)} \cos\psi^{(-)}, v_C^{(-)} \sin\psi^{(-)})$, and the posterior velocity can be expressed in terms of the posterior longitudinal velocity and angular velocity as

$$\dot{x}_C^{(+)} = \cos\theta_0 u^{(+)} - l \sin\theta_0 \dot{\theta}^{(+)},$$

$$\dot{y}_C^{(+)} = \sin\theta_0 u^{(+)} + l \cos\theta_0 \dot{\theta}^{(+)}.$$

Substituting the above formulas into (63), the posterior velocities $u^{(+)}$, $\dot{\theta}^{(+)}$ can be then determined by exactly the same equations as (57).

It is noted that in Lagrange's method, the quasi-velocity u is not introduced directly, which reflects the fact that the treatment of quasi-velocity is a little awkward in Lagrange's formulation. Moreover, the computation of kinetic energy can only accommodate the regular constraints. If the impulsive constraints occur, additional terms have to be included in the equations. Since the proposed methodology somewhat has similar characters as the Kane's approach for the finite-force problem, the comparison of different methods in [28] can be partly transported here. In particular, it has been shown in [27,42] that Kane's approach is very effective in generating equations of motion for multibody systems through symbolic computations. The same advantage of applying the proposed method to deal with impulsive motion of multibody systems subject to impulsive constraints is conceivable accordingly. Using the method to tackle the impulsive constraints, it is not necessary to include the impulsive constraint forces in the equations, which may contain some parameters, e.g., Lagrange's multipliers, to be determined in the process.

8 Conclusions

Although there are many approaches dealing with the constrained impulsive problems for mechanical systems, the principle of velocity variation for impulsive motions lays down the foundation of the methodology discussed in this paper. The classical principles for finite-force problems may not be suitable to be applied directly to the impulsive problems, due to the irregularity of the occurrences of the impulsive constraints. From the fundamental principle, it was shown that, instead of the classical notion of conservation of momenta for systems without applied impulses, the more appropriate law should be the conservation of virtual-privileged momenta, respecting the behavior of the system in the virtual motion. With the inclusion of quasi-velocities, such as the angular velocity of a rigid body, or the longitudinal velocity of a vehicle, the selection of independent privileged velocities seems to be more natural. The methodology discussed in this paper provides a systematic tool in dealing with various types of impulsive constraints, and can be applied to solve a variety of constrained impulsive problems for mechanical systems.

References

[1] Appell, P., 1923, *Traité de Mécanique Rationnelle*, Vol. 2, 4th Ed., Gauthier-Villars, Paris.
[2] Whittaker, E. T., 1937, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge, UK.
[3] Pars, L. A., 1965, *A Treatise on Analytical Dynamics*, Heinemann, London.
[4] Goldstein, H., 1980, *Classical Mechanics*, Addison-Wesley, Redding, MA.
[5] Rosenberg, R. M., 1977, *Analytical Dynamics of Discrete Systems*, Plenum Press, New York.
[6] Meirovitch, L., 1970, *Methods of Analytical Dynamics*, McGraw-Hill, New York.
[7] Kane, T. R., 1962, ''Impulsive Motions,'' ASME J. Appl. Mech., **29**, pp. 715–718.
[8] Papastavridis, J. G., 1989, ''Impulsive Motion of Ideally Constrained Mechanical Systems via Analytical Dynamics,'' Int. J. Eng. Sci., **27**(12), pp. 1445–1461.
[9] Stronge, W. J., 1990, ''Rigid Body Collision With Friction,'' Proc. R. Soc. London, Ser. A, **431**, pp. 169–181.
[10] Souchet, R., 1993, ''Analytical Dynamics of Rigid Body Impulsive Motions,'' Int. J. Eng. Sci., **31**(1), pp. 85–92.
[11] Bahar, L. Y., 1994, ''On The Use of Quasi-Velocities in Impulsive Motion,'' Int. J. Eng. Sci., **32**(11), pp. 1669–1686.
[12] Ibort, A., de Leon, M., Lacomba, E. A., de Diego, D. Martin, and Pitanga, P., 1997, ''Mechanical Systems Subjected to Impulsive Constraints,'' J. Phys. A, **30**, pp. 5835–5854.
[13] Pfeiffer, F., and Glocker, C., 1996, *Multibody Dynamics With Unilateral Contacts*, John Wiley and Sons, New York.
[14] Brogliato, B., 1996, *Nonsmooth Impact Mechanics: Models, Dynamics, and Control*, Springer-Verlag, London.
[15] Lacomba, E. A., and Tulczyjew, W. M., 1990, ''Geometric Formulation of Mechanical Systems With One-sided Constraints,'' J. Phys. A, **23**, pp. 2801–2813.
[16] Genot, F., and Brogliato, B., 1999, ''New Results on Painleve Paradoxes,'' Eur. J. Mech. A/Solids, **18**, pp. 653–677.
[17] van der Schaft, A. J., and Schumacher, J. M., 1999, ''Complementarity Modeling of Hybrid Systems,'' IEEE Trans. Autom. Control, **43**(4), pp. 483–490.
[18] Neimark, Ju. I., and Fufaev, N. A., 1972, *Dynamics of Nonholonomic Systems* (Vol. 33 of *Translations of Mathematical Monographs*) American Mathematical Society, Providence, RI.
[19] Greenwood, D. T., 1988, *Principles of Dynamics*, Prentice-Hall, Englewood Cliffs, NJ.
[20] Ibort, A., de Leon, M., Lacomba, E. A., Marrero, J. C., Martin de Diego, D., and Pitanga, P., 1998, ''Geometric Formulation of Mechanical Systems Subjected to Time-Dependent One-Sided constraints,'' J. Phys. A, **31**, pp. 2655–2674.
[21] Cortés, J., de León, M., de Diego, D. Martín, and Martínez, S., 2001, ''Mechanical Systems Subjected to Generalized Non-holonomic Constraints,'' Proc. R. Soc. London, Ser. A, **457**, pp. 651–670.
[22] Jourdain, P. E. B., 1909, ''Note on an Analogue of Gauss' Principle of Least Constraint,'' Quart. J. Pure Appl. Math., **40**, pp. 153–157.
[23] Roberson, R. E., and Schwertassek, R., 1988, *Dynamics of Multibody Systems*, Springer-Verlag, New York.
[24] Moon, F. C., 1998, *Applied Dynamics*, John Wiley and Sons, New York.
[25] Wang, L. S., and Pao, Y. H., 2003, ''Jourdain's Variational Equation and Appell's Equation of Motion for Nonholonomic Dynamical Systems,'' Am. J. Phys., **73**(1), pp. 72–82.
[26] Pao, Y. H., and Wang, L. S., 2002, ''A Unified Variational Principle for Dynamics of Particles and of Continua,'' *submitted for journal publication*.
[27] Lesser, M., 1995, *The Analysis of Complex Nonlinear Mechanical Systems: A Computer Algebra Assisted Approach*, World Scientific, Singapore.
[28] Kane, T. R., and Levinson, D. A., 1980, ''Formulation of Equations of Motion for Complex Spacecraft,'' AIAA J. Guidance Control, **3**(2), pp. 99–112.
[29] Scott, D., 1988, ''Can a Projection Method of Obtaining Equations of Motion Compete With Lagrange's Equations?'' Am. J. Phys., **56**, pp. 451–456.
[30] Blajer, W., 1992, ''A Projection Method Approach to Constrained Dynamic Analysis,'' ASME J. Appl. Mech., **59**, pp. 643–649.
[31] Ibort, A., de Leon, M., Lacomba, E. A., Marrero, J. C., de Diego, D. Martin, and Pitanga, P., 2001, ''Geometric Formulation of Carnot's Theorem,'' J. Phys. A, **34**, pp. 1691–1712.
[32] Arnold, V. I., 1989, *Mathematical Methods of Classical Mechanics*, 2nd Ed., Springer-Verlag, New York.
[33] Abraham, R., and Marsden, J. E., 1978, *Foundations of Mechanics*, 2nd Ed., Benjamin/Cummings, Reading, MA.
[34] Vershik, A. M., and Faddeev, L. D., 1981, ''Lagrangian Mechanics in Invariant Form,'' Sel. Math. Sov., **1**(4), pp. 339–350.
[35] Tulczyjew, W. M., 1989, *Geometric Formulation of Physical Theories: Statics and Dynamics of Mechanical Systems*, Bibliopolis, Naples.
[36] Gibbs, J. W., 1879, ''On the Fundamental Formulas of Dynamics,'' Am. J. Math., **2**, pp. 49–64.
[37] Mach, E., 1960, *The Science of Mechanics*, Open Court, La Salle, IL.
[38] Arnold, V. I., ed., 1988, *Dynamical Systems III* (Vol. 3 of *Encyclopedia of Mathematical Sciences*), Springer-Verlag, New York.
[39] Chetaev, N. G., 1932, ''On the Principle of Gauss,'' Kazan Univ. Bull. Phys. Math. Soc., **6**, pp. 68–71.
[40] Smith, C. E., and Liu, P. P., 1992, ''Coefficients of Restitution,'' ASME J. Appl. Mech., **59**, pp. 963–969.
[41] Weatherburn, C. E., 1921, *Elementary Vector Analysis*, Open Court, La Salle, IL.
[42] Junkins, John L., ed., 1990, *Mechanics and Control of Large Flexible Structures*, American Institute of Aeronautics and Astronautics, Washington, DC.

Dynamic Response of Kirchhoff Plate on a Viscoelastic Foundation to Harmonic Circular Loads

L. Sun
Assistant Professor,
Transportation and Operation Research Program,
Department of Civil Engineering,
The Catholic University of America,
Washington, DC 20064
e-mail: sun1@cua.edu

In this paper Fourier transform is used to derive the analytical solution of a Kirchhoff plate on a viscoelastic foundation subjected to harmonic circular loads. The solution is first given as a convolution of the Green's function of the plate. Poles of the integrand in the integral representation of the solution are identified for different cases of the foundation damping and the load frequency. The theorem of residue is then utilized to evaluate the generalized integral of the frequency response function. A closed-form solution is obtained in terms of the Bessel and Hankel functions corresponding to the frequency response function of the plate under a harmonic circular load. The result is partially verified by comparing the static solution of a point source obtained in this paper to a well-known result. This analytical representation permits one to construct fast algorithms for parameter identification in pavement nondestructive test. [DOI: 10.1115/1.1577598]

1 Introduction

Nondestructive testing (NDT) has been extensively used in pavement engineering since the 1980s, [1–4], to evaluate pavement structural parameters. The most commonly used NDT device for pavement structural evaluation are falling weight deflectometer and Dynaflect. Falling weight deflectometer applies an impulse load to pavement surfaces, while Dynaflect applies a steady-state vibrating harmonic load, [3]. Given that a typical physical model for rigid pavements (e.g., cement concrete pavements) is a Kirchhoff plate resting on an elastic Winkler foundation, [5,6], the mathematical problem involved here thus becomes to estimate the parameters of governing equation of the plate provided that the applied load is known (i.e., the inverse problem). The structural evaluation is then achieved by identifying structural parameters based on pavement response to applied dynamic loads.

Because of the complexity involved in the inverse problem, in current practice a widely used technique is to use the forward analysis of a plate under a static load. By comparing measured dynamic response and calculated static response using optimization techniques, pavement structural parameters are eventually determined while selecting a pavement structure whose calculated response is most closely to the measured maximum response in terms of certain objective functions, [7]. Clearly, pavement response under dynamic loads such as applied by Dynaflect is significantly different from pavement response under static loads.

Finite element procedures have been developed to calculate numerically the response of a plate to dynamic loads, [8–10]. However, in terms of efficiency, the computation using finite element methods is time-consuming. Computational efficiency can be improved if analytical solutions are available and used for numerical calculation. Achenbach et al. [11] investigated the response of an infinite plate to harmonic plane waves. Freund and Achenbach [12] and Oien [13] investigated the response of a semi-infinite plate on an elastic half-space. In their study the displacement of the plate is assumed to be harmonic. As a result, the time and spatial coordinates become separated and the governing partial differential equation turns out to be an ordinary differential equation. The solution is then obtained using the Bubnov-Galekin method and series expansion of the vibrational modes of the plate. Arnold et al. [14] and Warburton [15] conducted similar studies by means of integral transform methods. Using integral representation of the general solution of the plate provided by Bycroft [16], Krenk and Schmidt [17] studied the steady-state response of finite plate on an elastic half-space.

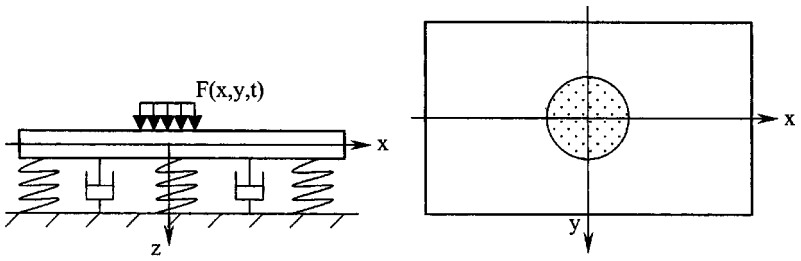
In our previous work, [18–20], dynamic response of Bernoulli-Eular beam (one-dimensional situation) to specific loading condition has been studied. However, the analysis of a Kirchhoff plate on a viscoelastic foundation to a circular harmonic load applied by a Dynaflect has not been available in the literature. To this end, as a continuous effort in this paper the author extends previous work to deal with cases and the dynamic response of a ridge pavement structure (two-dimensional situation) under a circular harmonic load. The availability of such analytical solutions will enable one to construct fast algorithm for parameter identification problem, a core issue in pavement NDT.

This paper is organized as follows. In Section 2, the governing equation is established with the exploration of associated foundation models. In Section 3, the Green's function of a Kirchhoff plate is derived analytically using integral transform method. In Section 4, by integrating the Green's function with respect to time-spatial dimensions, we obtain the frequency response function corresponding to a harmonic circular load and concentrated load. In Section 5 we address several special cases such as the static solution and Winkler foundation, which is also used to verify the correctness of our result.

2 The Governing Equation of the Plate

Figure 1 depicts the coordinate system and significant dimensions. Three assumptions are commonly made to simplify the mathematical model of a Kirchhoff plate. These assumptions are (1) the strain component ε_z in the perpendicular direction of the plate is sufficiently small such that it can be ignored; (2) the stress components τ_{zx}, τ_{zy}, and σ_z are far less than the other stress

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, Jan. 25, 2000; final revision, Sept. 15, 2002. Associate Editor: V. K. Kinra. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Chair, Department of Mechanics 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.



(a) side view (b) over view

Fig. 1 A plate on a viscoelastic foundation subjected to a circular load

components, therefore, the deformation caused by τ_{zx}, τ_{zy}, and σ_z can be negligible; and (3) the displacement parallel to the horizontal direction of the plate is zero, [6].

Denote the displacement of the plate in the z-direction by $W(x,y,t)$. Based on these assumptions and the fundamental equations of elastodynamics, the governing equation for the deflection of the Kirchhoff plate can be derived by considering the balance of all the forces acting on the element $(x,x+dx;y,y+dy)$. These forces are the impressed force distribution $F(x,y,t)$, the shearing force, the restoring force from the foundation $q(x,y,t)$, and the inertial force $\rho h\partial^2 W/\partial t^2$. The well-known result is

$$D\nabla^2\nabla^2 W(x,y,t)+\rho h\frac{\partial^2}{\partial t^2}W(x,y,t)=F(x,y,t)-q(x,y,t) \tag{1}$$

where the Laplace operator $\nabla^2=\partial^2/\partial x^2+\partial^2/\partial y^2$, $D=Eh^3/[12(1-\mu^2)]$ is stiffness of the plate, h is thickness of the plate, ρ is density of the plate, and E and μ are Young's elastic modulus and Poisson ratio of the plate, respectively.

The most widely used foundation model in rigid pavement design is Winkler foundation, [6,20,21], which assumes the reactive pressure to be proportional to the deflection of the plate, i.e., $q=KW$ where K is the modulus of subgrade reaction. A constant K implies a linear elasticity of the subgrade. When the damping effect of the subgrade is considered, the restoring force becomes $q=KW+C\partial W/\partial t$. This is a viscoelastic foundation consisting of a spring of strength K and a dashpot of strength, C, placed parallel, as shown in Fig. 1. Substitution of the restoring force into Eq. (1) gives

$$D\nabla^2\nabla^2 W(x,y,t)+KW(x,y,t)+C\frac{\partial}{\partial t}W(x,y,t)+\rho h\frac{\partial^2}{\partial t^2}W(x,y,t)=F(x,y,t). \tag{2}$$

3 The Green's Function

According to the mathematical physics theory, the Green's function is a fundamental solution of a partial differential equation, [22,23] For the present problem, the Green's function is defined as the solution of Eq. (1) given that the external excitation $F(x,y,t)$ is characterized by

$$F(\mathbf{x})=\boldsymbol{\delta}(\mathbf{x}-\mathbf{x}_0) \tag{3}$$

in which $\mathbf{x}=(x,y,t)$, $\mathbf{x}_0=(x_0,y_0,t_0)$, $\delta(\mathbf{x}-\mathbf{x}_0)=\delta(x-x_0)\delta(y-y_0)\delta(t-t_0)$, and $\delta(\cdot)$ is the Dirac-delta function, defined by

$$\int_{-\infty}^{\infty}\delta(x-x_0)f(x)dx=f(x_0). \tag{4}$$

Define the three-dimensional Fourier transform and its inversion, [24],

$$\tilde{f}(\boldsymbol{\xi})=\mathbf{F}[f(\mathbf{x})]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\mathbf{x})\exp(-i\boldsymbol{\xi}\mathbf{x})d\mathbf{x} \tag{5a}$$

$$f(\mathbf{x})=\mathbf{F}^{-1}[\tilde{f}(\boldsymbol{\xi})]=(2\pi)^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{f}(\boldsymbol{\xi})\exp(i\boldsymbol{\xi}\mathbf{x})d\boldsymbol{\xi} \tag{5b}$$

where $\boldsymbol{\xi}=(\xi,\eta,\omega)$, $\mathbf{F}[\cdot]$ and $\mathbf{F}^{-1}[\cdot]$ are the Fourier transform and its inversion, respectively. To solve the Green's function, we apply three-dimensional Fourier transform to both sides of Eq. (2)

$$D(\xi^2+\eta^2)^2\tilde{G}(\boldsymbol{\xi};\mathbf{x}_0)+K\tilde{G}(\boldsymbol{\xi};\mathbf{x}_0)+iC\omega\tilde{G}(\boldsymbol{\xi};\mathbf{x}_0)-\rho h\omega^2\tilde{G}(\boldsymbol{\xi};\mathbf{x}_0)=\tilde{F}(\boldsymbol{\xi}) \tag{6}$$

in which $\tilde{\mathbf{F}}(\boldsymbol{\xi})$ is the Fourier transform of $F(\mathbf{x})$, and the displacement response $W(\mathbf{x})$ has been replaced by the symbol $G(\mathbf{x};\mathbf{x}_0)$ to indicate the Green's function. In the derivation of Eq. (6) the following property of Fourier transform is used:

$$\mathbf{F}[f^{(n)}(t)]=(i\omega)^n\mathbf{F}[f(t)]. \tag{7}$$

Since $\tilde{F}(\boldsymbol{\xi})$ is the representation of $F(\mathbf{x})$ in the frequency domain, $\tilde{F}(\boldsymbol{\xi})$ needs to be evaluated as well. This can be achieved by applying three-dimensional Fourier transform on both sides of (3)

$$\tilde{F}(\boldsymbol{\xi})=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\mathbf{x}-\mathbf{x}_0)\exp(-i\boldsymbol{\xi}\mathbf{x})d\mathbf{x}=\exp(-i\boldsymbol{\xi}\mathbf{x}) \tag{8}$$

in which the property of the Dirac-delta function, i.e., Eq. (4), is utilized while evaluating the above integral. Substituting this result (8) into Eq. (6) gives

$$\tilde{G}(\boldsymbol{\xi};\mathbf{x}_0)=\exp(-i\boldsymbol{\xi}\mathbf{x}_0)[D(\xi^2+\eta^2)^2+K+iC\omega-\rho h\omega^2]^{-1}. \tag{9}$$

The Green's function given by (9) is in the frequency domain and needs to be converted back to the time domain. To this end, take the inverse Fourier transform of Eq. (9)

$$G(\mathbf{x};\mathbf{x}_0)=(2\pi)^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp[i\boldsymbol{\xi}(\mathbf{x}-\mathbf{x}_0)][D(\xi^2+\eta^2)^2+K+iC\omega-\rho h\omega^2]^{-1}d\boldsymbol{\xi}. \tag{10}$$

Equation (10) is the Green's function of a plate on the viscoelastic foundation. The Green's function serves as a fundamental solution of a partial differential equation. It can be very useful when dealing with circular loads.

