Transactions of the ASME®

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

Journal of Applied Mechanics

Published Bimonthly by The American Society of Mechanical Engineers

VOLUME 69 • NUMBER 3 • MAY 2002

TECHNICAL PAPERS

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

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

Journal of Applied Mechanics Volume 69, Number 3 MAY 2002

BRIEF NOTES

ANNOUNCEMENTS

Journal of Applied Mechanics

Announcement

Robert M. McMeeking Appointed Editor of the *Journal of Applied Mechanics*

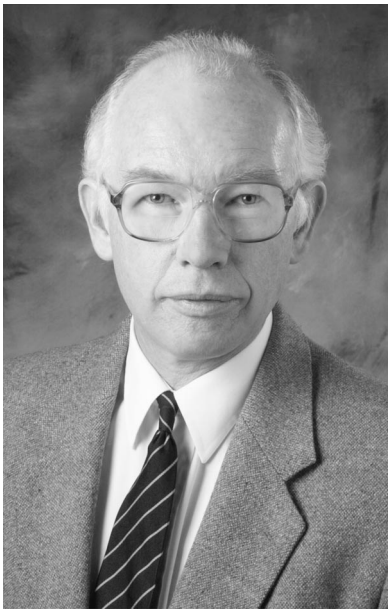
Dr. Robert M. McMeeking, Professor of Mechanical Engineering at the University of California at Santa Barbara, has been appointed Editor of the *Journal of Applied Mechanics* for a five-year term starting on July 1, 2002. He succeeds Dr. Lewis T. Wheeler who has served in this position since January 1993.

After nearly ten years at the helm of the Journal, Lewis Wheeler will step down on June 30, 2002. We thank him for his leadership and service to the international mechanics community in general and to the Applied Mechanics Division of the ASME in particular. Under his leadership the Journal has maintained its position as a premier journal in mechanics. Recent changes enacted including the increase of the page limit on manuscripts from six to nine pages, the switch from quarterly to bimonthly publication, a faster processing time and a new printing process should help make the Journal an even more attractive place to publish.

Robert McMeeking earned a Bachelor of Science degree in Mechanical Engineering from the University of Glasgow and Master of Science and Doctor of Philosophy degrees in Solid Mechanics from Brown University. From 1978 to 1985 he was on the faculty of the Department of Theoretical and Applied Mechanics at the University of Illinois at Urbana-Champaign. Since 1985 he has been a Professor of Mechanical Engineering and Environmental Engineering at UC-Santa Barbara where he has also served two terms as Department Chair. He is a Fellow of the ASME and has served as Associate Editor of the *Journal of Applied Mechanics* for six years.

McMeeking's research interests span the field of solid mechanics with particular emphasis on mechanics of materials and computational aspects. His publications include works in finite deformation plasticity, inelastic fracture, toughening mechanisms in ceramics, bimaterial fracture, powder consolidation, failure of composites, ferroelectrics, etc. Robert McMeeking will bring to the editorship a strong connection with materials science which is an important driver of many modern applied mechanics efforts.

This appointment represents the conclusion of a more than twelve-month search conducted by the Executive Committee in consultation with past editors of the Journal and past AMD chairs. McMeeking was selected from a strong field of candidates. His selection and this five-year appointment were approved by the ASME Publications Committee at the 2001 IMECE meeting. On behalf of the AMD Executive Committee we extend a warm welcome to Robert McMeeking to this most prominent appointment and assure him the support of the Division in his efforts to bring the Journal to even higher levels of excellence.



Stelios Kyriakides
for the AMD Executive Committee

Notice on JAM Correspondence

Effective July 1, 2002, all correspondence concerned with previously submitted articles and all new manuscript submissions should be directed to:

Professor R. M. McMeeking, Editor
Journal of Applied Mechanics
Department of Mechanical and Environmental Engineering
University of California
Santa Barbara, CA 93106

W. Chen
Mem. ASME

F. Lu

Department of Aerospace and Mechanical Engineering,
The University of Arizona,
Tucson, AZ 85721-0119

D. J. Frew
Assoc. Mem. ASME

M. J. Forrestal
Fellow ASME

Sandia National Laboratories,
Albuquerque, NM 87185-1174

Dynamic Compression Testing of Soft Materials

Low-strength and low-impedance materials pose significant challenges in the design of experiments to determine dynamic stress-strain responses. When these materials are tested with a conventional split Hopkinson pressure bar, the specimen will not deform homogeneously and the tests are not valid. To obtain valid data, the shape of the incident pulse and the specimen thickness must be designed such that the specimens are in dynamic equilibrium and deform homogeneously at constant strain rates. In addition, a sensitive transmission bar is required to detect the weak transmitted pulses. Experimental results show that homogeneous deformations at nearly constant strain rates can be achieved in materials with very low impedances, such as a silicone rubber and a polyurethane foam, with the experimental modifications presented in this study.
[DOI: 10.1115/1.1464871]