4 The Frequency Response Function

4.1 Integral Representation. We use the Green's function obtained in the previous section to construct the frequency response function (FRF). Denote $W(\mathbf{x})$ as the solution of Eq. (1) in

which the external load is a harmonic circular load with its center located at the origin of the coordinate system, i.e.,

$$F_{\mathrm{FRF}}(\mathbf{x})=(\pi r_0^2)^{-1}H(r_0^2-x^2-y^2)\exp(i\Omega t) \tag{11}$$

in which Ω is frequency of the harmonic load. The steady-state response can be expressed as

$$W(\mathbf{x})=H_{\mathrm{Circle}}(\mathbf{x},\Omega)\exp(i\Omega t). \tag{12}$$

Denote $H_{\mathrm{Circle}}(\mathbf{x},\Omega)$ the frequency response function (FRF) of the plate. Expression (12) simply says that both the response and external excitation possess identical frequency Ω, though response of the plate may have a phase difference with the external excitation reflected in the $H_{\mathrm{Circle}}(\mathbf{x},\Omega)$. The solution of Eqs. (1) and (11) can be constructed by integrating the Green's function over all dimensions, i.e.,

$$W(\mathbf{x})=\int_{-\infty}^{t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(\mathbf{x}_0)G(\mathbf{x};\mathbf{x}_0)d\mathbf{x}_0. \tag{13}$$

Take (10) and (11) into (13) and apply the property of the Dirac-delta function twice.

$$W(r,\Omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty}\frac{J_1(\zeta r_0)J_0(\zeta r)\exp(i\omega t)\exp[i(\omega-\Omega)t_0]}{\pi r_0(D\zeta^4+K+iC\omega-\rho h\omega^2)}\times d\zeta d\omega dt_0 \tag{14}$$

Here, $J_0(\cdot)$ and $J_1(\cdot)$ are the Bessel functions of the first kind, [25]. Since

$$\int_{-\infty}^{\infty}\exp[i(\Omega-\omega)t_0]dt_0=2\pi\delta(\Omega-\omega), \tag{15}$$

substituting (15) into formula (14) gives

$$W(r,\Omega)=\exp(i\Omega t)\int_0^{\infty}\frac{J_1(\zeta r_0)J_0(\zeta r)}{\pi r_0(D\zeta^4+K+iC\Omega-\rho h\Omega^2)}d\zeta d\omega. \tag{16}$$

Comparing (16) to (12) it is straightforward that

$$H_{\mathrm{Circle}}(r,\Omega)=\int_0^{\infty}\frac{J_1(\zeta r_0)J_0(\zeta r)}{\pi r_0(D\zeta^4+K+iC\Omega-\rho h\Omega^2)}d\zeta. \tag{17}$$

The frequency response function of the plate to a concentrated harmonic load $F_{\mathrm{Point}}(\mathbf{x})=\delta(x)\delta(y)\exp(i\Omega t)$ can be obtained by simply taking the limit $r_0\to 0$ on both sides of (17), i.e.,

$$H_{\mathrm{Point}}(r,\Omega)=\frac{1}{2\pi}\int_0^{\infty}\frac{J_0(\zeta r)}{D\zeta^4+K+iC\Omega-\rho h\Omega^2}\zeta d\zeta. \tag{18}$$

Now we have obtained the frequency response function $H_{\mathrm{Circle}}(\mathbf{x},\Omega)$ and $H_{\mathrm{Point}}(r,\Omega)$ in the rectangular and cylindrical coordinate systems, respectively. In general, the frequency response function given by (17) and (18) are complex functions. The Bessel function of the first kind $J_0(z)$ can be given in integration representation, [25],

$$J_0(z)=\frac{i}{\pi}\int_0^{\infty}[\exp(-iz\cosh\gamma)-\exp(iz\cosh\gamma)]d\gamma. \tag{19}$$

Realizing the property of even function, we can also rewrite identity (19) as

$$J_0(z)=\frac{i}{\pi}\int_{-\infty}^{\infty}\exp(-iz\cosh\gamma)d\gamma. \tag{20}$$

Substituting this expression into Eq. (18) gives

$$H_{\mathrm{Circle}}(r,\Omega)=\frac{i}{\pi^2 r_0 D}\int_0^{\infty}\int_{-\infty}^{\infty}\frac{J_1(\zeta r_0)\exp(-i\zeta r\cosh\gamma)}{\zeta^4+(K-\rho h\Omega^2+iC\Omega)/D}d\zeta d\gamma. \tag{21}$$

4.2 Roots of the Characteristic Equation. Before the integration (21) can be further evaluated, it is necessary to investigate the roots of the characteristic equation of type

$$\zeta^4+(K-\rho h\Omega^2+iC\Omega)/D=0. \tag{22}$$

Characteristic Eq. (22) is a fourth-order algebraic equation, roots of which are dependent upon parameters related to plate structure, foundation, and loading condition. We separate our discussion into two categories: no damping effect (Winkler foundation) and with damping effect (viscoelastic foundation). Within each individual category three cases are separately addressed because the relationship between load frequency and eigenfrequencies may result in different scenarios.

4.2.1 No damping ($C=0$). This case corresponds to a plate on a Winkler foundation. Define equivalent stiffness as $\bar{K}=|(K-\rho h\Omega^2)/D|$ and resonance frequency as $\Omega_0=\sqrt{K/\rho h}$.

a. $\Omega<\Omega_0$. Equation (22) becomes $\zeta^4+\bar{K}=0$. All the four roots of this equation possess complex values and can be given by $\zeta_j=\sqrt[4]{\bar{K}}\exp[i(1+2j)\pi/4]$ with $j=0,1,2,3$.
b. $\Omega=\Omega_0$. Equation (22) becomes $\zeta^4=0$ and all four roots degrade and becomes $\zeta=0$.
c. $\Omega>\Omega_0$. Equation (22) becomes $\zeta^4-\bar{K}=0$. Two of the four roots are imaginary and the other two are real valued, which are given by $\zeta_j=\sqrt[4]{\bar{K}}\exp[i(j\pi/2)]$ with $j=0,1,2,3$, respectively.

4.2.2 With Damping ($C\neq 0$). Define the equivalent damping coefficient $\bar{C}=C\Omega/D$. Three cases are discussed as follows:

a. $\Omega<\Omega_0$. Equation (22) becomes $\zeta^4+\bar{K}+i\bar{C}=0$. All four roots possess complex values and can be given by $\zeta_j=\sqrt[8]{\bar{K}^2+\bar{C}^2}\exp[i(\vartheta+\pi+2j\pi)/4]$ with $j=0,1,2,3$ in which $\tan\vartheta=\bar{C}/\bar{K}>0$.
b. $\Omega=\Omega_0$. Equation (22) becomes $\zeta^4+i\bar{C}=0$. In this case all four roots possess complex values and can be given by $\zeta_j=\sqrt[4]{\bar{C}}\exp[i(3\pi+4j\pi)/8]$ with $j=0,1,2,3$, respectively.
c. $\Omega>\Omega_0$. Equation (22) becomes $\zeta^4-\bar{K}+i\bar{C}=0$. All four roots possess complex values and can be given by $\zeta_j=\sqrt[8]{\bar{K}^2+\bar{C}^2}\exp[i(\vartheta+2j\pi)/4]$ with $j=0,1,2,3$ in which $\tan\vartheta=-\bar{C}/\bar{K}<0$.

4.3 Closed-Form Representation. According to the residue theorem (Saff and Snider, [26]) the residues of the integrand of (32) in the upper half-plane contribute to the following integration:

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{J_1(\zeta r_0)\exp(-i\zeta r\cosh\gamma)}{\zeta^4\pm\bar{K}+i\bar{C}}d\zeta&=2\pi i\sum_{j=1}^{J}\mathrm{Res}\left[\frac{J_1(\zeta r_0)\exp(-i\zeta r\cosh\gamma)}{\zeta^4\pm\bar{K}+i\bar{C}}\right]\Bigg|_{\zeta=\zeta_j}\\&=\frac{\pi i}{2}\sum_{j=1}^{J}\frac{J_1(\zeta r_0)\exp(-i\zeta r\cosh\gamma)}{\zeta^3}\Bigg|_{\zeta=\zeta_j,\,\mathrm{Im}(\zeta)>0}\end{aligned} \tag{23}$$

in which J is the number of poles whose imaginary parts are positive, and ζ_j represents the poles of the integrand of Eq. (23).

Based on the previous analysis, it is clear that two complex

roots exist in the upper half-plane of the complex ζ-plane for all the cases of $C\neq 0$ and the case $\Omega<\Omega_0$ of $C=0$. For $\Omega>\Omega_0$ and $C=0$ two poles are located on the real axis, while for $\Omega=\Omega_0$ and $\Omega=\Omega_0$ only one pole is located on the real axis. Since the residue theorem cannot be directly applied in the sense of Riemann integral, the concept of Cauchy principal value (p.v.) of the integration (23) has to be introduced for these two cases, [26]. In Fig. 2(*a*) and (*b*) two integral contours are respectively provided for $\Omega=\Omega_0$ of $C=0$ and $\Omega>\Omega_0$ of $C=0$.

As shown in Fig. 2(*a*), the contour consists of three portions for the case $\Omega=\Omega_0$ and $C=0$. The integral of the left-hand side of Eq. (23) now becomes $\int_{-\infty}^{\infty} J_1(\zeta r_0)\exp(-i\zeta r\cos\gamma)/\zeta^4\, d\zeta$. Since no pole is embraced by the closed contour, the theorem of residue says that the integral along this closed contour becomes zero, i.e.,

$$\oint = \text{p.v.}\int_{-R}^{R} + \int_{C_1} + \int_{C_2} = 0 \tag{24}$$

in which the abbreviation p.v. means the Cauchy principal value of the integration. For those ζ values on C_1, they can be expressed by $\zeta=\varepsilon\exp(i\beta)$ and $d\zeta=i\varepsilon\exp(i\beta)d\beta$. The integration $\int_{C_1}$ as $\varepsilon\to 0$ is then given by

$$\lim_{\varepsilon\to 0}\int_{\pi}^{0}\frac{J_1[\varepsilon\exp(i\beta)r_0]\exp[-i\varepsilon r\exp(i\beta)\cos\gamma]}{\varepsilon^4\exp(i4\beta)}i\varepsilon\exp(i\beta)d\beta = \lim_{\varepsilon\to 0} i\int_{\pi}^{0}\frac{J_1[\varepsilon\exp(i\beta)r_0]\exp[-i\varepsilon r\exp(i\beta)\cos\gamma]}{\varepsilon^3\exp(i3\beta)}d\beta. \tag{25}$$

By applying the Maclaurin expansion and the L'Hospital rule to the limit (25), it is found that a singularity with an order $O(\varepsilon^{-1})$ exists. For those ζ values on C_2, they can be expressed by $\zeta=R\exp(i\beta)$ and $d\zeta=iR\exp(i\beta)d\beta$. The integration $\int_{C_2}$ as $R\to\infty$ is then given by

$$\lim_{R\to\infty}\int_{\pi}^{0}\frac{J_1[R\exp(i\beta)r_0]\exp[-iRr\exp(i\beta)\cos\gamma]}{R^4\exp(i4\beta)}iR\exp(i\beta)d\beta = \lim_{R\to\infty} i\int_{\pi}^{0}\frac{J_1[R\exp(i\beta)r_0]\exp[-iRr\exp(i\beta)\cos\gamma]}{R^3\exp(i3\beta)}d\beta = 0. \tag{26}$$

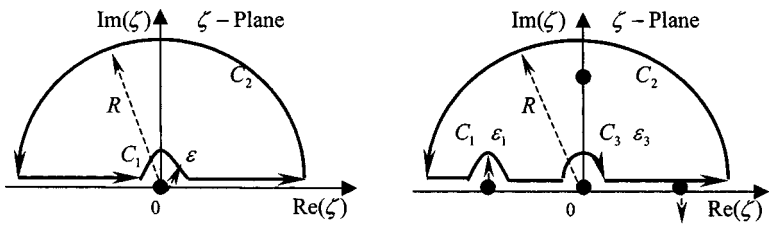
Comparison among (24), (25), and (26) shows

$$\text{p.v.}\int_{-\infty}^{\infty} = \lim_{R\to\infty}\int_{-R}^{R} = -\lim_{\varepsilon\to 0}\int_{C_1} \sim O(\varepsilon^{-1}). \tag{27}$$

For the case $\Omega>\Omega_0$ and $C=0$, as shown in Fig. 2(*b*), the contour consists of four portions. One pole is within the range of the closed contour. Suppose that a tiny amount of damping is present, which is the case in reality, the pole that is currently exactly located on the positive part of the real axis will be actually pushed down into the fourth quadrant of the complex plane. Hence, its contribution to the integral vanishes. So we now have

$$\begin{aligned}\oint &= \text{p.v.}\int_{-R}^{R} + \int_{C_1} + \int_{C_2} + \int_{C_3} \\ &= 2\pi i\,\text{Res}\left[\frac{J_1(\zeta r_0)\exp(-i\zeta r\cos\gamma)}{\zeta^4-\overline{K}}\right]\Bigg|_{\zeta=i\sqrt[4]{\overline{K}}} \\ &= \frac{\pi i J_1(\zeta r_0)\exp(-i\zeta r\cos\gamma)}{2\zeta^3}\Bigg|_{\zeta=i\sqrt[4]{\overline{K}}}.\end{aligned} \tag{28}$$

For those ζ values on C_2 and C_3, one can prove $\lim_{R\to 0}\int_{C_2}=0$ and $\lim_{\varepsilon_3\to 0}\int_{C_3}=0$ in the same manner as in the previous discussion. For those ζ values on C_1, they can be expressed by $\zeta=-\sqrt[4]{K}+\varepsilon_1\exp(i\beta)$ and $d\zeta=i\varepsilon_1\exp(i\beta)d\beta$. The integration $\int_{C_1}$ as $\varepsilon_1\to 0$ is then given by



(a) $\Omega=\Omega_0$ and $C=0$ (b) $\Omega>\Omega_0$ and $C=0$

Fig. 2 The integral contours for evaluating the Cauchy principal value in Eq. (23)

$$\lim_{\varepsilon_1\to 0}\int_{\pi}^{0}\frac{J_1\{r_0[-\sqrt[4]{\overline{K}}+\varepsilon_1\exp(i\beta)]\}\exp\{-ir[-\sqrt[4]{\overline{K}}+\varepsilon_1\exp(i\beta)]\cosh\gamma\}}{[-\sqrt[4]{\overline{K}}+\varepsilon_1\exp(i\beta)]^4-\overline{K}}i\varepsilon_1\exp(i\beta)d\beta$$
$$=-\pi i\mathrm{Res}\left[\frac{J_1(\zeta r_0)\exp(-i\zeta r\cosh\gamma)}{\zeta^4-\overline{K}}\right]\Bigg|_{\zeta=-\sqrt[4]{\overline{K}}}=\frac{-\pi iJ_1(\zeta r_0)\exp(-i\zeta r\cosh\gamma)}{4\zeta^3}\Bigg|_{\zeta=-\sqrt[4]{\overline{K}}}. \tag{29}$$

Comparison among (28) and (29) gives

$$\text{p.v.}\int_{-\infty}^{\infty}=\lim_{R\to\infty}\int_{-R}^{R}=\frac{\pi iJ_1(\zeta r_0)\exp(-i\zeta r\cosh\gamma)}{2\zeta^3}\Bigg|_{\zeta=i\sqrt[4]{\overline{K}}}$$
$$+\frac{\pi iJ_1(\zeta r_0)\exp(-i\zeta r\cosh\gamma)}{4\zeta^3}\Bigg|_{\zeta=-\sqrt[4]{\overline{K}}}. \tag{30}$$

So far, all cases have been analyzed and the left-hand side of Eq. (23) has been represented in closed-form expressions for different cases. With the help of these closed-form expressions, the frequency response function given by (21) is now ready to be further evaluated. Realize that an integral representation of Hankel function is, [25],

$$\frac{2i}{\pi}\int_0^{\infty}\exp(-izrch\gamma)d\gamma=H_0^{(2)}(zr) \tag{31}$$

where $H_0^{(2)}(\cdot)$ is the Hankel function of the second kind (i.e., the Bessel function of the third kind). Applying this expression and the closed-form representation of the inner integral of Eq. (21), we are eventually able to write the integration (21) in a closed-form expression, as illustrated by (32):

$$H_{\text{Circle}}(r,\Omega)=\frac{i}{4r_0D}\sum_{n=a,b}\frac{J_1(\zeta_n r_0)H_0^{(2)}(\zeta_n r)}{\zeta_n^3} \tag{32}$$

where poles ζ_a and ζ_b are provided in Table 1.

5 Verification Through Special Cases

It is of interest to examine the static solution through applying the results (17) and (18). The derivation of static solution can also be of great value in terms of verifying if the general result given by Eq. (32) is correct. For a static load, $F_{\text{Circle\_sta}}(\mathbf{x})=(\pi r_0^2)^{-1}H(r_0^2-x^2-y^2)$ and $F_{\text{Point\_sta}}(\mathbf{x})=\delta(x)\delta(y)$. The static solution of Eq. (1) corresponding to the static load can be achieved by letting $\Omega=0$ in (17) and (18)

$$H_{\text{Circle\_sta}}(r)=\int_0^{\infty}\frac{J_1(\zeta r_0)J_0(\zeta r)}{\pi r_0(D\zeta^4+K)}d\zeta \tag{33}$$

$$H_{\text{Point\_sta}}(r)=\frac{1}{2\pi}\int_0^{\infty}\frac{J_0(r\zeta)}{D\zeta^4+K}\zeta d\zeta \tag{34}$$

This expression (34) is exactly identical to a known result, [27–29].

If damping is ignored in Eqs. (17) and (18), the frequency response functions become

$$H_{\text{Circle}}(r,\Omega)=\int_0^{\infty}\frac{J_1(\zeta r_0)J_0(\zeta r)}{\pi r_0(D\zeta^4+K-\rho h\Omega^2)}d\zeta \tag{35}$$

$$H_{\text{Point}}(r,\Omega)=\frac{1}{2\pi}\int_0^{\infty}\frac{J_0(\zeta r)}{D\zeta^4+K-\rho h\Omega^2}\zeta d\zeta. \tag{36}$$

Clearly, $H_{\text{Circle}}(r,\Omega)$ and $H_{\text{Point}}(r,\Omega)$ given by (35) and (36) are real functions. It implies that no phase difference exists between the response and the external excitation.

If the radius of the circular load approaches zero, the frequency response functions given by (32) becomes

$$H_{\text{Point}}(r,\Omega)=\frac{i}{8D}\sum_{n=a,b}\frac{H_0^{(2)}(\zeta_n r)}{\zeta_n^2} \tag{37}$$

where the following limit is used:

$$\lim_{r_0\to 0}\frac{J_1(\zeta_n r_0)}{\zeta_n r_0}=1/2. \tag{38}$$

In practice, the vibratory devices used for pavement nondestructive test generate harmonic loads with frequency 5~60 Hz, [3]. These frequencies usually fall into the low frequency range of $\Omega<\Omega_0$. Under such condition, it is appropriate to write (37) as

$$H_{\text{Point}}(r,\Omega)=(64\overline{K}D)^{-1/2}\{H_0^{(2)}[\sqrt[4]{\overline{K}}\exp(i\pi/4)r]$$
$$+H_0^{(2)}[\sqrt[4]{\overline{K}}\exp(i3\pi/4)r]\}. \tag{39}$$

6 Conclusion

In this paper we derived a closed-form solution of dynamic response of a Kirchhoff plate on a viscoelastic foundation subjected to impulse and harmonic circular loads. The solution utilizes the Bessel and Hankel functions. The result has been partially verified by comparing the static solution of a point source obtained in this paper to a well-known result. This analytical expression permits one to construct fast algorithms for parameter identification in pavement nondestructive test.

Table 1 Poles that contribute to the harmonic response of the plate

Damping	Frequency	Poles ζ_a and ζ_b
$C=0$	$\Omega<\Omega_0$	$\zeta_a=\overline{K}^{1/4}e^{i\pi/4}$ and $\zeta_b=\overline{K}^{1/4}e^{i3\pi/4}$
$C=0$	$\Omega=\Omega_0$	$O(\varepsilon^{-1})$
$C=0$	$\Omega>\Omega_0$	$\zeta_a=i\overline{K}^{1/4}$ and $\zeta_b=-\overline{K}^{1/4}$
$C\neq 0$	$\Omega>\Omega_0$	$\zeta_a=(\overline{K}^2+\overline{C}^2)^{1/8}e^{i(\vartheta+\pi)/4}$, $\zeta_b=(\overline{K}^2+\overline{C}^2)^{1/8}e^{i(\vartheta+3\pi)/4}$ and $\tan\vartheta=\overline{C}/\overline{K}$
$C\neq 0$	$\Omega=\Omega_0$	$\zeta_a=\overline{C}^{1/4}e^{i3\pi/8}$ and $\zeta_b=\overline{C}^{1/4}e^{i7\pi/8}$
$C\neq 0$	$\Omega>\Omega_0$	$\zeta_a=(\overline{K}^2+\overline{C}^2)^{1/8}e^{i(\vartheta+2\pi)/4}$, $\zeta_b=(\overline{K}^2+\overline{C}^2)^{1/8}e^{i(\vartheta+4\pi)/4}$ and $\tan\vartheta=-\overline{C}/\overline{K}$

Acknowledgments

The author is grateful to Prof. Vikram Kinra, the Associate Editor, for his constructive comments and suggestions, which enhance the content and the presentation of the original manuscript.

References

[1] Bush, A. J., 1980, ''Nondestructive Testing for Light Aircraft Pavements, Phase II. Development of the Nondestructive Evaluation Methodology,'' Report NO. FAA RD-80-9, Final Report, Federal Aviation Administration.
[2] Uzan, J., and Lytton, R., 1990, ''Analysis of Pressure Distribution Under Falling Weight Deflectometer Loading,'' J. Transport. Eng., ASCE, **116**, No. 2.
[3] Haas, R., Hudson, W. R., and Zaniewski, J., 1994, *Modern Pavement Management*, Krieger, Malabar, FL.
[4] Hudson, W. R., Haas, R., and Uddin, W., 1997, *Infrastructure Management: Integrating Design, Construction, Maintenance, Rehabilitation, and Renovation*, McGraw-Hill, New York.
[5] Westergaard, H. M. S., 1926, ''Stresses in Concrete Pavements Computed by Theoretical Analysis,'' Public Roads, **7**(2), Apr.
[6] Yoder, E. J., and Witczak, M. W., 1975, *Principles of Pavement Design*, John Wiley and Sons, New York.
[7] Scullion, T., Uzan, J., and Paredes, M., 1990, ''MODULUS: A Microcomputer-Based Backcalculation System,'' Transp. Res. Rec., 1260.
[8] Taheri, M. R., 1986, ''Dynamic Response of Plates to Moving Loads,'' Ph.D. thesis, Purdue University, West Lafayette, IN.
[9] Kukreti, A. R., Taheri, M., and Ledesma, R. H., 1992, ''Dynamic Analysis of Rigid Airport Pavements With Discontinuities,'' J. Transport. Eng., ASCE, **118**(3), pp. 341–360.
[10] Zaghloul, S. M., White, T. D., Drnevich, V. P., and Coree, B., 1994, ''Dynamic Analysis of FWD Loading and Pavement Response Using a Three Dimensional Dynamic Finite Element Program,'' Transportation Resource Board, Washington, D.C.
[11] Achenbach, J. D., Keshava, S. P., and Herrman, G., 1966, ''Waves in a Smoothly Jointed Plate and Half Space,'' J. Eng. Mech., **92**(2), pp. 113–129.
[12] Freund, L. B., and Achenbach, J. D., 1968, ''Waves in a Semi-Infinite Plate in Smooth Contact With a Harmonically Distributed Half Space,'' Int. J. Solids Struct., **4**, pp. 605–621.
[13] Oien, M. A., 1973, ''Steady Motion of a Plate on an Elastic Half Space,'' ASME J. Appl. Mech., **40**(2), pp. 478–484.
[14] Arnold, R. N., Bycroft, G. N., and Warburton, G. B., 1955, ''Force Vibrations of a Body on an Infinite Elastic Solid,'' ASME J. Appl. Mech., **77**, pp. 391–400.
[15] Warburton, G. B., 1957, ''Forced Vibration of a Body on an Elastic Stratum,'' ASME J. Appl. Mech., **79**, pp. 55–57.
[16] Bycroft, G. N., 1956, ''Force Vibrations of a Rigid Circular Plate on a Semi-Infinite Elastic Space and on an Elastic Stratum,'' Philos. Trans. R. Soc. London, Ser. A, **248**, pp. 327–368.
[17] Krenk, S., and Schmidt, H., 1981, ''Vibration of an Elastic Circular Plate on an Elastic Half-Space—A Direct Approach,'' ASME J. Appl. Mech., **48**, pp. 161–168.
[18] Sun, L., 2001, ''A Closed-Form Solution of Bernoulli-Euler Beam on Viscoelastic Foundation Under Harmonic Line Loads,'' J. Sound Vib., **242**(4), pp. 619–627.
[19] Sun, L., 2001, ''Closed-Form Representation of Beam Response to Moving Line Loads,'' ASME J. Appl. Mech., **68**, pp. 348–350.
[20] Sun, L., and Deng, X., 1997, ''Random Response of Beam Under a Moving Random Load in the Line Source Form,'' Acta Mech. Sin., **29**(3), pp. 365–368.
[21] Kenney, J. T., 1954, ''Steady-State Vibrations of Beam on Elastic Foundation for Moving Load,'' ASME J. Appl. Mech., **21**, p. 359.
[22] Sun, L., and Greenberg, B., 2000, ''Dynamic Response of Linear Systems to Moving Stochastic Sources,'' J. Sound Vib., **229**(**4**), pp. 957–972.
[23] Morse, P. M., and Feshbach, H., 1953, *Methods of Theoretical Physics: Part I and II*, McGraw-Hill, New York.
[24] Eringen, A. C., and Suhubi, E. S., 1975, *Elastodynamics, Vol. I and II*, Academic Press, New York.
[25] Watson, G. N., 1966, *A Treatise on the Theory of Bessel functions*, 2nd Ed., Cambridge University Press, London.
[26] Saff, E. B., and Snider, A. D., 1993, *Fundamentals of Complex Analysis for Mathematics, Science, and Engineering*, 2nd Ed., Prentice-Hall, New York.
[27] Zhu, Z., Wang, B., and Guo, D., 1985, *Pavement Mechanics*, People's Transport Publishing, Beijing, China.
[28] Timoshenko, S., and Woinowsky-Krieger, S., 1968, *Theory of Plates and Shells*, 2nd Ed., McGraw-Hill, New York.
[29] Ugural, A. C., 1981, *Stresses in Plates and Shells*, McGraw-Hill, New York.

T. Y. Ng
Institute of High Performance Computing,
National University of Singapore,
1 Science Park Road,
01-01 The Capricorn,
Singapore Science Park II,
Singapore 117528
and
School of Mechanical and Production Engineering,
Nanyang Technological University,
50 Nanyang Avenue,
Singapore 639798

H. Li

K. Y. Lam

C. F. Chua

Institute of High Performance Computing,
National University of Singapore,
1 Science Park Road,
01-01 The Capricorn,
Singapore Science Park II,
Singapore 117528

Frequency Analysis of Rotating Conical Panels: A Generalized Differential Quadrature Approach

Based on the generalized differential quadrature (GDQ) method, this paper presents, for the first instance, the free-vibration behavior of a rotating thin truncated open conical shell panel. The present governing equations of free vibration include the effects of initial hoop tension and the centrifugal and Coriolis accelerations due to rotation. Frequency characteristics are obtained to study in detail the influence of panel parameters and boundary conditions on the frequency characteristics. Further, qualitative differences between the vibration characteristics of rotating conical panels and that of rotating full conical shells are investigated. To ensure the accuracy of the present results using the GDQ method, comparisons and verifications are made for the special case of a stationary panel. [DOI: 10.1115/1.1577600]

1 Introduction

A comprehensive literature search will reveal that the amount of research work conducted with regard to the vibration of rotating/nonrotating cylindrical shells and panels are indeed extensive. Comparatively, the number of research articles available on the vibration of either rotating or stationary full-circular conical shell as well as stationary open conical shell panels remains few. Further, it is noted here that there is presently no technical paper available in the open literature on the vibration analysis of rotating open conical shell panels. This paper aims to address the void of information on this problem.

By employing the generalized differential quadrature (GDQ) method, [1], this paper solves the free vibration problem of general rotating thin truncated open conical shell panels. The governing partial differential equations of motion derived include the effects of initial hoop tension as well as the centrifugal and Coriolis accelerations. Employing the GDQ method and imposing the boundary conditions, these equations are transformed to a set of numerical eigenvalue equations, which are then solved for the natural frequencies.

2 Theoretical Formulation

Figure 1 shows an isotropic truncated open thin conical shell panel rotating about its rotational axis of symmetry at a constant angular velocity Ω. The half vertex angle is denoted by α, the subtended angle by β, the thickness by h and the slant length of the shell by L. The symbols a and b are the mean radii for the smaller and larger ends of the shell, respectively. The middle surface of the shell is taken as the reference surface for our ground-based orthogonal coordinate system (x,θ,z). The components of displacement in the meridional x, circumferential θ, and normal z directions, are given by u, v, and w, respectively.

The governing equations of motion in terms of forces, moments and displacements for the resultants of free vibration of a rotating conical shell can be written as follows (see Lam and Li [2]):

$$N_{x,x}+r^{-1}N_{x\theta,\theta}+r^{-2}N_\theta^0(u_{,\theta\theta}-rw_{,x}\cos\alpha)+r^{-1}(N_x-N_\theta)\sin\alpha+2\rho h\Omega v_{,t}\sin\alpha-\rho hu_{,tt}=0 \tag{1}$$

$$N_{x\theta,x}+r^{-1}N_{\theta,\theta}+r^{-1}M_{x\theta,x}\cos\alpha+r^{-2}M_{\theta,\theta}\cos\alpha+r^{-2}N_\theta^0(ru_{,x\theta}+u_{,\theta}\sin\alpha+rv_{,x}\sin\alpha)+2r^{-1}N_{x\theta}\sin\alpha-2\rho h\Omega(u_{,t}\sin\alpha+w_{,t}\cos\alpha)-\rho hv_{,tt}=0 \tag{2}$$

$$M_{x,xx}+2r^{-1}M_{x\theta,\theta x}+r^{-2}M_{\theta,\theta\theta}+2r^{-1}M_{x,x}\sin\alpha-r^{-1}M_{0,x}\sin\alpha+r^{-2}N_\theta^0(w_{,\theta\theta}-ru_{,x}\cos\alpha)+r^{-2}N_\theta^0(w\cos^2\alpha+u\sin\alpha\cos\alpha)-r^{-1}N_\theta\cos\alpha+2\rho h\Omega v_{,t}\cos\alpha-\rho hw_{,tt}=0 \tag{3}$$

where the subscript variable after a comma indicates partial differentiation with respect to that variable, and

$$\rho=\rho(x,\theta)=h^{-1}\int_{-h/2}^{h/2}\rho^*(x,\theta,z)dz \tag{4}$$

$$N_\theta^0=\rho h\Omega^2r^2=\rho h\Omega^2(a+x\sin\alpha)^2,\quad r=r(x)=(a+x\sin\alpha) \tag{5}$$

with $\rho^*(x,\theta,z)$ being the mass density of the conical shell, and $\rho(x,\theta)$ the average density in the normal z direction. N_θ^0 is defined as the initial hoop tension due to the centrifugal force effect. The three terms, $u_{,tt}$, $v_{,tt}$, and $w_{,tt}$, on the left-hand side of the Eqs. (1) to (3) are the relative accelerations. The four terms, $2\Omega v_{,t}\sin\alpha$, $2\Omega u_{,t}\sin\alpha$, $2\Omega w_{,t}\cos\alpha$, and $2\Omega v_{,t}\cos\alpha$, are the Coriolis accelerations. The implicit terms related to N_θ^0, $u\Omega^2$, $v\Omega^2$, and $w\Omega^2$, are the centrifugal accelerations. $\mathbf{N}^T=\{N_x,N_\theta,N_{x\theta}\}$ and $\mathbf{M}^T=\{M_x,M_\theta,M_{x\theta}\}$ are the internal force

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Dec. 2, 2001; final revision, Nov. 26, 2002. Associate Editor: O. O' Reilly. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, 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.

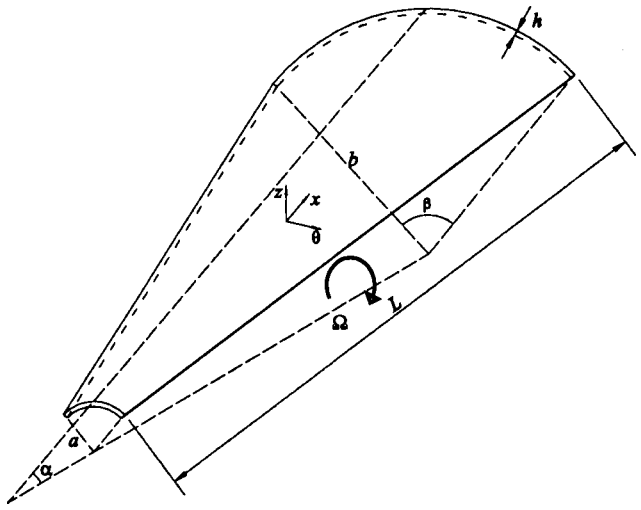


Fig. 1 Geometry of a thin rotating conical panel

and moment vectors which can be expressed by a linear-elastic constitutive relationship, [1]. The strains and curvatures follow those of Love thin shell theory. Substituting the constitutive relations into the governing Eqs. (1) to (3), a set of partial differential governing equations with variable coefficients expressed by the displacements is derived and written in the matrix form as

$$\mathbf{LU}=\mathbf{0} \tag{6}$$

where $\mathbf{U}^T=\{u(x,\theta,t),v(x,\theta,t),w(x,\theta,t)\}$ is the displacement vector. $\mathbf{L}=[L_{ij}]$ $(i,j=1,2,3)$ is a 3×3 differential operator matrix of **U**, see [1] for details.

The governing Eq. (6) can be used for a rotating conical shell with arbitrary boundary conditions. In this paper, the simply supported boundary condition $(u=0,w=0,N_\theta=0,M_\theta=0)$ is applied on both the straight edges (at $\theta=0$, β). For both the curve edges (at $x=0$, L), four boundary condition types are considered, which are namely: clamped $(u=0,v=0,w=0,w_{,x}=0)$ at both edges (Cs-Cl); simply-supported $(v=0,w=0,N_x=0,M_x=0)$ at both edges (Ss-Sl); simply supported at small edge and clamped at large edge (Ss-Cl); and clamped at small edge and simply supported at large edge (Cs-Sl).

3 Generalized Differential Quadrature (GDQ) Implementation

Unlike traditional numerical techniques such as the finite element method, which may require a large number of nodes or elements for accurate results at points of interest, the generalized differential quadrature (GDQ) method is a global numerical approximate technique, [1], requiring only a sparse grid point distribution to achieve similar accuracy. The basic concept is that the derivative of a sufficiently smooth function with respect to a coordinate direction at a discrete point can be approximated by a weighted linear sum of the functional values at all the discrete points in that direction. Details of the GDQ implementation can be found in [1].

In this paper, a cosine distribution of discrete grid points in the meridional x direction is used. Also, in satisfying the simply supported boundary conditions on the straight edges $(\theta=0,\beta)$, the displacement field can be written as follows:

$$\mathbf{U}=\begin{Bmatrix} u(x,\theta,t)\\ v(x,\theta,t)\\ w(x,\theta,t)\end{Bmatrix}=\begin{Bmatrix} U(x)\sin(\lambda\theta)e^{(i(\lambda\theta+\omega t))}\\ V(x)\cos(\lambda\theta)e^{(i(\lambda\theta+\omega t))}\\ W(x)\sin(\lambda\theta)e^{(i(\lambda\theta+\omega t))}\end{Bmatrix} \tag{7}$$

where $\lambda=(n\pi/\beta)$, ω(rad/s) is the natural circular frequency of the rotating conical panels, and n is an integer representing the circumferential wave number of the panel. Substituting Eq. (7) into the set of partial differential governing Eqs. (6) in temporal-spatial domain, a set of ordinary differential equations with spatial variable coefficients in the meridional x direction is derived as

$$\mathbf{L}^*\mathbf{U}^*=\mathbf{0} \tag{8}$$

where $\mathbf{U}^{*T}=\{U(x),V(x),W(x)\}$ is the unknown modal spatial function vector describing the distribution of vibrational amplitude in the x direction. $\mathbf{L}^*=[L_{ij}^*]$ $(i,j=1,2,3)$ is a 3×3 differential operator matrix. By the GDQ procedure, the approximate governing equations in the discrete form is

$$\mathbf{L}^*\mathbf{U}^*|_{x=x_i}=\mathbf{R}_{3\times11}\mathbf{U}^{**}_{11\times1}|_{x=x_i}=\mathbf{0} \quad (i=1,2,3,\ldots N) \tag{9}$$

where N is the number of total discrete grid points including the points at both edges in the meridional x direction. **R** is the 3×11 complex-coefficient matrix that is function of discrete grid point $x=x_i$. $\mathbf{U}^{**}$ is an 11-order column vector consisting of the eigenmodes in x.

Table 1 Comparison of frequency parameter, $f=\omega b\sqrt{\rho h/A_{11}}$, with results generated from MSC/NASTRAN for a conical panel. Numbers in parenthesis denote discrepancies against finite element method results.

$\alpha=20°$, $\beta=60°$, $h=0.015a$, $L=20a$, $\nu=0.3$ and axial wave number $m=1$							
Boundary Conditions	Methodology	Circumferential wave number, n					
		2	3	4	5	6	7
C_S-C_L	GDQ	0.0894(-0.17%)	0.1096(-0.27%)	0.1460(-0.46%)	0.1948(-0.75%)	0.2554(-1.14%)	0.3277(-1.59%)
	NASTRAN	0.0892	0.1093	0.1453	0.1933	0.2525	0.3226
S_S-C_L	GDQ	0.0894(-0.17%)	0.1096(-0.27%)	0.1460(-0.46%)	0.1948(-0.75%)	0.2554(-1.14%)	0.3277(-1.59%)
	NASTRAN	0.0892	0.1093	0.1453	0.1933	0.2525	0.3226
C_S-S_L	GDQ	0.0638(2.65%)	0.0909(0.83%)	0.1299(0.16%)	0.1801(-0.24%)	0.2419(-0.76%)	0.3147(-1.22%)
	NASTRAN	0.0655	0.0917	0.1301	0.1797	0.2401	0.3109
S_S-S_L	GDQ	0.0638(2.65%)	0.0909(0.83%)	0.1299(0.17%)	0.1801(-0.24%)	0.2419(-0.76%)	0.3147(-1.22%)
	NASTRAN	0.0655	0.0917	0.1301	0.1797	0.2401	0.3109
$\beta=60°$, $h=0.02a$, $L=25a$, $\nu=0.3$, and mode $(m,n)=(1,2)$							
Boundary Conditions	Methodology	Vertex angle, α					
		10°		20°	30°	40°	
C_S-C_L	GDQ	0.0850(-0.36%)		0.0937(-0.25%)	0.1104(-0.15%)	0.1240(3.87%)	
	NASTRAN	0.0847		0.0934	0.1102	0.1290	
$\alpha=30°$, $h=0.02a$, $L=20a$, $\nu=0.3$, and mode $(m,n)=(1,4)$							
Boundary Conditions	Methodology	Subtended angle, β					
		15°		30°	45°	60°	
C_S-C_L	GDQ	1.4619(-3.42%)		0.4191(-1.65%)	0.2273(-0.91%)	0.1634(-0.41%)	
	NASTRAN	1.4136		0.4123	0.2253	0.1627	