1 Introduction

Materials with low-strength and low-impedance, such as elastomers and polymeric foams, have found a wide range of applications for shock absorption. Rubbers have traditionally been used for shock-absorbing components; e.g., engine mounts, suspensions, and gaskets in the automotive industry. More recently, elastomers have been used to improve the drop-tolerance of portable electronic products, such as notebook computers and cellular phones. Accidental drops have been the main failure mode for many of these portable products. Polymeric foams have also been used to design crash-resistant components in air and ground vehicles. With tight restrictions on the size and weight faced by designers, it is essential to have accurate material models that predict the dynamic response and energy dissipation of these soft materials under impact. When the dynamic response of an elastomeric or polymeric structure used as a shock absorber needs to be predicted, the finite element analysis method is commonly used as the analysis tool for materials that experience nonlinear deformation and very large strains (Vallee and Shukla [1]). All material models need reliable experimental data to determine the material constants and to check the accuracy of the model over the range of its application. It is therefore essential to understand the mechanical response and failure behavior of soft materials under dynamic loading conditions.

When very large strains and high strain rates are expected in deformation processes, current standard dynamic experimental techniques for testing elastomers and polymeric foam materials are inadequate. Rotating eccentric test machines (Clamroch [2]) are only capable of developing large specimen strains at relatively low frequencies. Vibrating machines (Harris [3]) can often develop sinusoidal or other waveform input at very high frequencies but cannot achieve the desired large sample strains. Ultrasonic wave techniques produce small strains and usually have significant material damping as the waves travel through soft materials. The dynamic destructive Charpy test cannot provide an accurate and complete stress-strain history. The split Hopkinson pressure bar (SHPB), originally developed by Kolsky [4], has been widely used and modified to determine the dynamic properties of a variety of engineering materials, such as metals (Gray [5]), concrete (Ross et al., [6]), and ceramics (Chen and Ravichandran [7–9]). Walley et al. [10] summarized conventional SHPB tests for a range of hard polymers. Instead of the single data point obtained with the Charpy test, SHPB experiments provide complete dynamic stress-strain curves as a function of strain rates. However, if the specimen is a soft material with a low mechanical impedance, such as silicone rubbers and polymeric foams, the conventional split Hopkinson pressure bar technique needs to be modified before reliable dynamic data can be produced. Because of the drastic impedance mismatch at the soft sample/transmission bar interface, the transmitted signal can be too weak to be accurately measured by the surface strain gages on a steel bar. Viscoelastic bars have been used to obtain a transmitted signal with sufficiently high signal-to-noise ratio (Gamby and Chaoufi [11], Wang et al. [12], Zhao et al. [13], and Sawas et al. [14]). However, dispersion and attenuation corrections in data reduction and the dependence of bar material on temperature, moisture level, and aging factors have brought uncertainties into the data from viscoelastic bar experiments. Furthermore, it takes longer time for the specimen to achieve dynamic equilibrium when sandwiched between two low-impedance bars (Frew et al. [15]). To avoid these uncertainties, magnesium bars (Gray and Blumenthal [16]), a hollow aluminum transmission bar (Chen et al. [17]), and a quartz-crystal-embedded aluminum transmission bar (Chen et al. [18]) have been developed and validated.

The effects of specimen thickness for low-impedance specimens must also be thoroughly investigated because stress waves attenuate when traveling through the soft specimens. A thick specimen acts as a shock absorber instead of a representative volume for material properties. In his original work, Kolsky [4] pointed out that a thick specimen would invalidate the assumption that the axial stresses on both sides of the specimen were nearly equal. Kolsky then studied the thickness effects on the dynamic compressive stress-strain behavior of polythene as he varied the specimen thickness from 2.68 mm to 0.25 mm. The peak strength of the thickest specimen was only one-fifth that of the thinnest one. Recently, Dioh et al. [19] pointed out that specimen thickness was an important parameter when the SHPB technique was used to test glassy polymers. Gray et al. [16,20] performed SHPB experiments with several thicknesses of Adiprene L-100 rubber and showed that dynamic equilibrium was achieved during later stages of the experiments for the thinner samples.