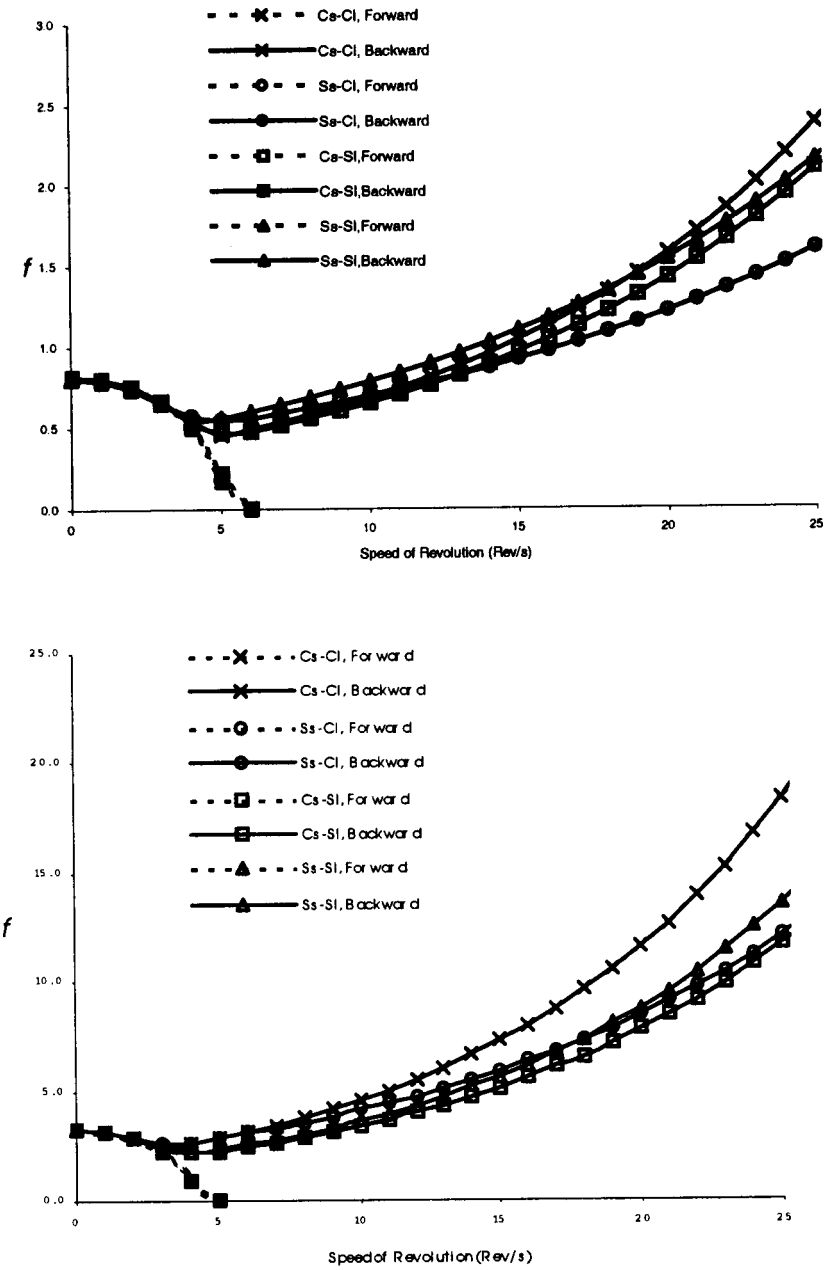


Fig. 2 Variation of frequency parameter $f=\omega b\sqrt{\rho h/A_{11}}$ for a rotating conical panel with different boundary conditions against revolution speed (rps), for different n [$n=1$ (top) and $n=5$ (bottom)], with panel parameters $\alpha=30°$, $\beta=15°$, $h/a=0.015$, $L/a=5$, and $m=1$

By imposing Eq. (9) on every discrete grid point $x=x_i$ ($i=1,2,\ldots N$), and then rewriting resulting equation in terms of yields an equation in the following matrix form:

$$[\omega^2\mathbf{H}_1+\omega\mathbf{H}_2+\mathbf{H}_3]\mathbf{d}=\mathbf{0} \tag{10}$$

where $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$ are the $N^*\times N^*$ numerical complex-coefficient system matrices ($N^*=3\times N-8$), and $\mathbf{d}$ is the N^*-order eigenmode column vector. Equation (10) is a nonstandard eigenvalue equation. For a given frequency, it can be transformed into a standard form of eigenvalue equation as

$$\left(\begin{bmatrix}\mathbf{0} & \mathbf{I}\\ -\mathbf{H}_3 & -\mathbf{H}_2\end{bmatrix}-\omega\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{H}_1\end{bmatrix}\right)\begin{Bmatrix}\mathbf{d}\\ \omega\mathbf{d}\end{Bmatrix}=\mathbf{0} \tag{11}$$

where $\mathbf{I}$ is an $N^*\times N^*$ identity matrix. From the eigenvalue Eq. (11), a total of $2N^*$ real and complex eigenvalues $\omega_i(i=1,2,\ldots 2N^*)$ can be obtained. We select the two real egenvalues with the lowest absolute values, one positive and the other negative, and these correspond, respectively, to the backward and forward traveling waves.

4 Numerical Results and Discussions

To facilitate comparison of data, all frequency parameters presented in this section are in the dimensionless form, $f=\omega b\sqrt{\rho h/A_{11}}=\omega b\sqrt{(1-\nu^2)\rho/E}$. Comparisons studies against results generated by the commercial finite element solver MSC/NASTRAN using eight-noded shell elements are first conducted. For these finite element results, well converged results were obtained using a 200×40 grid size or 8000 elements. The tabulated results of these comparison studies are presented in Table 1 for a range of parametric cases. For all four boundary

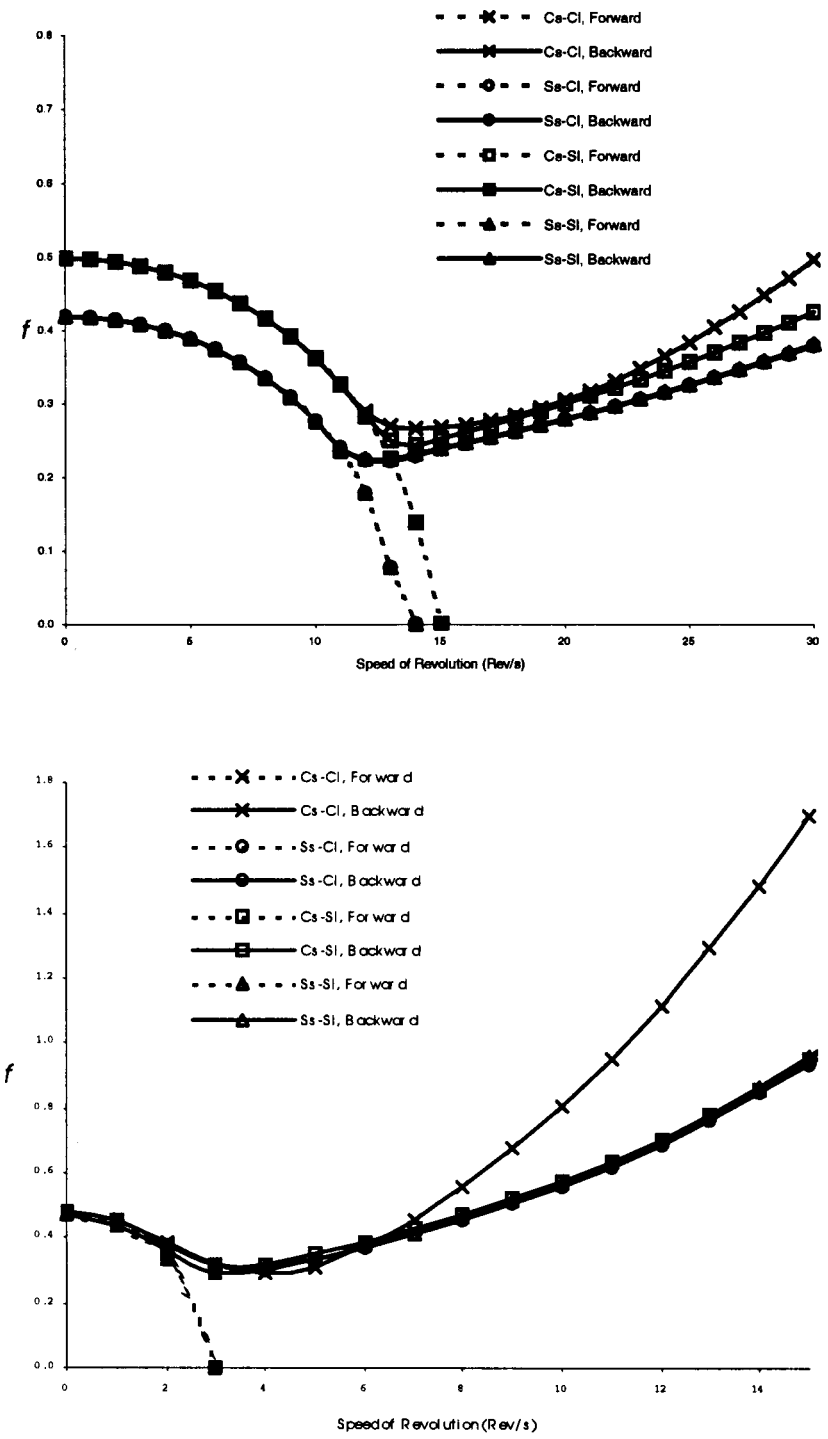


Fig. 3 Variation of frequency parameter $f=\omega b\sqrt{\rho h/A_{11}}$ for a rotating conical panel with different boundary conditions against revolution speed (rps), for different n [$n=1$ (top) and $n=5$ (bottom)], with panel parameters $\alpha=60°$, $\beta=120°$, $h/a=0.015$, $L/a=5$ and $m=1$

condition cases considered here, there is monotonic convergence as the number of grid points were increased. It is found that the use of 15 grid points produced sufficiently converged results with generally less than 1% difference when compared with results using 13 grid points. It is also observed that present GDQ results agree very well the finite element results for all four boundary condition cases with less than 4% difference between the two sets of results.

Figures 2 and 3 are two sets of graphical plots showing the changes in the vibration characteristics of the rotating conical panels when the boundary condition varies. Each figure shows the frequency behavior of the rotating conical panels at circumferential wave number $n=1$ and 5. Figure 2 shows the effects of the boundary conditions when both the subtended angle β and the half vertex angle α are low at values of 15° and 30°, respectively. Figure 3 shows the effects of the boundary conditions when both the subtended angle β and the half vertex angle α are high at values of 120°, and 60°, respectively.

In Figs. 2 and 3, the present results reveal a major qualitative difference from that of rotating complete (or full) truncated coni-

cal shells. For full conical shells, bifurcation of the natural frequencies into the forward and backward waves occur immediately upon the presence of rotation, see Lam and Li [1–3]. For the present conical panels, however, it is observed that this bifurcation does not occur until a certain rotating speed is reached. Once bifurcation occurs, the forward wave is observed to decrease rapidly to the critical speed value. This phenomenon is again distinct from that observed in the vibration of rotating full conical shells, see Lam and Li [1–3], where if critical speeds exist, it is generally as a result of the forward waves decreasing monotonically towards zero, see Ng and Lam [4] for detailed discussion on critical speed analysis. The boundary conditions, in the order which produces the highest to the lowest frequency parameters f, are (Cs-Cl)>(Ss-Cl)>(Cs-Sl)>(Ss-Sl). This is intuitively correct as the stiffer structures possess the higher frequencies. Further, in the same order, a rightward shift of the frequencies is observed, with the stiffer cases presenting more pronounced shifts.

In Fig. 3, it can be observed that, at low rotating speeds, the initial frequency parameters f are different for different boundary conditions. However, the same conclusions cannot be drawn from Fig. 2, in which the difference in the initial frequency parameters f are small for the different boundary conditions. It can thus be concluded that the effects of boundary conditions on the frequency behavior is small for relatively low speed rotating conical panels with relatively low values of both the subtended angle β and the half vertex angle α.

5 Conclusions

The free vibration study of thin rotating truncated conical panels has been carried out, the present results represent the first to be published for this class of problem. The generalized differential quadrature (GDQ) method was formulated for this analysis where the effects of initial hoop tension and the centrifugal and Coriolis accelerations due to rotation were all considered. A major qualitative difference, from that of rotating complete (or full) truncated conical shells, in that bifurcation of the forward and backward waves in rotating shell panels does not occur until a certain rotating speed is reached, was discovered as a result of this work. The effects of panel parameters and boundary conditions on the frequency characteristics of these rotating conical panels have also been examined.

References

[1] Li, Hua., and Lam, K. Y., 2001, ''Orthotropic Influence on Frequency Characteristics of a Rotating Composite Laminated Conical Shell by the Generalized Differential Quadrature Method,'' Int. J. Solids Struct., **38**, pp. 3995–4015.

[2] Lam, K. Y., and Li, Hua., 1997, ''Vibration Analysis of a Rotating Truncated Circular Conical Shell,'' Int. J. Solids Struct., **34**, pp. 2183–2197.

[3] Lam, K. Y., and Li, Hua., 2000, ''Generalized Differential Quadrature for Frequency of Rotating Multilayered Conical Shell,'' J. Eng. Mech., **126**, pp. 1156–1162.

[4] Ng, T. Y., and Lam, K. Y., 1999, ''Vibration and Critical Speed of a Rotating Cylindrical Shell Subjected to Axial Loading,'' Appl. Acoust., **56**, pp. 273–282.

Journal of Applied Mechanics

Brief Notes

A Brief Note is a short paper that presents a specific solution of technical interest in mechanics but which does not necessarily contain new general methods or results. A Brief Note should not exceed 2500 words *or equivalent* (a typical one-column figure or table is equivalent to 250 words; a one line equation to 30 words). Brief Notes will be subject to the usual review procedures prior to publication. After approval such Notes will be published as soon as possible. The Notes should be submitted to the Editor of the JOURNAL OF APPLIED MECHANICS. Discussions on the Brief Notes should be addressed to the Editorial Department, ASME International, Three Park Avenue, New York, NY 10016-5990, or to the Editor of the JOURNAL OF APPLIED MECHANICS. Discussions on Brief Notes appearing in this issue will be accepted until two months after publication. Readers who need more time to prepare a Discussion should request an extension of the deadline from the Editorial Department.

Comparison of Stresses in Center-Wound Rolls From Two Linear Elastic Models

W. R. Debesis and S. J. Burns
Department of Mechanical Engineering, Materials Science Program, University of Rochester, Rochester, NY 14627-0133

Two linear elastic models for describing stresses in center-wound rolls have been compared. One model includes wound on residual strains from the web while the other uses a clamped on stress boundary condition. The stress and displacement solutions to these two models, in the linear elastic limit, are shown to have analytical differences, and similarities.
[DOI: 10.1115/1.1571855]

Introduction

Web materials are wound onto rolls for storage, distribution, processing, and operations. A tension is applied to the web as the roll is being wound which produces stresses inside the roll. These stresses can cause defects and damage to the material in the roll, [1]. It is important to know these stresses as a function of winding tension so roll stresses can be predicted and defects can be avoided. The wound-on-strain (WOS) model and the clamped-on-stress (COS) models are both intended to describe stresses in center-wound rolls.

Linear elastic solutions for shrink-fitting reinforcing hoops onto gun cylinders were explored in the early part of the 20th century. Shrink-fitting was a common method used to strengthen large gun barrels. A close tolerance, thin outer cylinder is made to expand by heating and is then shrunk over an existing barrel when it is allowed to cool. The thermal shrinkage creates residual stresses between the thin cylinders that make up the gun barrel. These stresses are formed because neither cylinder can relax to its stress free shape.

Burns, Meehan, and Lambropoulos [2] and Altmann [3] formulated analytic elastic solutions for describing stresses in center-wound rolls. Burns et al. WOS solution for roll stresses was recently developed; the analysis converted the accreted residual, wound on strains from the web into an elastic boundary value problem which yielded the stresses in the roll. Altmann developed his displacement based intralayer accretion solution by balancing forces for a thin cylindrical shell clamped onto a roll as a stressed layer in the roll. His COS model has been the basis for most linear and nonlinear analyses on this topic, [4–10]. The WOS and COS analyses are formulated in stress and strain, respectively. The WOS and COS solutions were thought to be compatible since both analyses are based on the same deformation formulas and adaptations of Hooke's law. The stresses in center-wound rolls for assumed linear elastic media should be the same since both models describe the same problem. In the linear elastic limit, they were initially thought to be identical, i.e., elastic solutions. We show here in the linear elastic limit, differences between the models. Moreover, stresses in center-wound rolls are well known to be more complex as the media is both nonlinear and nonelastic. COS and WOS differences should be clarified before undertaking the complications of a realistic constitutive law for nonlinear, nonelastic stacked sheets.

Hakiel [5] and Benson [6] have nonlinear, nonanalytic solutions for the wound roll problem. Hakiel uses the same boundary conditions as the COS model while Benson utilizes a method which allows the outer lap to relax. The linear elastic differences between the COS and WOS models are as large as the nonlinear media effects are on accreting layers. If the COS model does not provide a good description of roll stresses then subsequent nonlinear work based on this model are suspect. This paper discusses how the inclusion of residual strains into the elastic solution compares with the inclusion of a clamped on stress boundary condition.

Stress Analysis

Boutaghou and Chase [4] by using Maxwell's reciprocal theorem, simplified Altmann's equations for the COS model. Here we are restricted to plane stress solutions that are assumed to be linear elastic. Equation (1) is the COS model's formula for radial stresses and Eq. (2) is the formula for circumferential stresses.

$$\sigma_r = -\frac{r^{2\beta}+a}{r^{\beta+1}} \times \int_r^R \frac{\sigma_w(t)t^\beta}{t^{2\beta}+a}\,dt \tag{1}$$

$$\sigma_\theta = \sigma_w(r) - \beta\left(\frac{r^{2\beta}-a}{r^{\beta+1}}\right) \times \int_r^R \frac{\sigma_w(t)t^\beta}{t^{2\beta}+a}\,dt \tag{2}$$

The symbol r is defined as the radius ratio, i.e., the actual radius divided by the outer radius of the core, and R is the outer radius

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Aug. 14, 2001; final revision, Dec. 16, 2002. Associate Editor: V. K. Kinra.

ratio of the roll. The stress of the web as it is being wound onto the roll is a function of the radius ratio for profiled rolls, and it is represented by $\sigma_w(r)$. The symbol t is an integration variable, and the constant a is used to simplify the formula after the first boundary condition is applied. β is a material property defined by Eq. (3).

$$\beta^2=\frac{E_\theta}{E_r} \tag{3}$$

E_r is the radial modulus of elasticity in the roll and E_θ is the circumferential modulus of elasticity in the roll.

The WOS model's formulas, as done by Burns et al., are given by Eq. (4) for radial stress and by Eq. (5) for circumferential stress.

$$\sigma_r=\frac{1}{r}\left\{\left[B\left(r^\beta-\frac{R^{2\beta}}{r^\beta}\right)\right]+\frac{1}{2\beta}\left[r^{-\beta}\int_r^R t^\beta\sigma^*(t)dt - r^\beta\int_r^R t^{-\beta}\sigma^*(t)dt\right]\right\} \tag{4}$$

$$\sigma_\theta=\frac{1}{r}\left\{\left[\beta B\left(r^\beta+\frac{R^{2\beta}}{r^\beta}\right)\right]-\frac{1}{2}\left[r^{-\beta}\int_r^R t^\beta\sigma^*(t)dt + r^\beta\int_r^R t^{-\beta}\sigma^*(t)dt\right]\right\} \tag{5}$$

The symbols r, R, and t are the same in the WOS model and the COS model, and B is a constant found by applying the inner boundary condition to the WOS solution.

The web stress in the WOS model is given by Eq. (6), which is an expression for the strains caused by the tension put on the web during the winding process.

$$\sigma^*(r)=\left(\frac{d}{dr}[r\sigma_w(r)]+\nu\sigma_w(r)\right) \tag{6}$$

The symbol ν is Poisson's ratio of the web, and E_θ is as defined before. $\sigma_w(r)$ is as defined previously.

Both solutions must be in equilibrium, which can be checked by Eq. (7).

$$\sigma_\theta=r\frac{d\sigma_r}{dr}+\sigma_r \tag{7}$$

Both models use Eq. (7) to help form second-order differential equations that are solved to formulate the equations for radial stresses; Eq. (7) is applicable to all linear and nonlinear solutions. Notice that according to Eq. (7) if the slope of the radial stress in the radial direction is zero, then the circumferential stress must equal the radial stress. This does not occur in any of the examples in this paper, but is good way to check all linear and nonlinear solutions graphically.

Both the COS and WOS models used here are based on second-order differential equations formulated from the plane stress form of Hooke's law and radial displacements. The COS model uses Hooke's law first then Eq. (7) to formulate a second-order differential equation for displacement. In the WOS model Eq. (7) is put into Hooke's law to form a second-order differential equation for radial stress. Another difference in these models is how the web stress is modeled. The WOS solution models the web stress as residual strains, which leads to the development of a particular solution. The COS model does not include the web stress residual strains in the development of its differential equation, instead the web stress is modeled in one of the boundary conditions. Radial stresses at the inside and outside of the roll are used as boundary conditions therefore the arbitrary constants are found using the formulas for radial stresses only. Circumferential stresses are obtained from Hooke's law and are checked by applying Eq. (7). Both models use similar methods of solution, yet they don't agree analytically nor do they give similar results for stresses in wound rolls as explained below.

Examples

The first example shows stresses for both models when the web is being held at a constant stress as the roll is wound.

$$\sigma_w(r)=\sigma_o \tag{8}$$

σ_o is the initial web stress. The material properties applied to the solutions are found in Table 1. The radial stresses for both models divided by the initial web stress are shown in Fig. 1(*a*). The circumferential stresses for both models divided by the initial web stress are shown in Fig. 1(*b*).

By looking at Figs. 1(*a*) and 1(*b*) it can be seen that the two models do not agree. The COS model's radial stresses in the roll are larger than the WOS model's radial stresses. The COS model's circumferential stresses are smaller than the WOS model's circumferential stresses at the core and become greater than them at the outside of the roll.

The next example is for constant torque. A moment placed on the core of the roll is kept constant during the winding process. The web stress decreases as the roll is wound. Constant torque is

Table 1 Physical properties

Property	Value
Modulus of elasticity in radial direction (E_r)	690 MPa
Modulus of elasticity in circumferential direction (E_θ) (E_{22})	4.14 GPa
Poisson's ratio of the web (ν)	0.28
Modulus of Elasticity of the core (E_c)	6.14 GPa
Outer radius ratio (R)	4
Initial web stress (σ_o)	2.30 MPa

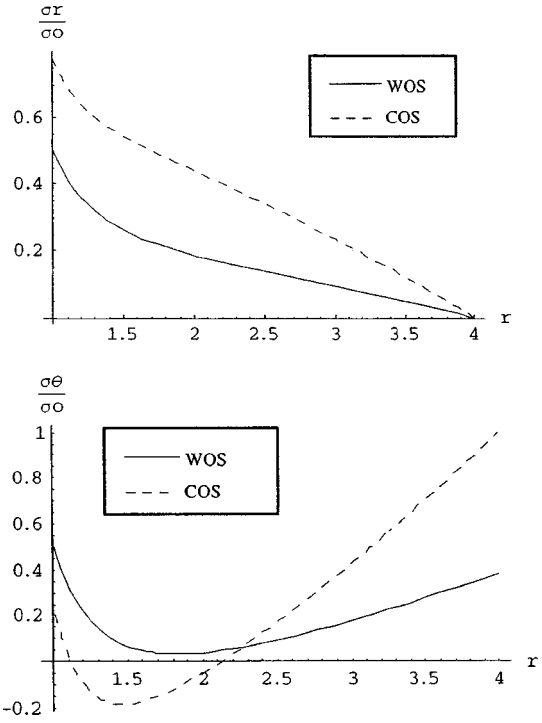


Fig. 1 (*a*) Radial stress ratio versus radius ratio for constant web tension, (*b*) circumferential stress ratio versus radius ratio for constant web tension

often how many machines are designed to wind rolls making this a very practical example. The web stress used in this example is stated in Eq. (9).

$$\sigma_w(r) = \frac{\sigma_o}{r} \tag{9}$$

Figures 2(*a*) and 2(*b*) show that the radial and circumferential stresses, respectively, both solutions are divided by the initial web stress. Again, the COS model's radial stresses are greater than the WOS model's radial stresses. In this example the COS model has only positive values for circumferential stresses. This shows that these two models have major differences and they are affected differently by changes in web stresses.

Comparing Figs. 1(*a*), 1(*b*), 2(*a*), and 2(*b*) it is evident that these solutions are very different even though they are based on the same equations. Timoshenko and Goodier [11] present a solution for residual thermal strains in a disk, as does Case. When the thermal strains in these solutions are replaced by residual strains of a web stress (see the Appendix), the solution equals the WOS models solution. In Fig. 1(*b*), the COS model's circumferential stress has a negative value in the roll close to the core while the WOS model's does not. This is an important difference because if the circumferential stress is in compression defects are more likely to occur in the roll. Reasons why the two solutions differ will be discussed in the next section.

Figures 3(*a*) and 3(*b*) show the WOS model's differential equation in radial stress without the residual strains from the particular solution and the COS model's differential equation in displacement solved using the boundary conditions of the COS model. The models use different differential equations yet come to the same stresses, when the web stress is modeled the same in each solution. Thus, the difference between the two analytical solutions is how the web stress is modeled and which boundary conditions

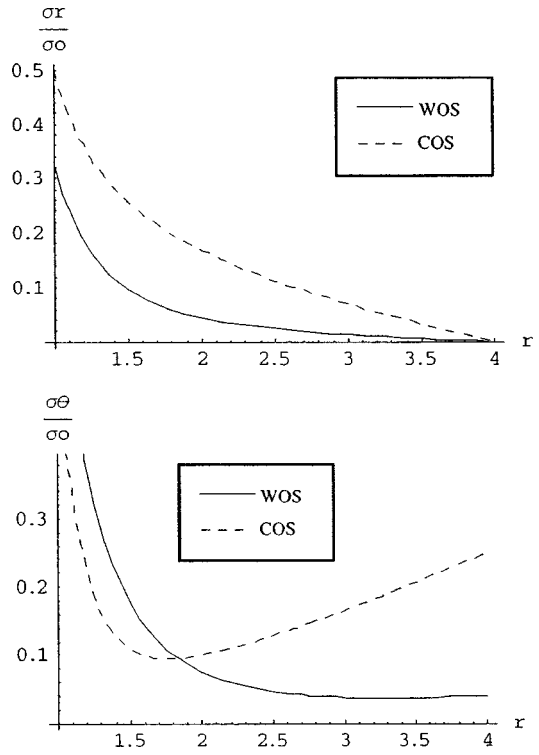


Fig. 2 (*a*) Radial stress ratio versus radius ratio for constant torque, (*b*) circumferential stress ratio versus radius ratio for constant torque

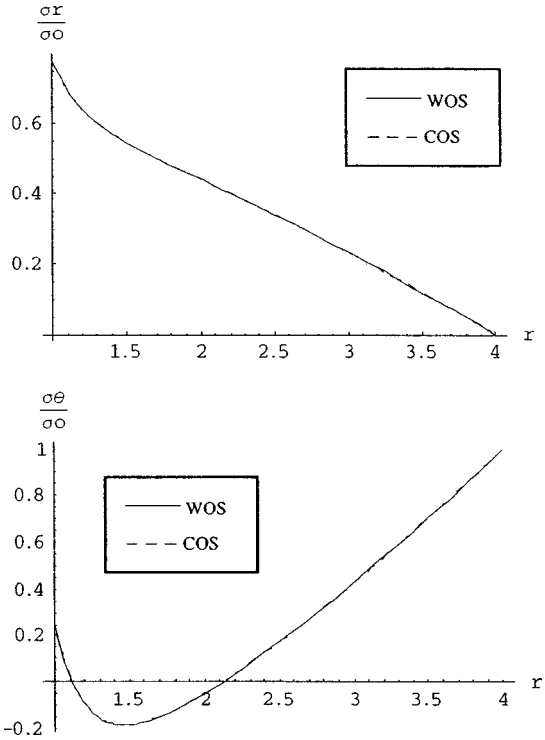


Fig. 3 (*a*) Radial stress ratio versus radius ratio of two solutions without residual strains, (*b*) circumferential stress ratio versus radius ratio of two solutions without residual strains

are used. Figures 3(*a*) and 3(*b*) show that both models have common parts as would be expected since they arise from the same equations and assumptions.

Inner Boundary Condition

Both models use Eq. (10) as a boundary condition, but it can be seen in Figs. 1(*a*) and 2(*a*) that the radial stresses of the two models are not equal at the core.

$$\sigma_r(r=1) = E_c \times u(r=1) \tag{10}$$

Equation (10) is used to find one arbitrary constant in both solutions. The symbol E_c is the modulus of elasticity of the core as already mentioned. The WOS model's solution for radial stress includes the residual strains caused by the web stress when this boundary condition is applied, while in the COS model's solution the web stress has not been factored into the equation. The inclusion of the web stress in this boundary condition causes the stresses of the two models to have different values at the core. In Fig. 3(*a*) and 3(*b*) it can be seen that the WOS and COS solutions equal each other at the core when the residual strains are not modeled in the WOS solution. Modeling the web stress as residual strains allows the material to relax once it is on the roll. This decreases the radial stress and increases the circumferential stress at the core. Why the web material is allowed to relax in the WOS model and not in the COS model will be discussed in the next section.

Outer Boundary Condition

The radial stress at the outside of any roll must be zero. No forces are applied externally to the roll so the radial stress must

equal zero. The outer boundary condition used in the WOS model and by Timoshenko and Goodier is

$$\sigma_r(R)=0. \tag{11}$$

Both models show that the radial stress is zero at the outer surface of the roll. The COS model's arbitrary constants are zero at the outside of the roll because they are integrated from r to the outer radius which causes the radial stress to be zero at the outside. When Eq. (11) is used as a boundary condition in the COS model's solution both arbitrary constants are zero and the solution is trivial.

Equation (12) is the boundary condition used by the COS model to find the second arbitrary constant.

$$\sigma_r(s)=-Q \tag{12}$$

The web stress in the COS model is applied as the second boundary condition. Where Q is defined as the incremental load change of pressure applied to the outside of the roll at $r=s$

$$Q=\frac{\sigma_w(s)}{s}\Delta s. \tag{13}$$

Equation (13) is defining Q as the radial pressure caused by the circumferential stress of the lap (Δs) on the outside of the roll at $r=s$. Equation (13) also states this circumferential stress is equal to the web stress. This is shown in Fig. 1(b) where the COS model's circumferential stress approaches the web stress at the outside of the roll. In the COS model the web material is not allowed to relax once it is on the roll. If the roll is rigid then the stress placed on it from the winding process holds constant. Equation (13) states that the roll and the outermost lap do not interact with each other.

The outer layer's circumferential stress would equal the web stress if the roll acts as a rigid body. This can only occur if the modulus of elasticity in the radial direction is much larger than the modulus of elasticity in the circumferential direction. By looking at Table 1 it can be seen that the modulus of elasticity in the radial direction is much less than the modulus of elasticity in the circumferential direction. Thus, it is elastically difficult to envision the outer lap's circumferential stress to equal the web stress.

If the roll is not rigid then for Eq. (13) to be true the web material must be inelastic. Both models have the web material modeled elastically in the roll. If the web material is elastic it will relax on the roll. The WOS model allows for this, as does Benson's model, [6]. For an accurate model the material in the winding process won't be elastic in the roll and inelastic in the web.

The difference in the models is not only different boundary conditions but also how the web stress is modeled in each solution. The WOS model includes the web stress as residual strains, which results in forming a particular solution. Modeling the web stress as residual strains allows the material to have the same constitutive behavior inside and outside the roll. In the COS model the web stress is applied though a boundary condition. This boundary condition gives the radial pressure caused by the circumferential stress in the outer most lap and this circumferential stress is always equal to the web stress. If this is true then either the roll is rigid or the outer most lap of the material is not allowed to relax and there is no friction between the web and the roll.

Conclusion

The tension that is put on the web while the roll is being wound creates strains in the material. The WOS model and the COS model are the same except for how they model the web stress. Modeling the web stress as residual strains allows the material to have the same constitutive behavior through out the winding process.

Reference [1] includes thermal residual strains in the solution for a disk, as do the solutions for the shrink fitting of gun barrels, [12]. The thermal strains can be thought of as web stresses for the wound roll problem with the winding process at constant temperature. The material in the winding process should have the same constitutive behavior through out the whole process. Modeling residual strains is an easy method to accomplish this.

It is unlikely that the outer lap of the roll is unable to relax. In Benson's nonlinear solution, [6,7], the outer lap is allowed to relax. Timoshenko's solution allows the thermal strains to convert to mechanical stresses as they cool. The outside lap of the roll will not relax if the roll is rigid, or the web material is perfectly compliant. This assumption is only present in one of the boundary conditions of the COS model. Modeling the web stress as residual strains precludes the WOS model to the have a particular solution which makes it possible to use two exact and consistent boundary conditions.

In the nonlinear problem the material will have different constitutive behavior in the roll and in the web. The constitutive law of the roll is typically very complaint for early laps. Thus it does not act like a rigid body while the web is elastic. In the linear problem the web and the roll have the same elastic properties.

All strains present in the winding process should be included in the inner boundary condition in the wound roll analysis for a realistic solution. Utilizing the method of residual strains to model the stresses brings continuity to the problem. If the roll is to be considered rigid this must be represented in forming the differential equation and not just in one boundary condition. The WOS model is a more accurate method for modeling the stresses in wound rolls in the elastic limit because the web stress is modeled as a residual strain.

Appendix

Timoshenko and Goodier [11] analyze stresses from thermal strains in a disk. This method can be converted to the wound roll problem, which will result in a differential equation in displacement that models the web stress as residual strains.

Timoshenko and Goodier sets up Hook's law in the following form, as does the WOS model:

$$\varepsilon_{rr}+\varepsilon_{rs}=\frac{\sigma_r}{E_r}-\frac{\nu\times\sigma_\theta}{E_\theta} \tag{14}$$

$$\varepsilon_{\theta\theta}+\varepsilon_{\theta s}=\frac{\sigma_\theta}{E_\theta}-\frac{\nu\times\sigma_r}{E_\theta}. \tag{15}$$

E_r is the radial modulus of elasticity, E_θ is the circumferential modulus of elasticity, and ν is Poisson's ratio in the circumferential direction. σ_r and σ_θ are the radial stress and the circumferential stress, respectively. ϵ_{rr} is the mechanical strain in the radial direction and ϵ_{rs} is the residual (formerly the thermal) parts of the strain in the radial direction. The mechanical strain in the circumferential direction is represented by $\epsilon_{\theta\theta}$ and $\epsilon_{\theta s}$ represents the residual strain in the circumferential direction. All parts of the strains must be added together to get the total strain. In Eq. (14) the ν_r/E_r term has been replaced with a ν/E_θ term which according to Maxwell's reciprocal theorem are equal.

Solving Eqs. (14) and (15) simultaneously for radial stress and circumferential stress, and then putting these equations into equi-

librium by using Eq. (7) form a differential equation. The strains can then be defined and placed into this equation. ϵ_{rr} and $\epsilon_{\theta\theta}$ are the parts of the mechanical radial and circumferential strains caused by displacements in the roll and are given as functions of radial displacements only (Eqs. (16) and (17)).

$$\varepsilon_{rr}=\frac{du(r)}{dr} \tag{16}$$

$$\varepsilon_{\theta\theta}=\frac{u(r)}{r} \tag{17}$$

Timoshenko and Goodier use Eqs. (18) and (19) for their thermal strains. Here thermal strains are replaced by the strains produced by the web stress, Eqs. (20) and (21).

$$\varepsilon_{rs}=\alpha T \tag{18}$$

$$\varepsilon_{\theta s}=\alpha T \tag{19}$$

$$\varepsilon_{rs}=-\frac{\nu}{E_\theta}\sigma_w(r) \tag{20}$$

$$\varepsilon_{\theta s}=\frac{\sigma_w(r)}{E_\theta} \tag{21}$$

The moduli of elasticity and Poisson's ratio are assumed to be equal in the web and in the roll.

The differential equation of the displacement $u(r)$ formed after the strains are substituted in is shown in Eq. (22).

$$\frac{d}{dr}\left[\frac{d[u(r)]}{dr}\right]+\frac{1}{r}\frac{d[u(r)]}{dr}-\beta^2\frac{u(r)}{r^2}=\frac{\sigma_w(r)}{r}\left(\frac{\beta^2-\nu^2}{E_\theta}\right), \tag{22}$$

where β is defined the same as before in Eq. (3).

Equation (26) is the resulting solution for $u(r)$ that includes the homogeneous and the particular solution

$$u(r)=Ar^\beta+Br^{-\beta}+\frac{\beta^2-\nu^2}{2E_\theta\beta}\left(r^{-\beta}\int_r^R t^\beta\sigma_w(r)dt-r^\beta\int_r^R t^{-\beta}\sigma_w(t)dt\right). \tag{23}$$

Equation (23) is then substituted back into Eqs. (14) and (15) to get the equations for the stresses. The equations for circumferential stress and radial stress are given in Eqs. (24) and (25) after Eq. (11) has been used as a boundary condition to solve for *A*.

$$\sigma_r(r)=-\frac{1}{r}\left(\frac{BE_\theta(r^{-\beta}-r^\beta R^{-2\beta})}{\beta+\nu}+\frac{1}{2\beta}\left(r^\beta(\beta+\nu)\times\int_r^R t^{-\beta}\sigma_w(t)dt+r^{-\beta}(\beta-\nu)\int_r^R t^\beta\sigma_w(t)dt\right)\right) \tag{24}$$

$$\sigma_\theta(r)=\frac{1}{r}\left(\frac{BE_\theta\beta(r^{-\beta}-r^\beta R^{-2\beta})}{\beta+\nu}+\frac{1}{2}\left(-r^\beta(\beta+\nu)\times\int_r^R t^{-\beta}\sigma_w(t)dt+r^{-\beta}(\beta-\nu)\int_r^R t^\beta\sigma_w(t)dt\right)\right)+\sigma_w(r) \tag{25}$$

Equation (26) shows the solution for *B* when Eq. (10) is used as the second boundary. Recall that Eq. (10) is used as a boundary condition in both models. In Timoshenko's and Goodier's example the disk has no inner radius thus the boundary condition he uses is the stresses equal zero at the center of the disk.

$$B=\frac{\left(R^{2\beta}(\beta+\nu)\left((\beta+\nu)(E_\theta+E_c(-\beta+\nu))\int_1^R t^{-\beta}\sigma_w(r)dt+(\beta-\nu)(E_\theta+E_c(\beta+\nu))\int_1^R t^\beta\sigma_w(t)dt\right)\right)}{(2E_\theta\beta(E_\theta(-1+R^{2\beta})+E_c((1+R^{2\beta})\beta+(-1+R^{2\beta})\nu)))} \tag{26}$$

Equations (24) and (25) don't look exactly like Eqs. (4) and (5), but if the web stress is defined in symbolic form and after some algebraic manipulation Eq. (24) will be equal to Eq. (4) as will Eq. (25) be equal to Eq. (5). The equations look different because the latter solves a differential equation for $u(r)$ and the former solves a differential equation in $\sigma_r(r)$. The COS models use a differential equation in displacement as does this Appendix. This shows that the differences in the two models are not in the analytic formalism but rather how the web stress is modeled.

References

[1] Smith, Duane R., 1995, *Roll and Web Defect Terminology*, Tappi Press.

[2] Burns, S. J., Meehnan, R. R., and Lambropoulos, J. C., 1999, "Strain-Based Formulas for Stresses in Profiled Center-Wound Rolls," Tappi J., **82**, pp. 159–167.

[3] Altmann, H. C., 1968, "Formulas for Computing the Stresses in Center-Wound Rolls," Tappi J., **51**, pp. 176–179.

[4] Boutaghou, Z.-E., and Chase, T. R., 1991, "Formulas for Generating Prescribed Residual-Stress Distributions in Center-Wound Rolls," ASME J. Appl. Mech., **58**, pp. 836–840.

[5] Hakiel, Z., 1987, "Nonlinear Model for Wound Roll Stresses," Tappi J., **70**, pp. 113–117.

[6] Benson, R. C., 1995, "A Nonlinear Wound Roll Model Allowing for Large Deformation," ASME J. Appl. Mech., **62**, pp. 853–859.

[7] Benson, R. C., 1996, Errata, ASME J. Appl. Mech., **63**, p. 418.

[8] Zabaras, N. et al., 1994, "A Hypoelastic Model for Computing the Stresses in Center-Wound Rolls of Magnetic Tape," ASME J. Appl. Mech., **61**, pp. 290–295.

[9] Yagoda, H. P., 1980, "Resolution of a Core Problem in Wound Rolls," ASME J. Appl. Mech., **47**, pp. 847–854.

[10] Quallis, W. R., and Good, J. K., 1997, "Thermal Analysis of a Round Roll," ASME J. Appl. Mech., **64**, pp. 871–876.

[11] Timoshenko, S., and Goodier, J. N., 1951, *Theory of Elasticity*, 2nd Ed., McGraw-Hill, New York, pp. 406–407.

[12] Case, J., 1925, *Strength of Materials*, Arnold, pp. 449–455.

Rate-Dependent Transition From Thermal Softening to Hardening in Elastomers

Z. Chen[1]
Department of Civil and Environmental Engineering, University of Missouri-Columbia, Columbia, MO 65211
Mem. ASME
e-mail: chenzh@missouri.edu

J. L. Atwood
Department of Chemistry, University of Missouri-Columbia, Columbia, MO 65211

Y.-W. Mai
Center for Advanced Materials Technology (CAMT), School of Aerospace, Mechanical and Mechatronic Engineering J07, The University of Sydney, Sydney NSW 2006, Australia
Fellow ASME

The thermal-mechanical properties of the materials currently used in packaging are being reexamined as the electronic packaging industry moves towards chip scale packages and wafer scale packages. The rate-dependent transition of elastic modulus and viscosity from thermal softening to thermal hardening with rising temperature, which does not involve any phase change, has been observed in certain elastomers. An explanation about this interesting phenomenon is given based on thermodynamic considerations. A theoretical analysis is performed to show the limitation of existing viscoelastic models in predicting the transition. It appears that macroscopic material properties should be reexamined based on the physics behind the interaction between ordinary elasticity and entropic elasticity. [DOI: 10.1115/1.1571860]

The effects of strain rate and temperature on the deformation of polymers are the challenging research topics of current interests due to the increasing use of polymers in extreme environments. A spring in series with a dashpot, and a spring and dashpot in parallel have been commonly employed to describe the deformation of a liquid with elasticity, and that of a viscous solid, respectively. To the authors' knowledge, however, the effects of temperature and strain rate on elastic modulus and viscosity have not been investigated in a comparative and systematic manner in small and/or large deformation cases, as can be seen from the representative references [1,2]. Especially, there exists a lack of understanding of the interaction between the ordinary elasticity, which is almost entirely due to internal energy and decreases with the increase of temperature, and the entropic elasticity that rises with the increase of temperature.