We used pulse shaping techniques to obtain homogeneous de-

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, April 5, 2001; final revision, November 14, 2001. Editor: K. T. Ramesh. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston. TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

formations at a nearly constant strain rate for the soft specimen. Some of the advantages and necessities for shaping the incident pulse for split Hopkinson pressure bar experiments were discussed 20 years ago. Frantz et al. [21] and Follansbee [22] wrote review papers that discussed pulse shaping for SHPB experiments with metal samples. In these review papers, the authors emphasized that a slowly rising incident pulse is preferred to a pulse that rises steeply in order to minimize the effects of dispersion and allow the sample to achieve dynamic stress equilibrium. Frantz et al. [21] and Follansbee [22] discuss experimental techniques for pulse shaping and a numerical procedure [23] for correcting raw data for wave dispersion in the bars. To shape the incident pulse these authors [21,22] machined a large radius on the impact face of the striker bar or placed a tip material between the striker and incident bars. The tip material or pulse shaper was a disk slightly larger than the bars and 0.1 to 2.0 mm thick. The pulse shaper materials were paper, aluminum, brass, or stainless steel. Frantz et al. [21] present experimental results that show the advantages of pulse shaping for a 3041 stainless steel sample at an approximate strain rate of 4,500 s^{-1}. In addition, these authors [21] show that a properly chosen tip material or pulse shaper can also be used to generate a nearly constant strain rate in the sample. Gray [5] and Gray and Blumenthal [16] present additional information in recent survey papers that include these subjects. However, Duffy et al. [24] were probably the first authors to use pulse shapers to smooth pulses generated by explosive loading for the torsional split Hopkinson bar. A more recent and complete discussion on pulse shaping is given by the authors (Frew et al. [25]).

In this paper, limitations of the SHPB technique for testing soft materials are illustrated, and necessary modifications are presented to obtain valid stress-strain data. In particular, the thickness of RTV 630[1] silicone rubber specimens is systematically varied to determine the thickness effects. In addition, a pulse shaping technique is used to ensure a homogeneous deformation at a nearly constant strain rate in the soft specimen. It was found that a proper pulse shaper and a reduced thickness must both be carefully determined to achieve a dynamic equilibrium state of stress in the specimen and obtain a nearly constant strain rate. To show our techniques are reliable for extremely low-impedance materials, we also present dynamic stress-strain data for a low density (290 kg/m^3), rigid, and closed-cell polyurethane foam.

2 A Conventional Split Hopkinson Pressure Bar (SHPB) Experiment With an RTV Silicone Rubber

The working principles of the SHPB technique are well documented (Kolsky [4], Lindholm [26], and Gray [5]). The length-to-diameter ratios (aspect ratios) of the cylindrical specimens are typically 0.5–1.0 in order to minimize the inertial and end frictional effects. However, when the specimen is a low-strength and low-impedance material such as a silicone rubber or a polymeric foam, the limitations of the SHPB technique must be recognized before valid data can be obtained. For testing RTV 630 silicone rubber specimens, we used an aluminum SHPB apparatus (Chen and Zhang [27]) to capture the weak strain signals transmitted from the low-impedance specimens. As schematically shown in Fig. 1, the 7075-T6 aluminum elastic bars had 19-mm diameters. The incident, transmission, and striker bars were 1802, 762, and 305 mm long, respectively. The strain gages shown in Fig. 1 were located at 560 mm from the impact surface on the incident bar and 203 mm from the sample/bar interface on the transmission bar. The RTV 630 rubber specimens had a diameter and length of 12.7 mm with an aspect ratio of 1.0. As previously discussed, other investigators, such as Gray et al. [16,20], have shown that thinner samples are required for valid tests on some rubber-like materials. However, we use a sample with an aspect ratio of 1.0 such that high-speed digital photographs clearly show the sample deformation, which illustrates the problems associated with SHPB tests on soft materials.

According to the conventional SHPB theory (Gray [5]), the nominal strain rate $\dot{\varepsilon}(t)$ and nominal stress $\sigma(t)$ in the specimen are given by

$$\dot{\varepsilon}(t) = -\frac{2c_0}{L}\varepsilon_r(t), \tag{1}$$

$$\sigma(t) = \frac{A_0}{A_s}E\varepsilon_t(t), \tag{2}$$

where L and A_s are the original sample length and cross-sectional area; c_0, E, and A_0 are the elastic bar-wave velocity, Young's modulus, and cross-sectional area of the bars; $\varepsilon_r(t)$ and $\varepsilon_t(t)$ are measured reflected and transmitted strain signals on the bar surfaces. Figure 2 shows oscilloscope records of the incident, reflected, and transmitted strain signals for an experiment where the striker bar impacted the incident bar at 31 m/s. Equation (1) and the nearly constant amplitude of the reflected signal suggest that the dynamic deformation of the specimen was at a nearly constant strain rate. Equation (2) and the very low amplitude of the transmitted signal suggest a very low average stress in the specimen. However, Eqs. (1) and (2) are valid only if the sample is in dynamic stress equilibrium and undergoes a homogeneous deformation (Gray and Blumenthal [16]). We next show that the dynamic deformation process of the sample during this SHPB experiment does not undergo homogeneous deformation. Thus, conclusions drawn from the strain signals and Eqs. (1) and (2) are not valid.