In recent experiments to benchmark commercial die attach adhesives, it was found that the elastic modulus and complex viscosity (dynamic loss modulus) of certain kinds of silicone elastomeric adhesives would experience a rate-dependent transition from thermal softening to thermal hardening with rising temperature. The observed rate-dependent thermomechanical process is reversible without any phase change. This kind of physical phenomenon has not been documented in the open literature. Hence, a phenomenological experimental study is performed here to explore the effects of temperature and loading rate on the elastic modulus and viscosity of a selected silicone elastomer, DA6501, the inherent properties of which have already been investigated by several researchers, [3,4].

Silicone elastomer is the best candidate for the compliant layers in the new electronic packages due to its low modulus and small moisture absorption. The low rotational energy of silicone-oxygen bond in silicone attributes to the highly flexible inorganic backbone, which yields a low modulus and a low glass transition temperature for silicone elastomeric adhesives, [3]. DA6501, made by Dow Corning toray silicone (DCTS), is a kind of one-part adhesives packaged in a syringe, which can be cured fast at low temperature without voids. The specimens used in this study were cured at 150°C for three minutes, following the procedure recommended by DCTS. The cure mechanism of DA6501 is a hydrosilyration reaction in crosslink formation. Silicones have a glass transition point (T_g) of −120°C and a melt transition point (T_m) of −43°C. There are no additional phase transitions above T_m. Normally, the weight loss of silicone only occurs at a temperature above 200°C. The relationship between the dimensional change of silicones and temperature is linear above T_m so that a very stable profile can be found for the thermal expansion coefficient over a broad temperature range of −40°C to 300°C, [4].

Based on the above properties of DA6501, the effects of temperature and loading rate on the elastic modulus and viscosity were measured over a temperature range of −50°C to 150°C and by means of cyclic deformations at three different frequencies, 1 Hz, 10 Hz, and 100 Hz. A dynamic mechanical analyzer, TA Instruments DMA 2980, was employed to measure the changes of the elastic (storage) modulus and complex viscosity of cured DA6501 samples with rising temperature and frequency. The length, width and thickness of the samples were 15 mm, 5 mm and 2 mm, respectively, and the multifrequency–tension mode was selected based on the test procedure recommended in the user manual of DMA 2980. As can be seen in Fig. 1, the elastic modulus experiences a transition from thermal softening to thermal hardening over a temperature range of −50°C to 150°C. The range of change in the elastic modulus is about 5%. The corresponding transition in the complex viscosity is shown in Fig. 2, with the rate effect on the viscosity being different from that on the elastic modulus. Note that the same sample was used for tests conducted at three different frequencies. The transition is repeatable. To verify the experimental data, three DA6501 samples were tested under the same conditions, with the result that the rate-dependent thermomechanical responses of these samples are similar.

Another kind of die attach adhesives, DA7920, was also tested by using the above test procedure. A similar transition from thermal softening to thermal hardening could be found with the change of elastic modulus being in the scale of 10 MPa instead of 1 MPa as shown for DA6501 in Fig. 1. Thus, the transition from thermal softening to thermal hardening appears to be common for silicone elastomers. Based on the physics of rubber elasticity, the internal-energy-based elasticity (the ordinary elasticity), which softens with the increase of temperature, could be offset by the entropic elasticity that hardens with rising temperature. As discussed by Treloar in detail, [1], the elastic modulus is insensitive to the change in frequency while the viscosity is inversely proportional to the frequency. To the authors' knowledge, however, the rate-dependent transition from thermal softening to thermal hardening as observed here has not been documented in the open

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Apr. 23, 2002; final revision, Nov. 10, 2002. Associate Editor: M.-J. Pindera.

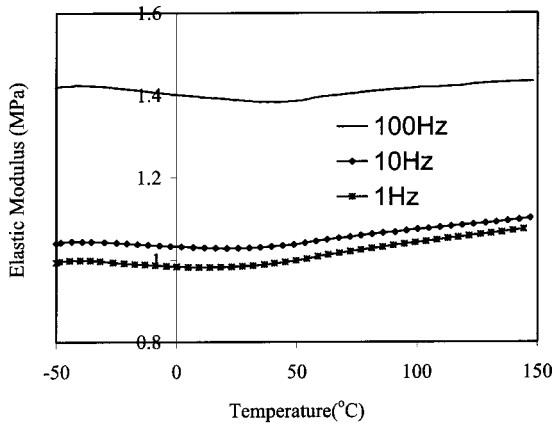


Fig. 1 Transition from thermal softening to thermal hardening of elastic modulus

literature. Especially, there exists a lack of quantitative understanding of the interaction between the ordinary elasticity and entropic elasticity. As demonstrated in the following theoretical analysis, the conventional constitutive models can not predict the observed rate-dependent transition.

For a viscoelastic solid as measured by the dynamic mechanical analyzer, the Kelvin-Voigt model takes the form of

$$s = \eta \frac{de}{dt} + Ee \tag{1}$$

where s denotes stress, e strain, and t time. The material constants E and η correspond, respectively, to a modulus of elasticity and a viscosity. For a cosinusoidal stress defined by

$$s = s_0 \cos \omega t \tag{2}$$

the general solution of Eq. (1) can be found to be

$$e = e_0 \cos(\omega t - \alpha) + c \tag{3}$$

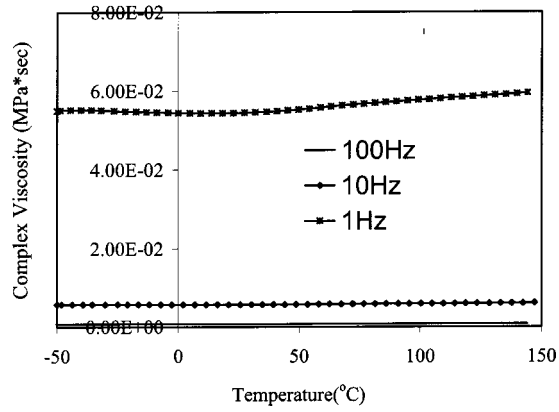


Fig. 2 Transition of complex viscosity corresponding to Fig. 1

with c being an integration constant. The amplitude e_0 and phase angle α are given by

$$e_0 = \frac{s_0}{E}\left[1 + \omega^2\left(\frac{\eta}{E}\right)^2\right]^{-1/2} \tag{4}$$

and

$$\alpha = \tan^{-1}\left(\omega \frac{\eta}{E}\right). \tag{5}$$

If $\tau = \eta/E$ is used to represent the time of relaxation, it can be found from Eq. (4) that the amplitude ratio of stress to strain is

$$\frac{s_0}{e_0} = E(1 + \omega^2 \tau^2)^{1/2}. \tag{6}$$

With the assumption that the time required for the molecules to change their configurations is governed by the rate of passage of chain segments across potential barriers, τ would exponentially decay with rising temperature, [1]. Based on the conventional constitutive models, therefore, the amplitude ratio of stress to strain would increase with the increase of loading frequency ω, but decrease with rising temperature, as can be seen from Eq. (6). However, the transition from thermal softening to hardening, as shown in Figs. 1 and 2, indicates the limitation of existing models.

Although the observed transition form thermal softening to hardening in the elastic modulus is small as can be seen from the experimental data, there exists an urgent need to accurately characterize the rate-dependent thermal-mechanical responses of the materials currently used in electronic packaging. In view of the recent claim that silicones maintain consistent properties without any transition throughout the temperature range of −40°C to 300°C, [3], the rate-dependent transition reported here deserves another look at the physics of the thermal-mechanical properties of silicon elastomers. Especially, a combined analytical, experimental and numerical effort is required in the further study to quantitatively understand the interaction between the ordinary elasticity and entropic elasticity.

Acknowledgments

The authors are grateful to Prof. F. Hsieh in the Department of Biological and Agricultural Engineering at the University of Missouri-Columbia, Prof. G. Kanel at the Institute of Problems of Chemical Physics of the Russian Academy of Sciences, Prof. R. Li in the Department of Physics and Materials Science at the City University of Hong Kong, and Dr. M. Kent and Dr. D. Adolf at Sandia National Laboratories for valuable discussions. The experimental assistance by Miss P. Chou is appreciated. The authors are also grateful to the reviewers for discerning comments on this paper. This research was sponsored in part by the National Science Foundation, and by the City University of Hong Kong.

References

[1] Treloar, L. R. G., 1975, *The Physics of Rubber Elasticity*, 3rd Ed., Clarendon Press, Oxford, UK.
[2] Hillmansen, S., Hobeika, S., Haward, R. N., and Leevers, P. S., 2000, "The Effect of Strain Rate, Temperature, and Molecular Mass on the Tensile Deformation of Polyethylene," Polym. Eng. Sci., **40**, p. 481.
[3] Lee, Y. J., 2000, "Silicone, Chip Scale Package and Its Reliability," *Proceedings of the Fifth Pan Pacific Microelectronics Symposium*, G. Chai and B. J. Han, eds., Maui, HI, Jan., Surface Mount Technology Association, Edina, MN.
[4] Wilson, S. W., Norris, A., Benson, E., Watson, M., and Mine, K., 2000, "Key Properties and Requirements for Materials Used in Chip Scale Packaging," *Proceedings of the Fifth Pan Pacific Microelectronics Symposium*, G. Chai and B. J. Han, eds., Maui, HI, Jan., Surface Mount Technology Association, Edina, MN.

Elastic Waves Induced by Surface Heating in a Half-Space

J. P. Blanchard
University of Wisconsin–Madison, 1500 Engineering Drive, Madison WI 53706-1609
e-mail: blanchard@engr.wisc.edu

Rapid surface heating will induce waves in an elastic material. Closed-form solutions for the resulting longitudinal and transverse thermal stresses are derived using Laplace Transforms. The model is one-dimensional, consisting of a half-space subjected to a step change in the surface heating. The transverse stress at the wave peak is found to exceed the surface stress for short times, while for long times the surface stress far exceeds either of the stresses at the wave peak. Both the longitudinal and transverse stresses at the peak of the wave reach steady-state values after a few dimensionless times. [DOI: 10.1115/1.1571861]

Introduction

Rapid surface heating, such as that created by a laser, can induce numerous phenomena in solids. Some of these include melting, vaporization, thermal waves, [1], and plastic deformation. In many applications, such as mirrors, such phenomena must be avoided in order to ensure a long life. In this paper, I derive a closed-form solution for one-dimensional elastic waves induced by a step change in surface heating. This creates a temperature field in which the surface temperature increases as the square root of the pulse time. It is assumed that the heat is deposited at the surface, there is no cooling, the heat transport is diffusional, and that the elastic and thermal equations are uncoupled.

Several previous works have analyzed thermoelastic waves in solids. Sternberg and Chakravorty [2] solved the problem for both a step and ramp change in surface temperature. Gladysz [3] solved the problem for a surface temperature changing as $t^2\exp(-\alpha t)$ while White [4] solved it for surface heating which varied harmonically and Boley and Weiner [5] solve the problem for convection boundary conditions. Gladysz obtains a series solution, while the others are provided in closed form. Other studies have included the effects of coupling of the elastic and thermal equations, as well as heat deposition below the surface and thermal waves. Bushnell and McCloskey [6] obtained a closed-form solution for elastic waves due to volumetric heating, modeling the deposition using a nonzero attenuation coefficient for the heat incident on the surface of the solid. However, they ignored diffusion, assuming that the temperature profile matched the deposition profile. Similarly, Mozina and Dovc [7] and Galka and Wojnar [8] each assumed a volumetric heating given in the form $Q'''=a\mu\exp(-\mu x)$ where μ is the attenuation coefficient of the heat incident on the surface, but they included diffusion. All of these papers provide closed-form solutions. Boley and Tolins [9] modeled the case with a step change in temperature, but included the coupling of the thermal and elastic equations by adding in a heating term which depended on the local strain rate. The solutions were obtained in terms of an integral and approximations were given in closed form for short and long times. Kao [10] solves the problem analytically for the non-Fourier case, in which the time scales of the heating are such that thermal waves are induced. Wang and Xu [11] included thermal waves, as well as volumetric heating and coupling of the thermal and elastic equations. Series solutions are obtained. These latter papers introduce undue complications for cases in which non-Fourier effects are insignificant and are of questionable value since the hyperbolic equation solved has not been validated experimentally and has been shown to yield nonphysical results for three-dimensional problems, [12]. The time scales for which the solution derived in this paper is valid are discussed in the Results section of this paper.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Apr. 28, 2002; final revision, Dec. 19, 2002. Associate Editor: R. C. Benson.

Modeling

I begin by considering thermoelastic deformation of a half-space, with x denoting the perpendicular distance from the surface. Following Sternberg and Chakravorty [2], I define the following dimensionless variables:

$$\xi=\frac{x}{a}$$

$$\tau=\frac{\kappa t}{a^2}$$

$$\phi=\frac{kT}{qa}$$

$$\hat{\sigma}_x=\frac{(1-2\nu)}{2(1+\nu)}\frac{k\sigma_x}{\alpha q a\mu} \tag{1}$$

where $\hat{\sigma}$ is the dimensionless axial stress, ϕ is the dimensionless temperature, ξ and τ are the dimensionless coordinates for space and time, respectively, κ is the thermal diffusivity, q is the surface heat load, α is the thermal expansion coefficient, μ is the shear modulus, ν is Poisson's ratio, and

$$a=\frac{\kappa}{c}$$

$$c^2=\frac{2(1-\nu)\mu}{(1-2\nu)\rho}. \tag{2}$$

Here c is the wave speed and ρ is the density of the solid.

With these definitions, the governing equations then become

$$\frac{\partial^2\phi}{\partial\xi^2}=\frac{\partial\phi}{\partial\tau}$$

$$\frac{\partial^2\hat{\sigma}_x}{\partial\xi^2}=\frac{\partial^2\hat{\sigma}_x}{\partial\tau^2}+\frac{\partial^2\phi}{\partial\tau^2}. \tag{3}$$

It is assumed here that the only nonzero displacement is perpendicular to the surface of the half-space. That is, $u_y=u_z=0$. The initial conditions are such that all temperatures, stresses, and time derivatives are nonexistent. The boundary conditions are that the temperatures and stresses vanish at x equals infinity, while at the surface

$$q=-k\frac{\partial T}{\partial x}$$

$$\sigma_x=0 \tag{4}$$

which implies

$$\frac{\partial\phi}{\partial\xi}=-1$$

$$\hat{\sigma}_x=0. \tag{5}$$

The solution for the dimensionless temperature is given by, [13],

$$\phi = 2\left(\sqrt{\frac{\tau}{\pi}} \exp\left[\frac{-\xi^2}{4\tau}\right] - \frac{\xi}{2} \operatorname{erfc}\left[\frac{\xi}{2\sqrt{\tau}}\right]\right) \tag{6}$$

and its second derivative can thus be found to be

$$\frac{\partial^2 \phi}{\partial \tau^2} = \exp\left[\frac{-\xi^2}{4\tau}\right] \frac{(\xi^2 - 2\tau)}{4\sqrt{\pi}\tau^{5/2}}. \tag{7}$$

The Laplace transform of this function is ([14], p. 246, #15)

$$\overline{\phi} = \sqrt{s} \exp(-\xi\sqrt{s}) \tag{8}$$

where s is the Laplace parameter and the bar over the function is meant to denote the transformed instance of the time function. Taking the transform of Eq. (3) and using the fact that all the initial values of the stress and its time derivatives are 0 gives

$$\frac{d^2\overline{\sigma}_x}{d\xi^2} - s^2\overline{\sigma}_x = \sqrt{s} \exp(-\xi\sqrt{s}). \tag{9}$$

Solving this equation and using the stress free boundary conditions yields

$$\overline{\sigma}_x = \frac{\exp(-\xi\sqrt{s})}{(1-s)\sqrt{s}} - \frac{\exp(-\xi s)}{(1-s)\sqrt{s}} = \overline{\sigma}_1 + \overline{\sigma}_2. \tag{10}$$

Inverting the first term in this function requires some manipulation. I begin by letting $p = \sqrt{s}$, giving us

$$\overline{f} = \frac{\exp(-\xi p)}{p(1-p^2)}. \tag{11}$$

The inverse of this function is ([14] p. 183, #22 and p. 170, #15)

$$f(\tau) = [1 - \cosh(\tau - \xi)] H(\tau - \xi) \tag{12}$$

where $H(z)$ is the Heaviside step function. Given this result for $f(\tau)$, one can then find the time dependence of the original function using ([14], p. 171, #29)

$$\hat{\sigma}_1 = \frac{1}{2\sqrt{\pi}t^{3/2}} \int_{\xi}^{\infty} u \exp\left(\frac{-u^2}{4\tau}\right) [1 - \cosh(u - \xi)] du \tag{13}$$

where the lower limit on the integral has been changed to reflect the step function in $f(\tau)$. Carrying out this integral and inverting the second term ([14], p. 221, #1 and p. 170, #15) gives

$$\hat{\sigma}_x = -\frac{1}{2} \exp(\tau - \xi) \left\{ 1 - \exp(2\xi) \operatorname{erfc}\left(\frac{2\tau + \xi}{2\sqrt{\tau}}\right) + \operatorname{erf}\left(\frac{2\tau - \xi}{2\sqrt{\tau}}\right) - 2 \operatorname{erf}(\sqrt{\tau - \xi}) H(\tau - \xi) \right\}. \tag{14}$$

Having obtained this stress, we can obtain the remaining stresses by taking advantage of the assumption that there is no displacement parallel to the surface. Defining two more dimensionless stresses as

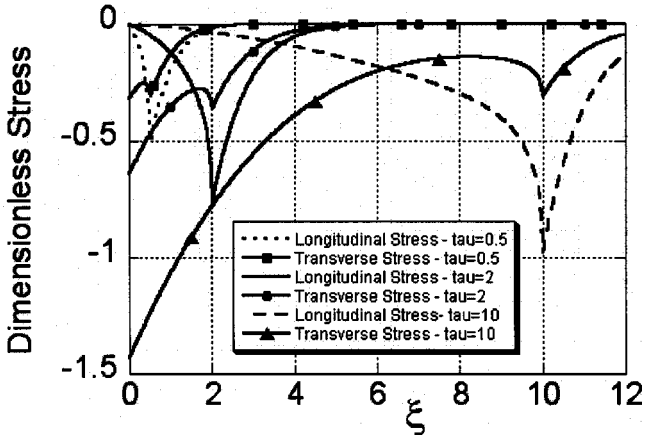


Fig. 1 Shape of stress fields at three different times. Transverse stresses assume $\nu = 0.3$.

$$\hat{\sigma}_y = \frac{(1-2\nu)(1-\nu)}{2(1+\nu)} \frac{k\sigma_y}{\alpha q a \mu}$$

$$\hat{\sigma}_z = \frac{(1-2\nu)(1-\nu)}{2(1+\nu)} \frac{k\sigma_z}{\alpha q a \mu} \tag{15}$$

we obtain

$$\hat{\sigma}_y = \hat{\sigma}_z = \nu\hat{\sigma}_x - (1-2\nu)\phi. \tag{16}$$

Substituting Eq. (14) into this expression gives us a solution for the two remaining normal stresses. This completes our solution for the stresses induced by surface heating on a half-space.

Since the longitudinal stress ($\hat{\sigma}_x$) is zero at the surface, the transverse stress ($\hat{\sigma}_y$) at the surface is given by

$$\hat{\sigma}_y = \hat{\sigma}_z = -(1-2\nu)\phi \tag{17}$$

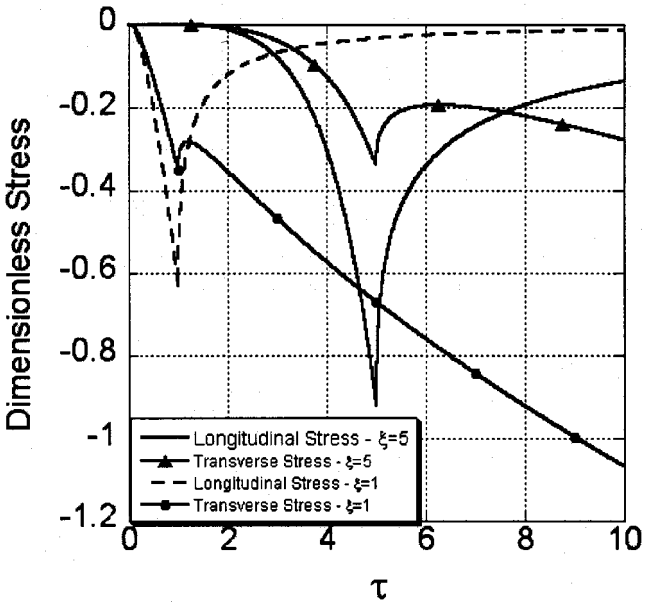


Fig. 2 Time dependence of stresses at two different locations relative to the surface. Transverse stresses assume $\nu = 0.3$.

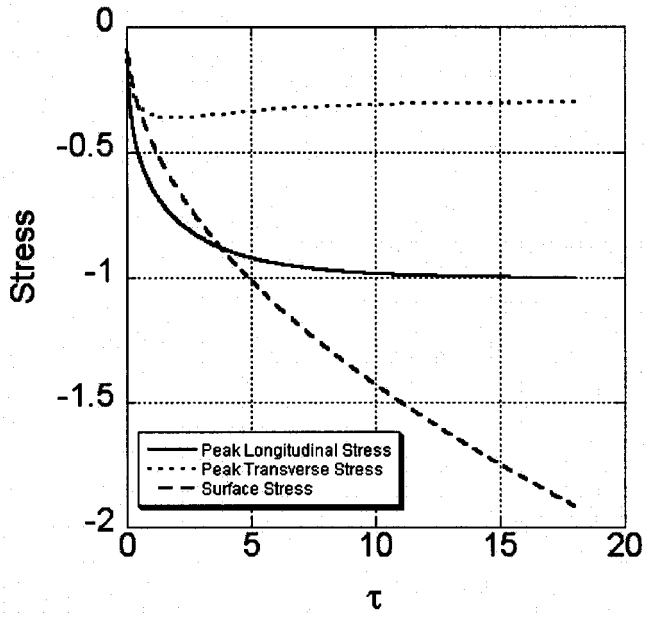


Fig. 3 Stresses versus time at the surface and at the peak of the wave. Transverse stresses assume $\nu=0.3$.

or

$$\hat{\sigma}_y=\hat{\sigma}_z=-2(1-2\nu)\left(\sqrt{\frac{\tau}{\pi}}\right). \tag{18}$$

The peak stress in the wave occurs at $\xi=\tau$. Substituting this into Eq. (14) gives

$$\hat{\sigma}_x=-\frac{1}{2}\left\{1-\exp(2\tau)\operatorname{erfc}\left(\frac{3\sqrt{\tau}}{2}\right)+\operatorname{erf}\left(\frac{\sqrt{\tau}}{2}\right)\right\}$$

$$\hat{\sigma}_y=-\frac{\nu}{2}\left\{1-\exp(2\tau)\operatorname{erfc}\left(\frac{3\sqrt{\tau}}{2}\right)+\operatorname{erf}\left(\frac{\sqrt{\tau}}{2}\right)\right\}-2(1-2\nu)$$
$$\times\left(\sqrt{\frac{\tau}{\pi}}\exp\left[\frac{-\tau}{4}\right]-\frac{\tau}{2}\operatorname{erfc}\left[\frac{\sqrt{\tau}}{2}\right]\right). \tag{19}$$

For long times the longitudinal stress approaches -1 while the transverse stress approaches ν. For short times, the leading terms for the longitudinal and transverse stresses are

$$\hat{\sigma}_x\approx 2\sqrt{\frac{\tau}{\pi}}$$

$$\hat{\sigma}_y\approx 2(1-\nu)\sqrt{\frac{\tau}{\pi}}. \tag{20}$$

Results

Typical wave shapes are shown in Fig. 1, which plots the two dimensionless stresses as a function of distance from the surface. These results are given for dimensionless times of 0.5, 1, and 10. As one would expect, the stresses are all compressive, and the peak stress in the wave occurs at $\xi=\tau$. Except at early times, the transverse stress peaks at the surface because that's where the temperature peaks. At early times, there is a local peak in the transverse stress where the wave front lies, and at this point the transverse stress is less than the longitudinal stress.

Figure 2 displays the time dependence of the stresses at various depths from the surface. Each of the pairs of the curves is given at dimensionless times of 1 and 5. It is clear from this figure that the long-term behavior is dominated by the quasi-static stress while the short-term behavior is dominated by inertial effects.

The peak in the longitudinal stress occurs at $\xi=\tau$ while the peak transverse stress occurs at the surface (except at short times). These peaks are plotted in Fig. 3, which gives both stresses at $\xi=\tau$ along with the surface stress at the same dimensionless time. It can be seen that beyond a dimensionless time of approximately 4, the surface stress exceeds the stress at the wave peak. The point where the two are equal can be found more precisely by numerically solving Eqs. (18) and (19), giving a value of 3.81 for a Poisson's ratio of 0.3.

By comparing the surface stress to the peak stress in the wave the results presented here are of interest for dimensionless times less than about 10. This corresponds to approximately 5 ps for aluminum and 25 ps for iron at room temperature. The pulse times of interest can also be much longer, because the stress wave will propagate much faster than the heat diffuses, and can cause spallation at a free surface at the back of a solid. In this case it will be the absolute magnitude of the peak stress in the wave, rather than its relation to the surface stress, that is of interest. On the other hand, the results are only valid for times long compared to the relaxation time associated with non-Fourier conduction. In most metals this relaxation time is less than 0.01 picoseconds at room temperature and even smaller at higher temperatures. Hence, there is a pulse length window from tens of fs to tens of ps (or greater) for which this solution is valid and meaningful.

Acknowledgments

This work was sponsored by the Naval Research Laboratory in support of the High Average Power Laser program.

References

[1] Joseph, D. D., and Preziosi, L., 1989, "Heat Waves," Rev. Mod. Phys., **61**(1), p. 41.
[2] Sternberg, E., and Chakravorty, J., 1959, "On Inertia Effects in a Transient Thermoelastic Problem," ASME J. Appl. Mech., **26**, p. 503.
[3] Gladysz, J., 1985, "Propagation of a Plane Wave in a Thermoelastic Half-Space Under Smooth Heating of Its Boundary," J. Therm. Stresses, **8**, p. 227.
[4] White, R., 1963, "Generation of Elastic Waves by Transient Surface Heating," J. Appl. Phys., **34**(12), p. 3559.
[5] Boley, B., and Weiner, J., 1960, *Theory of Thermal Stresses*, John Wiley and Sons, New York, p. 54.
[6] Bushnell, J., and McCloskey, D., 1968, "Thermoelastic Stress Production in Solids," J. Appl. Phys., **39**(12), p. 5541.
[7] Mozina, J., and Dovc, M., 1994, "One-Dimensional Model of Optically Induced Thermoelastic Waves," Mod. Phys. Lett. A, **8(28)**, p. 1791.
[8] Galka, A., and Wojnar, R., 1995, "One-Dimensional Dynamic Thermal Stresses Generated in an Elastic Half-Space by Laser Pulses," J. Therm. Stresses, **18**, p. 113.
[9] Boley, B., and Tolins, I., 1962, "Transient Coupled Thermoelastic Boundary Value Problems in the Half-Space," ASME J. Appl. Mech., **29**, p. 637.
[10] Kao, T., 1976, "On Thermally Induced Non-Fourier Stress Waves in a Semi-Infinite Medium," AIAA J., **14**(6), p. 818.
[11] Wang, X., and Xu, X., 2001, "Thermoelastic Wave Induced by Pulsed Laser Heating," Appl. Phys. A: Mater. Sci. Process., **73**, p. 107.
[12] Körner, C., and Bergmann, H. W., 1998, "The Physical Defect of the Hyperbolic Heat Conduction Equation," Appl. Phys. A: Mater. Sci. Process., **67**, p. 397.
[13] Carslaw, H., and Jaeger, J., 1959, *Conduction of Heat in Solids*, Oxford Clarendon Press, London, p. 75.
[14] Kaufman, R., 1966, *Table of Laplace Transforms*, W. B. Saunders, Philadelphia, PA.

Thermal Stresses in an Infinite Elastic Pipe Weakened by a Finite Cylindrical Crack

S. Itou
Professor, Department of Mechanical Engineering,
Kanagawa University, Rokkakubashi,
Kanagawa-ku, Yokohama 221-8686, Japan
e-mail: itous001@kanagawa-u.ac.jp
Fax: +81-45-491-7915

Axially symmetric thermal stresses in the vicinity of a finite cylindrical crack in an elastic pipe are calculated. The surfaces of the crack are assumed to be insulated. The outer surface of the pipe is heated so as to maintain a constant temperature T_d, and the inner surface of the pipe is cooled so as to maintain a constant temperature T_b. Expressions developed by Sharma are used to solve the problem. Stress intensity factors are defined and calculated numerically for several configurations of the pipe.
[DOI: 10.1115/1.1598475]

1 Introduction

A pipe is one type of mechanical part used in plant construction. If a low-temperature liquid flows in a metallic pipe in a high-temperature environment in order to remove heat, tensile stress may be produced in the pipe, possibly leading to development of a finite cylindrical crack. To the author's knowledge, thermal stresses resulting from a cylindrical crack have not been presented. Therefore, in the present paper, axially symmetric thermal stresses are solved for an elastic pipe weakened by a finite cylindrical crack, by use of the expressions provided by Sharma [1].

In a first step, the boundary conditions relating to the temperature field are reduced to dual integral equations by the Fourier transform technique. To satisfy the boundary conditions outside the crack, the temperature difference at the crack surfaces is expanded to a series of functions that diminish to zero outside the cracks. The unknown coefficients in the series are determined by the Schmidt method [2]. Next, the boundary conditions relating to the stress field are reduced to dual integral equations. To solve the equations, the differences in the displacements at the crack surfaces are newly expanded in a series of functions that diminish to zero outside the crack. The Schmidt method [3] is also used to solve the unknown coefficients in the series.

2 Fundamental Equation

With respect to the cylindrical coordinates (r,θ,z) shown in Fig. 1, consider a cylindrical crack located on $r=c$ and extending from $z=-a$ to $z=+a$. The inner surface and the outer surface of the hollow cylinder are denoted by $r=b$ and $r=d$, respectively. For the sake of convenience, the hollow cylinder is divided into an inner layer (1) denoted by $(b\leqslant r\leqslant c)$ and an outer layer (2) denoted by $(c\leqslant r\leqslant d)$.

If displacements u_r and u_z are defined by the potential functions ϕ_1, ϕ_2, and T^*, in a manner similar to that developed by Sharma [1],

$$u_r=\partial(\phi_1+\phi_2+T^*)/\partial r,\quad u_z=\partial(\beta_1\phi_1+\beta_2\phi_2+\mu\partial T^*)/\partial z, \tag{1}$$

the axially symmetric stresses can be expressed by

$$\begin{aligned}
\tau_{rr}&=(c_{11}\partial^2/\partial r^2+c_{12}\times 1/r\times\partial/\partial r)(\phi_1+\phi_2+T^*)\\
&\quad+c_{13}\partial^2/\partial z^2(\beta_1\phi_1+\beta_2\phi_2+\mu T^*)-b_1T,\\
\tau_{\theta\theta}&=(c_{12}\partial^2/\partial r^2+c_{11}\times 1/r\times\partial/\partial r)(\phi_1+\phi_2+T^*)\\
&\quad+c_{13}\partial^2/\partial z^2(\beta_1\phi_1+\beta_2\phi_2+\mu T^*)-b_1T,\\
\tau_{zz}&=c_{13}(\partial^2/\partial r^2+1/r\times\partial/\partial r)(\phi_1+\phi_2+T^*)\\
&\quad+c_{13}\partial^2/\partial z^2(\beta_1\phi_1+\beta_2\phi_2+\mu T^*)-b_2T,\\
\tau_{rz}&=c_{44}\partial^2/\partial r\partial z[(1+\beta_1)\phi_1+(1+\beta_2)\phi_2+(1+\mu)T^*],
\end{aligned} \tag{2}$$

with

$$b_1=(c_{11}+c_{12})\alpha_1+c_{13}\alpha_2,\quad b_2=2c_{13}\alpha_1+c_{33}\alpha_2, \tag{3}$$

where u_r and u_z are r and z components of displacement, c_{ij} are elastic moduli, and α_1 and α_2 are the coefficients of linear expansion along the r axis and the z axis, respectively. In Eq. (2), temperature T satisfies

$$(\partial^2/\partial r^2+1/r\times\partial/\partial r+\kappa^2\partial^2/\partial z^2)T=0 \tag{4}$$

with

$$\kappa^2=k_2/k_1, \tag{5}$$

where k_1, k_2 are the thermal conductivities along the r axis and the z axis, respectively. Potential functions ϕ_1 and ϕ_2 must satisfy the equation

$$(\partial^2\phi/\partial r^2+1/r\times\partial\phi/\partial r)+\nu^2\partial^2\phi/\partial z^2=0, \tag{6}$$

where ν_j^2 $(j=1,2)$ are the roots of the equation

$$c_{11}c_{44}\nu^4+(2c_{13}c_{44}-c_{11}c_{33}+c_{13}^2)\nu^2+c_{33}c_{44}=0. \tag{7}$$

In Eqs. (1) and (2), β_j^2 $(j=1, 2)$ are given by the equation

$$\beta_j=(c_{11}\nu_j^2-c_{44})/(c_{13}+c_{44}). \tag{8}$$

Functions T^* and μ can be calculated by use of the equations

$$\begin{aligned}
&c_{11}(\partial^2T^*/\partial r^2+1/r\times\partial T^*/\partial r)+[c_{44}+\mu(c_{13}+c_{44})]\partial^2T^*/\partial z^2\\
&\quad-b_1T=0,\\
&(c_{13}+c_{44}+\mu c_{44})(1/r\times\partial T^*/\partial r+\partial^2T^*/\partial r^2)+\mu c_{33}\partial^2T^*/\partial z^2\\
&\quad-b_2T=0.
\end{aligned} \tag{9}$$

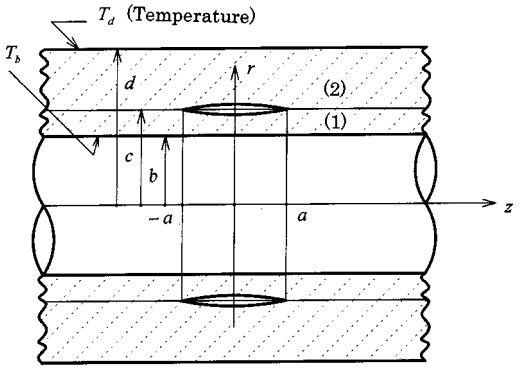


Fig. 1 Geometry and coordinate system

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, July 1, 2002; final revision, Apr. 2, 2003. Associate Editor: J. R. Barber.

3 Boundary Conditions

Consider the case where the outer surface of the pipe is heated so as to maintain a constant temperature T_d, and the inner surface of the pipe is cooled so as to maintain a constant temperature T_b. If the crack surfaces are assumed to be thermally insulated, the boundary conditions for the temperature field are given by the equations

$$T_1 = T_b, \quad \text{for } r=b, \quad |z|<\infty, \tag{10}$$

$$T_2 = T_d, \quad \text{for } r=d, \quad |z|<\infty, \tag{11}$$

$$\partial T_1/\partial r = \partial T_2/\partial r, \quad \text{for } r=c, \quad |z|<\infty, \tag{12}$$

$$\partial T_1/\partial r = 0, \quad \text{for } r=c, \quad |z|<a, \tag{13}$$

$$T_1 = T_2, \quad \text{for } r=c, \quad a<|z|, \tag{14}$$

where the variables denoted by subscripts 1, 2 are those for layers (1) and (2), respectively. When the crack faces are assumed to not come into contact, the boundary conditions for the stress field are given by the equations

$$\tau^b_{rr1}=0, \quad \tau^b_{rz1}=0, \quad \text{for } r=b, \quad |z|<\infty, \tag{15}$$

$$\tau^d_{rr2}=0, \quad \tau^d_{rz2}=0, \quad \text{for } r=d, \quad |z|<\infty, \tag{16}$$

$$\tau^c_{rr1}=\tau^c_{rr2}, \quad \tau^c_{rz1}=\tau^c_{rz2}, \quad \text{for } r=c, \quad |z|<\infty, \tag{17}$$

$$u^c_{r1}=u^c_{r2}, \quad u^c_{z1}=u^c_{z2}, \quad \text{for } c=0, \quad a<|z|, \tag{18}$$

$$\tau^c_{rr1}=0, \quad \tau^c_{rz1}=0, \quad \text{for } r=c, \quad |z|<a, \tag{19}$$

where superscripts b, c, and d denote variables at $r=b$, $r=c$, and $r=d$, respectively.

4 Analysis

To obtain the solutions, we introduce the Fourier transforms

$$\overline{f}(\xi) = \int_{-\infty}^{\infty} f(z)\exp(i\xi z)dz, \tag{20}$$

$$f(z) = 1/(2\pi)\int_{-\infty}^{\infty} \overline{f}(\xi)\exp(-i\xi z)d\xi. \tag{21}$$

In order to satisfy Eq. (14), the temperature difference at $r=c$ is expanded by the series

$$\pi(T^c_1 - T^c_2) = \sum_{n=1}^{\infty} c_n \cos[(2n-1)\sin^{-1}(z/a)] \quad \text{for } |z|<a,$$
$$\pi(T^c_1 - T^c_2) = 0 \quad \text{for } a<|z|, \tag{22}$$

where c_n are unknown coefficients and the superscript c denotes the values at $r=c$. Then, as is easily shown, the boundary conditions with respect to the temperature field can be reduced to the form

$$\sum_{n=1}^{\infty} c_n F_n(z) = -u(z), \quad \text{for } 0 \leq z < a, \tag{23}$$

Table 1 Material constants of steel

E (GPa)	205.9
ν'	0.3
$\alpha(\times 10^{-5}/°C)$	1.14
k [W/m °C)]	48.6

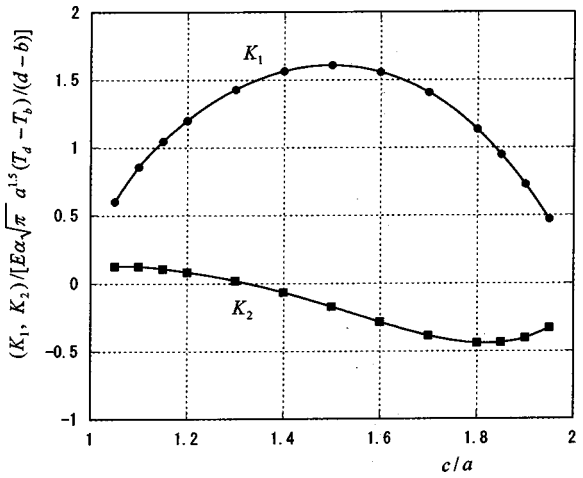


Fig. 2 Stress intensity factors K_1 and K_2 for $b/a=1.0$, $d/a=2.0$ versus c/a

where the expressions of the known functions $F_n(z)$ and $u(z)$ are omitted. Equation (23) can be solved for coefficients c_n by means of the Schmidt method [2], whereby the entire temperature field can be obtained.

Next, the stress field is obtained. Equation (18) shows that the displacements are continuous outside the crack. To satisfy these conditions, the differences in the displacements are expanded by the series

$$\pi(u^c_{z1} - u^c_{z2}) = \sum_{n=1}^{\infty} d_n \sin[2n\sin^{-1}(z/a)], \quad \text{for } r=c, \quad |z|<a,$$
$$= 0 \quad \text{for } r=c, \quad a<|z|, \tag{24}$$

$$\pi(u^c_{r1} - u^c_{r2}) = \sum_{n=1}^{\infty} e_n \cos[(2n-1)\sin^{-1}(z/a)],$$
$$\text{for } r=c, \quad |z|<a,$$
$$= 0 \quad \text{for } r=c, \quad a<|z|,$$

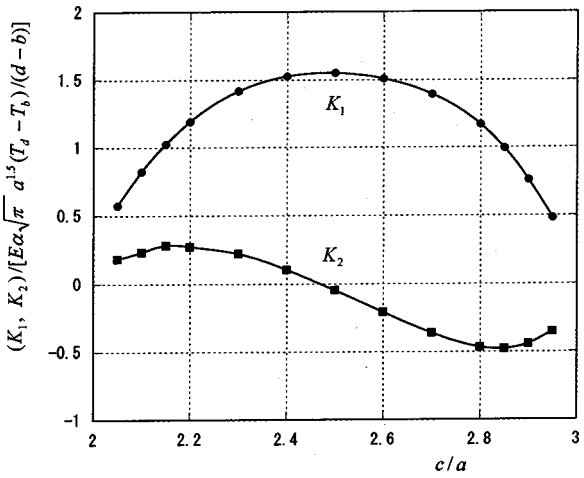


Fig. 3 Stress intensity factors K_1 and K_2 for $b/a=2.0$, $d/a=3.0$ versus c/a

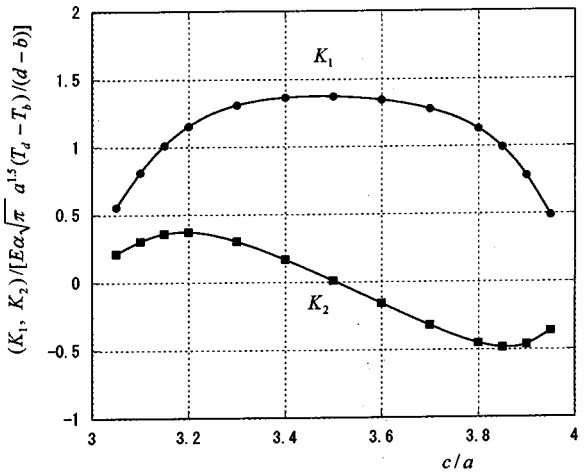


Fig. 4 Stress intensity factors K_1 and K_2 for $b/a=3.0$, $d/a=4.0$ versus c/a

where d_n, e_n are the unknown coefficients to be determined. Then, as is also easily shown, the boundary conditions with respect to the stress field can be reduced to the form

$$\sum_{n=1}^{\infty} d_n G_n(z)+\sum_{n=1}^{\infty} e_n H_n(z)=-U(z),$$
$$\sum_{n=1}^{\infty} d_n K_n(z)+\sum_{n=1}^{\infty} e_n L_n(z)=-V(x) \quad \text{for } 0\leqslant z<a, \tag{25}$$

where the expressions of the known functions $G_n(z)$, $H_n(z)$, $K_n(z)$, $L_n(z)$, $U(z)$, and $V(z)$ are omitted. Equation (25) can be solved for coefficients d_n, e_n by means of the Schmidt method [3], whereby the entire stress field can be obtained.