As shown in Fig. 1, a high-speed digital Imacon 468 camera was focused on the rubber test section of the SHPB setup. Figure 3 shows the high-speed camera record of the dynamic deformation process of the rubber sample during the SHPB experiment. The first image was taken before the loading pulse reached the specimen. The next seven images were taken at 10, 40, 70, 100, 130, 160, and 190 μs after the arrival of the incident pulse at the specimen. In each image, the grids are used as the background for reference. The dark rod on the right is the incident bar, which moves toward the left during loading (due to the optical system

[1]GE Silicone, 260 Hudson River Road, Waterford, NY, 12188.

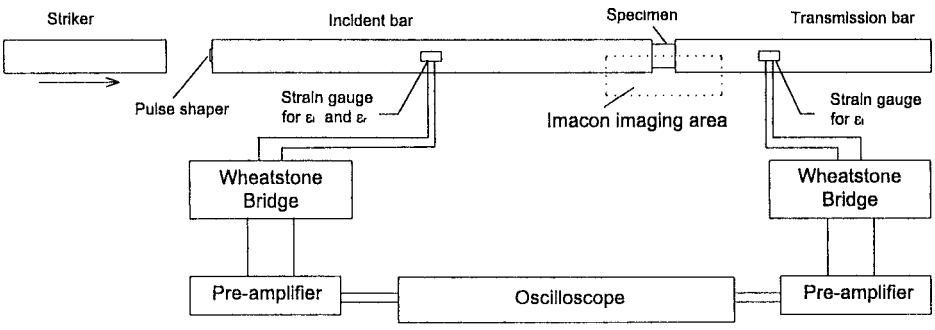


Fig. 1 A schematic illustration of a split Hopkinson pressure bar setup

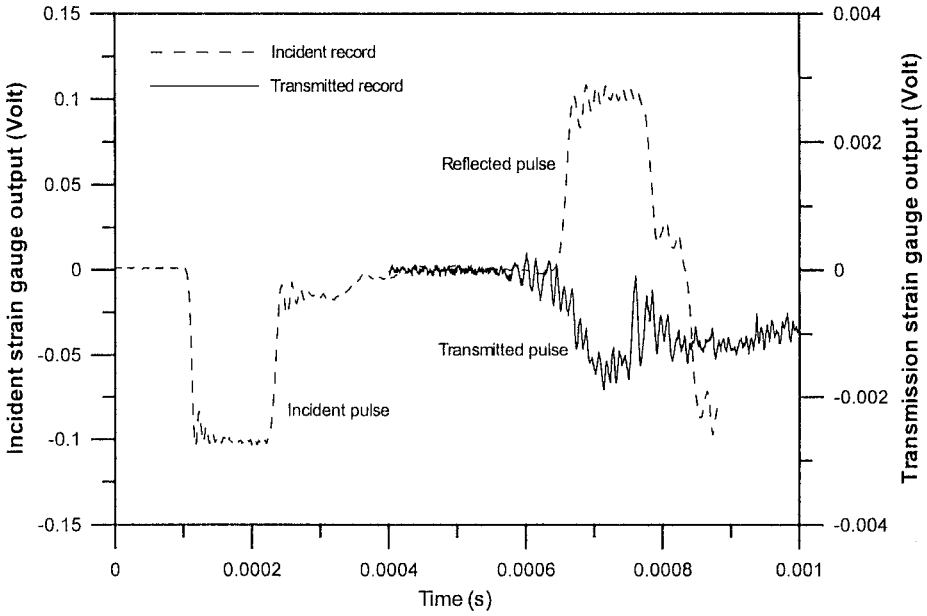


Fig. 2 Oscilloscope records of a SHPB experiment on a silicone rubber

setup, the moving direction appears to be opposite of that illustrated in Fig. 1). The aluminum bars are covered by dark tape to avoid strong glare on the images. The dark rod on the left is the transmission bar. The light-colored specimen is sandwiched between the dark bar-ends. The series of images in Fig. 3 show that, after impact, significant deformation occurs near the impact-end of the specimen as evidenced by the large lateral deformation near the incident bar. Meanwhile, the other end of the specimen has no visible deformation. The large, localized deformation then propagates toward the transmission bar and, when reaching the transmission bar/specimen interface, is reflected back into the specimen. During the entire experiment, the transmission bar/specimen interface nearly remained at its original position, which indicates that a very small axial force was transmitted into the transmission bar even though the specimen deformed significantly. The small transmitted signal shown in Fig. 2 is also consistent with this observation. However, the specimen recovered after the impact loading shows significant damage (Fig. 4), even though the stress in the specimen calculated from (2) and the weak transmitted signal shown in Fig. 2 is very small.