5 Stress Intensity Factors

Since coefficients c_n, d_n, e_n are now known, the entire temperature and stress fields can be obtained. In fracture mechanics, determining the stresses just ahead of the crack end is important. The stress singularities around the crack tip derive from the behavior of the integrand when the integration variable assumes an infinite value. Therefore the stress intensity factors can be determined as follows:

$$K_1=[2\pi(z-a)]^{1/2}\tau^c_{rr1}\big|_{z\to a+}$$
$$=\sum_{n=1}^{\infty} e_n(2n-1)(-1)^{n-1}Q^L_2/(\pi a)^{1/2},$$
$$K_2=[2\pi(z-a)]^{1/2}\tau^c_{rz1}\big|_{z\to a+}=\sum_{n=1}^{\infty} d_n(2n)(-1)^n Q^L_4/(\pi a)^{1/2}, \tag{26}$$

where the expressions of the known constants Q^L_2 and Q^L_4 are omitted.

6 Numerical Examples and Results

Numerical calculations are carried out for a steel pipe. Table 1 lists the material constants of the steel, where E is Young's modulus, ν' is Poisson's ratio, α is the coefficient of linear expansion, and k is the thermal conductivity. In the isotropic case, the constants in orthotropic elasticity can be expressed in terms of E, ν', α, and k as follows:

$$c_{11}=c_{33}=(1-\nu')E/[(1+\nu')(1-2\nu')],$$
$$c_{12}=c_{13}=\nu' E/[(1+\nu')(1-2\nu')],$$
$$c_{44}=E/(1+\nu'), \quad k_1=k_2=k, \quad \alpha_1=\alpha_2=\alpha. \tag{27}$$

The present analysis is based on orthotropic elasticity. This presents no problem in solving the temperature field for an isotropic material; however, Eq. (7) has two kinds of multiple roots, and the analysis is invalid for this case. However, if the values of c_{33} and c_{13} are replaced by values slightly larger than those of c_{11} and c_{12}, the solutions presented in this paper are still effective. This can be verified in the solution for a transversely isotropic material [4–7]. Here, the following values are assigned to c_{33} and c_{13}:

$$c_{33}=1.01\times c_{11}, \quad c_{13}=1.01\times c_{12}. \tag{28}$$

The semi-infinite integrals which appear in $F_n(z)$ in Eq. (23) and those which appear in $G_n(z)$, $H_n(z)$, $K_n(z)$, $L_n(z)$, $U(z)$, and $V(z)$ in Eq. (25) can be easily evaluated numerically by means of Filon's method, because the integrands decay rapidly as the integration variable ξ increases.

The Schmidt method has been applied to solve coefficients c_n in Eq. (23) and d_n, e_n in Eq. (25), truncating the infinite series to 15 terms. The left-hand side of Eq. (23) has been confirmed to coincide with the right-hand side of Eq. (23). The same applies to Eq. (25). Namely, the boundary conditions inside the crack are seen to have been completely satisfied.

$(d-b)/a=1.0$ is assumed. The results of the stress intensity factors $(K_1,K_2)/[E\alpha\sqrt{\pi}a^{1.5}(T_d-T_b)/(d-b)]$ are plotted in Figs. 2–4 for $(b/a=1,\ d/a=2)$, $(b/a=2,\ d/a=3)$, $(b/a=3,\ d/a=4)$, respectively.

One the basis of the numerical calculations outlined above, we can draw the following conclusions:

(1) K_1 has a large value near $r=(d-b)/2$, and becomes smaller as the crack approaches the outer surface or the inner surface. The value of K_2 is considerably smaller than that of K_1, and its sign changes near $r=(d-b)/2$. The value of K_2 has two peaks, which are located near the inner and outer surfaces of the pipe, respectively. The predominant value of the stress intensity factor is K_2 in the two-dimensional problem, whereas it is K_1 for a cylindrical crack in the axially symmetric problem.

(2) The peak value of K_1 near $r=(d-b)/2$ decreases slightly with increasing diameter of the pipe, whereas the absolute values of the two peaks of K_2 increase slightly.

References

[1] Sharma, B., 1958, "Thermal Stresses in Transversely Isotropic Semi-Infinite Elastic Solid," ASME J. Appl. Mech., **25**, pp. 86–88.
[2] Morse, P. M., and Feshbach, H., 1958, *Methods of Theoretical Physics 1*, McGraw-Hill, New York, p. 926.
[3] Yau, W. F., 1967, "Axisymmetric Slipless Indentation of an Infinite Elastic Cylinder," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **15**, pp. 219–227.
[4] Atsumi, A., and Itou, S., 1973, "Stresses in a Transversely Isotropic Slab Having a Spherical Cavity," ASME J. Appl. Mech., **40**, pp. 752–758.
[5] Atsumi, A., and Itou, S., 1974, "Stresses in a Transversely Isotropic Circular Cylinder Having a Spherical Cavity," ASME J. Appl. Mech., **41**, pp. 507–511.
[6] Atsumi, A., and Itou, S., 1974, "Stresses in a Transversely Isotropic Half Space Having a Spherical Cavity," ASME J. Appl. Mech., **41**, pp. 708–711.
[7] Atsumi, A., and Itou, S., 1976, "Axisymmetric Thermal Stresses in a Transversely Isotropic Circular Cylinder Having a Spherical Cavity," ASME J. Appl. Mech., **43**, pp. 431–433.

Benchmark Results for the Problem of Interaction Between a Crack and a Circular Inclusion

J. Wang, S. G. Mogilevskaya, and S. L. Crouch
Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive S.E., Minneapolis, MN 55455, USA

This paper is a reply to the challenge by Helsing and Jonsson (2002, ASME J. Appl. Mech., **69***, pp. 88–90) for other investigators to confirm or disprove their new numerical results for the stress intensity factors for a crack in the neighborhood of a circular inclusion. We examined the same problem as Helsing and Jonsson using two different approaches—a Galerkin boundary integral method (Wang et al., 2001, in Rock Mechanics in the National Interest, pp. 1453–1460) (Mogilevskaya and Crouch, 2001, Int. J. Numer. Meth. Eng.,* **52***, pp. 1069–1106) and a complex variables boundary element method (Mogilevskaya, 1996, Comput. Mech.,* **18***, pp. 127–138). Our results agree with Helsing and Jonsson's in all cases considered.* [DOI: 10.1115/1.1598473]

1 Introduction

Helsing and Jonsson [1] recently initiated a discussion on the accuracy of benchmark tables and graphical results presented in the applied mechanics literature. They considered the problem of a matrix crack interacting with an elastic inclusion (Fig. 1). Their converged results for the normalized stress intensity factors of the crack differed from those presented in papers by Erdogan, Gupta, and Ratwani [5], and by Cheeseman and Santare [6]. In view of this discrepancy, the authors challenged other investigators to confirm or disprove their new results. As a response, we would like to participate in this discussion and present our numerical solutions to the same problems.

2 Methods of Solution

We used two methods to solve the problem depicted in Fig. 1—a Galerkin boundary integral (GBI) method [2,3] and a complex variables boundary element method (CVBEM) [4]. These two methods are independent, although both of them are based on a complex hypersingular integral equation [7] written in terms of the tractions on the boundary of the inclusion and the displacement discontinuities along the crack. In the GBI method, the boundaries of the inclusion and the crack are not subdivided into elements; instead, global approximations are used for the unknown boundary parameters. The tractions along the boundary of the circular inclusion are represented by truncated complex Fourier series and the distributions of displacement discontinuity along the straight crack are approximated by a series of Chebyshev polynomials multiplied by a weight function, which takes into account the crack tip asymptotics. A Galerkin (weighted residual) procedure is used to construct a system of linear algebraic equations. In the CVBEM, the boundaries are discretized into small straight or circular arc elements, on each of which a local approximation is used. The unknown boundary parameters are approximated by complex Lagrange polynomials of the second degree. For the crack tip elements, square-root asymptotes are used. The linear system is obtained by using a collocation formulation, in which three collocation points on each element coincide with the nodes of the Lagrange polynomials. In both methods, all of the integrations are performed analytically.

3 Results and Discussion

The problems under investigation are shown in Fig. 1. We refer to the problem with the straight crack as case 1 and the one with the curved crack as case 2. The elastic properties of the matrix and the inclusion are μ_1, υ_1 and μ_2, υ_2, respectively. In both cases, following Helsing and Jonsson [1], we take $\mu_1=1$, $\upsilon_1=0.35$, and $\mu_2=23$, $\upsilon_2=0.30$.

For case 1, we recompute the results in Table 1 in Ref. [1] by using both methods mentioned above. In the GBI method, the unknowns are the complex coefficients involved in the truncated series. The number of terms of the complex Fourier series for the inclusion is $2n$ (n negative terms plus n positive terms) and the number of terms of the series of Chebyshev polynomials for the straight crack is m. In the CVBEM, the number of collocation points is k. The numbers of degrees of freedom for the two methods are $4n+2m$ and $2k$, respectively. The degrees of freedom are chosen such that convergence to a fixed numerical result is obtained. The stress intensity factors at the crack tips are normalized by dividing them by $\sigma_0\sqrt{\pi a}$. Our results for the normalized mode 1 and mode 2 stress intensity factors F_1 and F_2 at the left and right crack tips from the above two methods (denoted by GBI and CVBE) are shown in Tables 1 and 2, together with the results obtained by Helsing and Jonsson [1] (HJ) and Erdogan, Gupta, and Ratwani [5] (EGR).

It can be seen from Tables 1 and 2 that our results from two different methods agree closely with Helsing and Jonsson's for all values of c/a considered except for the case $c/a=3$, where we believe that the difference may be due to a misprint (i.e., we believe that their result should be -0.0035 rather than -0.035). The results from the GBI method depend on the number of terms of the truncated series. When the crack is not very close to the inclusion, the method converges very fast and only a few terms of the complex Fourier series and the series of Chebyshev polynomials give accurate solutions. For the cases where the inclusion and crack are very close to one another, many more terms are needed to take into account the strong interaction between them. For example, 80 and 100 terms are needed in the Fourier series and 20 terms in the series of Chebyshev polynomials to get the asymptotic solution for the two most extreme cases where $c/a=0.5$ and 0.3, respectively. In the CVBEM, 696 and 870 collocation points were required to get the converged results for these two cases. We find that the GBI method involves approximately one-sixth as many degrees of freedom as the CVBEM. For the GBI method the convergence of our results for F_1 and F_2 at the left tip of the crack with the increase of the number n is shown in Fig. 2 for the cases $c/a=1.0$ and 0.5. It can be seen from these figures that the convergence is much faster for $c/a=1.0$ than for $c/a=0.5$. For $c/a=2.0$, Helsing and Jonsson [1] gave a 16-digit reference value of F_1 at the right tip of the crack computed with approximately 600 discretization points. We obtained the first nine digits of that value with 20 terms in the Fourier series and 15 terms in the series of Chebyshev polynomials.

For case 2, we recomputed the results shown in Fig. 2 in Ref. [1] by using the CVBEM. For the circular inclusion, we used 100 circular arc elements for $R_c/R=1.1$ and 1.2 and 50 for $R_c/R>1.2$. For the circular arc crack, the numbers of elements for different cases were chosen such that all of the elements on the inclusion and the crack have approximately the same length. For example, for $R_c/R=1.1$, we had 48, 28, and 20 circular arc elements for $\theta=75°$, $\theta=45°$, and $\theta=30°$, respectively. The stress

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Apr. 12, 2002; final revision, Feb. 22, 2003. Associate Editor: B. M. Moran.

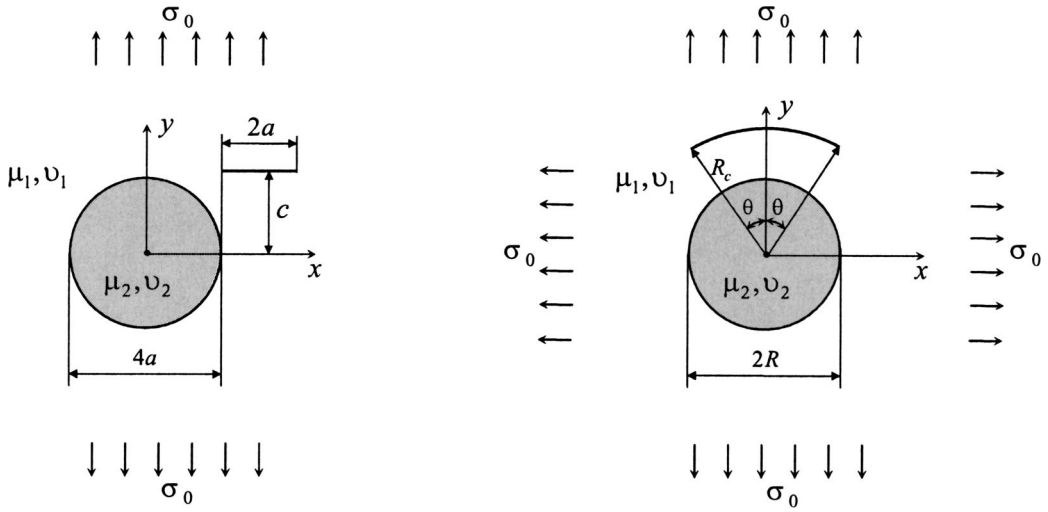


Fig. 1 A straight (left) or circular arc (right) crack outside an inclusion under uniaxial or biaxial tension. This figure corresponds to Fig. 1 in Ref. [1].

Table 1 Normalized stress intensity factors at the left tip of the straight crack

	F_1				F_2			
c/a	GBI	CVBE	HJ	EGR	GBI	CVBE	HJ	EGR
0.3	0.236	0.234	0.235	0.225	0.074	0.073	0.073	0.072
0.5	0.347	0.347	0.347	0.341	0.102	0.102	0.102	0.101
1.0	0.613	0.614	0.613	0.613	0.061	0.061	0.061	0.057
1.5	0.755	0.755	0.755	0.763	0.012	0.012	0.012	−0.007
2.0	0.830	0.830	0.830	0.845	0.018	0.018	0.018	−0.021
3.0	0.936	0.936	0.936	0.953	0.067	0.067	0.067	−0.001
4.0	1.003	1.003	1.003	1.014	0.079	0.079	0.079	0.002
8.0	1.043	1.043	1.043	1.043	0.032	0.032	0.032	−0.026

Table 2 Normalized stress intensity factors at the right tip of the straight crack

	F_1				F_2			
c/a	GBI	CVBE	HJ	EGR	GBI	CVBE	HJ	EGR
0.3	0.790	0.790	0.790	0.784	−0.023	−0.023	−0.023	−0.004
0.5	0.797	0.797	0.797	0.792	−0.037	−0.037	−0.037	−0.006
1.0	0.817	0.817	0.817	0.817	−0.067	−0.067	−0.067	−0.005
1.5	0.833	0.833	0.833	0.839	−0.074	−0.074	−0.074	0.008
2.0	0.850	0.850	0.850	0.860	−0.058	−0.058	−0.058	0.034
3.0	0.897	0.897	0.897	0.905	−0.004	−0.004	−0.035	0.089
4.0	0.947	0.947	0.947	0.951	0.032	0.032	0.032	0.117
8.0	1.022	1.022	1.022	1.020	0.032	0.032	0.032	0.088

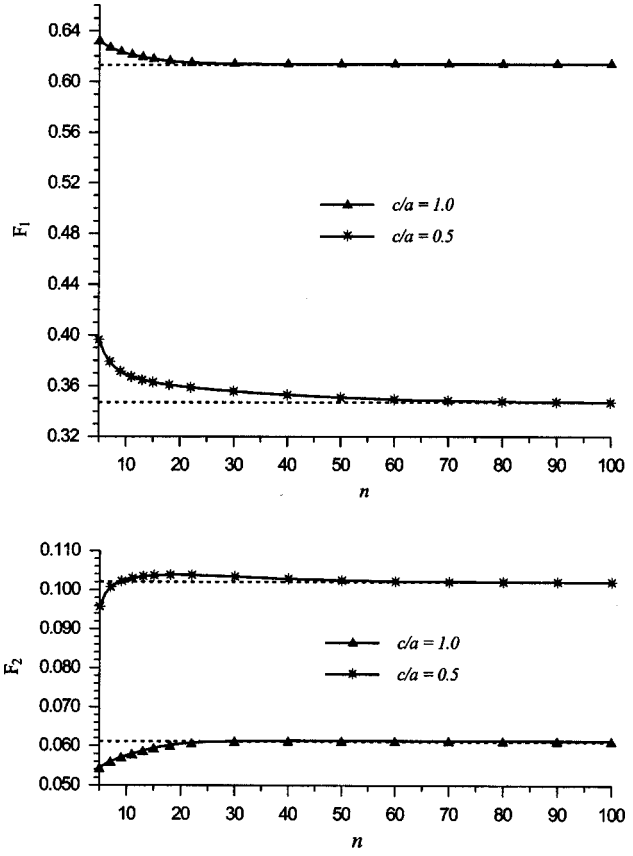


Fig. 2 Convergence of F_1 and F_2 at the left tip of the straight crack with increase of the number of terms of the Fourier series (the dotted lines denote Helsing and Jonsson's results for $c/a = 0.5$ and 1.0)

Table 3 Normalized stress intensity factors at the tips of the circular crack

	$\theta=30°$		$\theta=45°$		$\theta=75°$	
R_c/R	F_1	F_2	F_1	F_2	F_1	F_2
1.1	0.919	1.494	0.928	1.393	0.955	1.363
1.2	0.944	1.353	0.962	1.285	1.009	1.281
1.5	0.961	1.202	0.990	1.159	1.056	1.160
2.0	0.972	1.104	0.992	1.092	1.059	1.084
3.0	0.986	1.040	0.992	1.043	1.039	1.037
4.0	0.992	1.021	0.994	1.024	1.025	1.021
5.0	0.995	1.013	0.996	1.016	1.017	1.014
6.0	0.997	1.009	0.997	1.011	1.012	1.010

intensity factors are normalized by the corresponding values of the stress intensity factors in the absence of the inclusion. Our results for the normalized stress intensity factors F_1 and F_2 at the left tip of the crack were plotted and found to be in good agreement with Helsing and Jonsson's. We have tabulated our results (Table 3) to make it easier for others to make comparisons.

In this paper, we presented our numerical results for two problems involving interaction between a crack and a circular inclusion. Our two independent calculations confirmed the results in Table 1 in Ref. [1]. The comparison between the two methods used in this investigation showed that the Galerkin boundary integral method is much more efficient than the collocation boundary element method.

References

[1] Helsing, J., and Jonsson, A., 2002, "On the Accuracy of Benchmark Tables and Graphical Results in the Applied Mechanics Literature," ASME J. Appl. Mech., **69**, pp. 88–90.

[2] Wang, J., Mogilevskaya, S. G., and Crouch, S. L., 2001, "A Galerkin Boundary Integral Method for Nonhomogeneous Materials With Cracks," *Rock Mechanics in the National Interest*, D. Elsworth, J. Tinucci, and K. Heasley, eds., Balkema, Lisse, The Netherlands, pp. 1453–1460.

[3] 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.

[4] Mogilevskaya, S. G., 1996, "The Universal Algorithm Based on Complex Hypersingular Integral Equation to Solve Plane Elasticity Problems," Comput. Mech., **18**, pp. 127–138.

[5] Erdogan, F., Gupta, G. D., and Ratwani, M., 1974, "Interaction Between a Circular Inclusion and an Arbitrarily Oriented Crack," ASME J. Appl. Mech., **41**, pp. 1007–1013.

[6] Cheeseman, B. A., and Santare, M. H., 2000, "The Interaction of a Curved Crack With a Circular Inclusion," Int. J. Fract., **103**, pp. 259–277.

[7] Linkov, A. M., and Mogilevskaya, S. G., 1994, "Complex Hypersingular Integrals and Integral Equations in Plane Elasticity," Acta Mech., **105**, pp. 189–205.