The results shown in Figs. 2–4 indicate two distinct characteristics in dynamic compression experiments involving elastomeric specimens with the conventional SHPB technique: (1) nonequilibrium stress state and (2) nonhomogeneous deformation and nonconstant strain rate in the specimen. The propagation of the deformation wave in the specimen revealed in Fig. 3 clearly indicates a

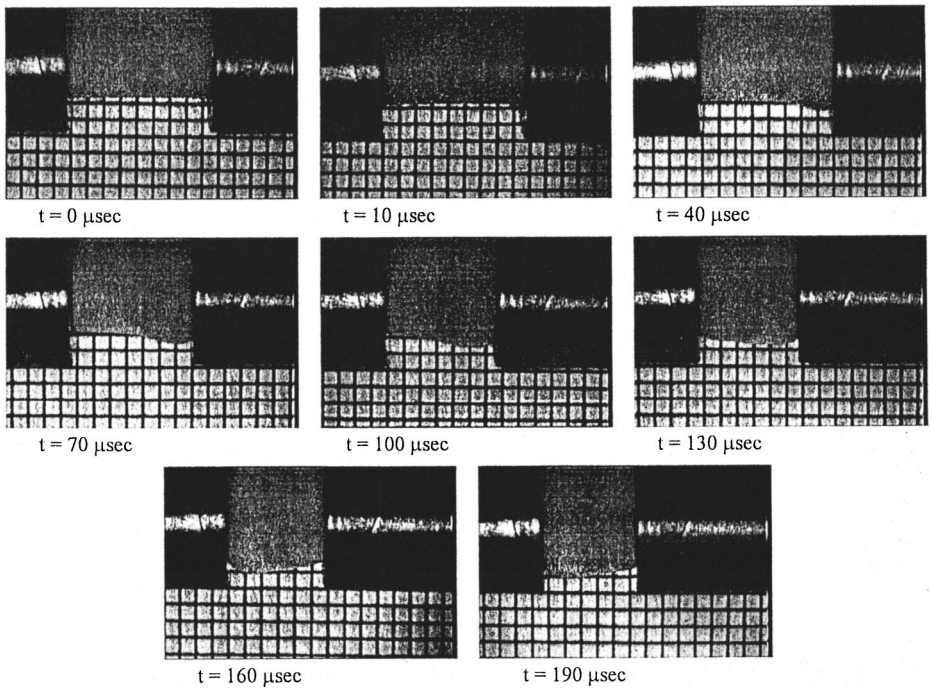


Fig. 3 Sequential images of a rubber deformation during a SHPB test

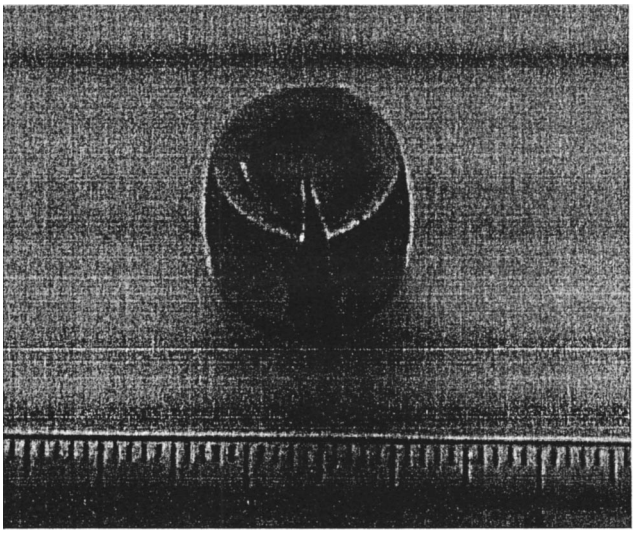


Fig. 4 An RTV 630 silicone rubber specimen recovered after a SHPB test (the scale is in centimeters)

nonhomogeneous deformation state and thus a nonequilibrium stress state over the entire duration of the experiment. Thus, the stress and strain measured through incident and transmission bars cannot be correlated to any unique stress-strain state in the specimen due to the nonuniform deformation. The nonequilibrium stress state is also demonstrated by the fact that the specimen is extensively damaged (Fig. 4), even though only a small load is transmitted through the specimen as shown by the small amplitude transmitted signal in Fig. 2 and by the nearly stationary position of the transmission bar in Fig. 3. Even though the reflected signal in Fig. 2 suggests that a constant strain rate has been reached in the specimen based on classical SHPB theory, the images in Fig. 3 show that the strain rate cannot be constant over the entire specimen.

The conventional SHPB has been an effective tool for investigating the dynamic flow behavior of ductile metals. However, when the specimen is a low-impedance and low-strength material, the results shown in Figs. 2–4 show that the reflected signal is not proportional to strain rate nor is the transmitted signal proportional to specimen stress. Thus, the reflected and transmitted pulses that are commonly used to obtain dynamic stress and strain histories in the specimen with (1) and (2) are both not valid for the experiment described in this section. Therefore, the limitations of SHPB in testing soft materials should be recognized and remedied before reliable results can be obtained. First, the images in Fig. 3 indicate that the rubber specimen is too thick to have a homogeneous deformation during dynamic compression, which is consistent with the conclusions by previous investigations [4,16,19]. The wave speeds in the rubbers are very low compared to those in metals and ceramics. A reduced specimen thickness is therefore necessary. Second, the rise time of the incident (loading) pulse is too short. The rise time in a conventional SHPB is typically about 10 μs, which is too short for the stress waves to travel back and forth inside the soft specimen more than three times to reach a dynamic stress equilibrium (Gray [5], Chen et al. [28], and Ravichandran and Subhash [29]). The shape of the loading pulse profile must be carefully controlled to facilitate an equilibrium state of stress and a homogeneous deformation at a nearly constant strain rate.

3 Modified SHPB Experiments With RTV 630 Silicone Rubber

In this section, we present modifications to the conventional SHPB technique for the valid compression testing of soft materials. We show that RTV 630 silicone rubber samples are in dynamic stress equilibrium and have nearly constant strain rates over most of the test durations.

3.1 Dynamic Stress Equilibrium. To check dynamic stress equilibrium directly, 0.25-mm thick circular piezoelectric transducers[2] (X-cut quartz crystal disks of the same diameter as the bars) were attached at both ends of the specimen to determine the axial load histories on the front and back surfaces of the specimen. The mechanical impedance of the self-generating quartz crystal transducer is very close to the mechanical impedance of the aluminum bars, which ensures that the introduction of the quartz disks does not affect the one-dimensional wave propagation in the bars (Chen et al. [18]). Quartz-crystal force transducers have been used by previous investigators to measure dynamic force profiles (Karnes and Ripperger [30], Wasley et al. [31], and Togami et al. [32]). To prevent the large lateral expansion of the soft specimen during axial compression from damaging the brittle quartz crystal, a thin aluminum disk the same diameter as the quartz disk was placed between the quartz crystal and the specimen. The aluminum disk also serves as an electrode to collect charges from the quartz crystal transducer. The modified experimental setup for recording stresses is schematically shown in Fig. 5. The quartz-crystal transducer signals and the strain gage signals from incident and transmission bars were recorded using a Tektronix TDS 420A digital storage oscilloscope through ADA400A differential amplifiers and Kistler 5010B charge amplifiers.

Dynamic stress equilibrium can also be checked by comparing the transmitted signal (1-wave) and the difference between the incident and reflected signals (2-wave) (Follansbee and Frantz

[2]Valpey-Fisher Corp., 1994, *The User's Guide to Ultrasound Products*, 75 South St., Hopkinton, MA 01748.

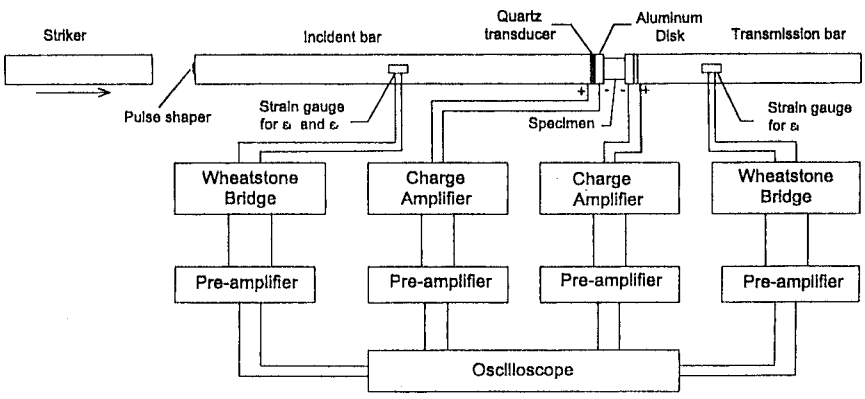


Fig. 5 A schematic illustration of a modified SHPB setup for stress equilibrium

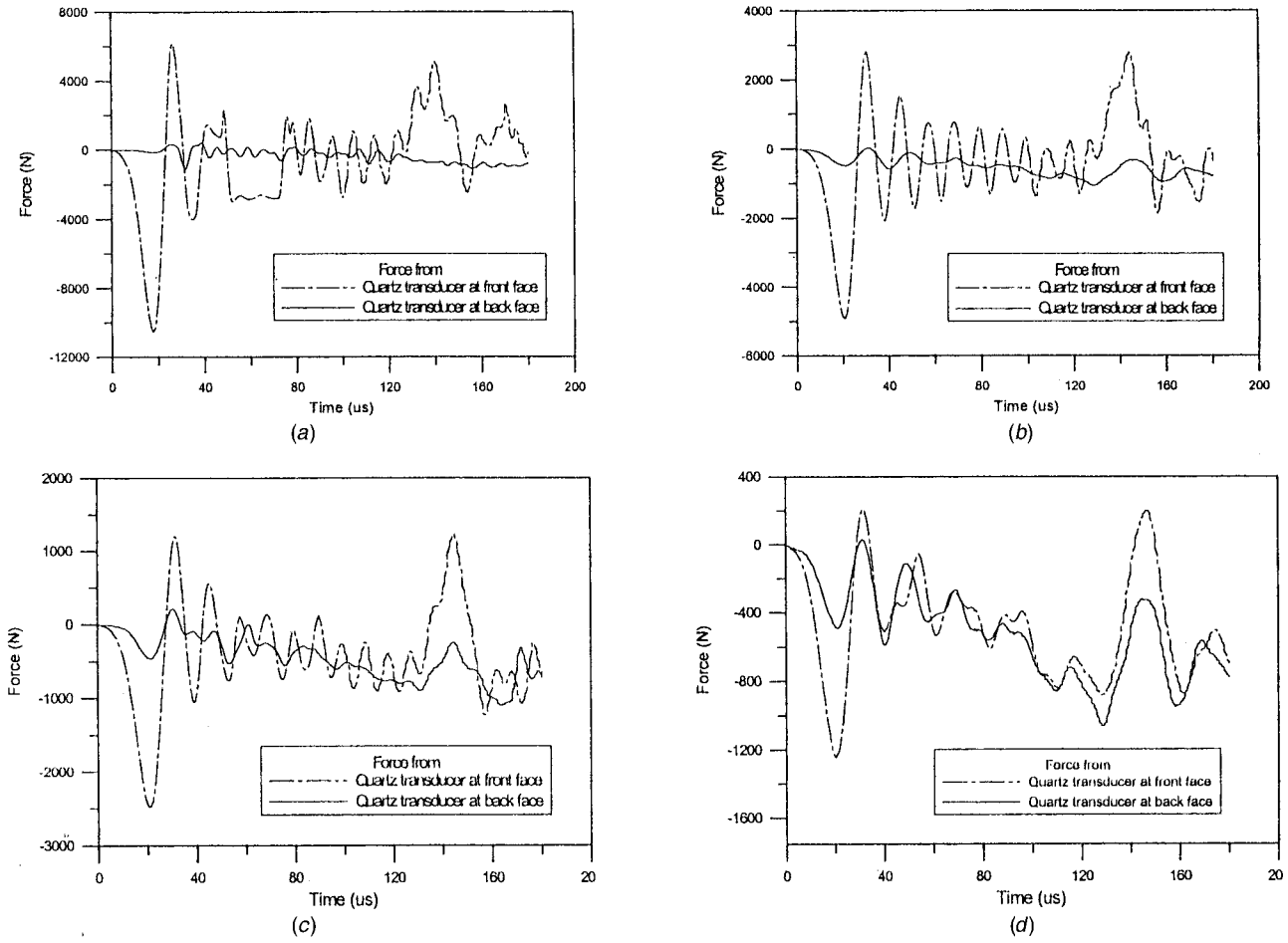


Fig. 6 Front and back-end force histories during a SHPB test without pulse shaping for various specimen thicknesses: (*a*) 12.7 mm, (*b*) 6.57 mm, (*c*) 3.08 mm, and (*d*) 1.53 mm

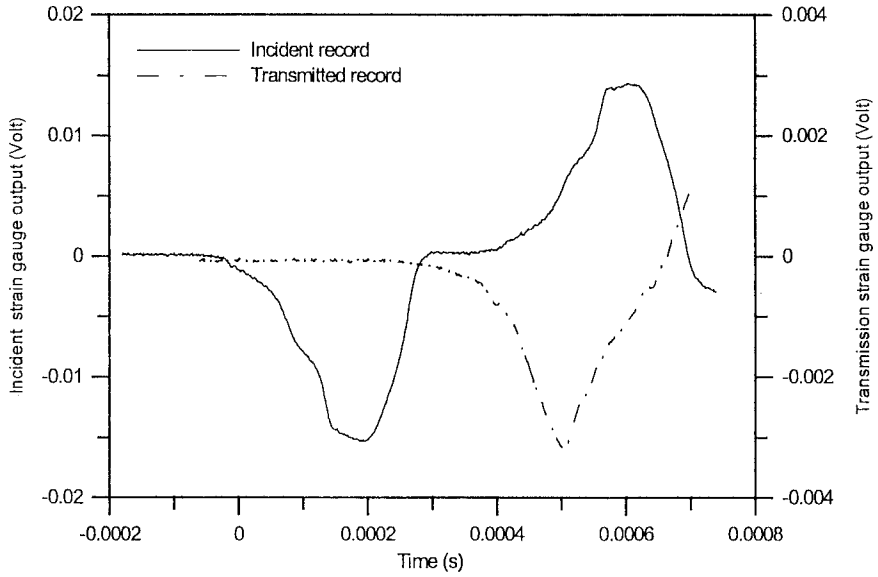


Fig. 7 A typical oscilloscope record of a dynamic compressive experiment on RTV 630 with a pulse shaper on an aluminum SHPB

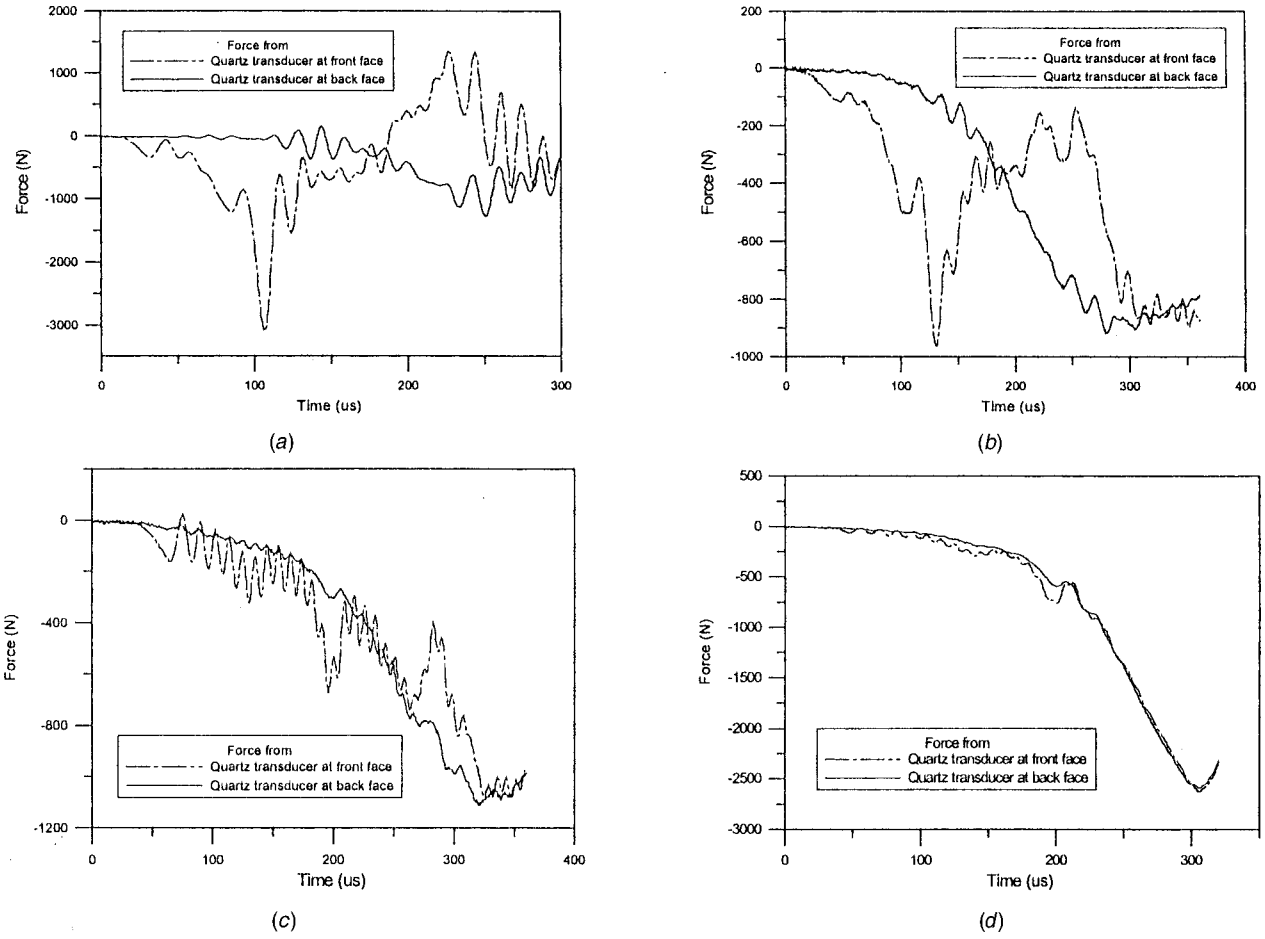


Fig. 8 Front and back-end force histories during a SHPB test with pulse shaping for various specimen thicknesses: (*a*) 12.7 mm, (*b*) 6.58 mm, (*c*) 3.06 mm, and (*d*) 1.53 mm

[23], Wu and Gorham [33], and Gray et al. [16,20]). The quartz-crystal method is a more direct measurement of the end forces on the specimen. The method is also more effective when the specimen is soft. The incident and reflected signals are nearly the same in this case, which makes 2-wave calculations very inaccurate.

3.2 Sample Thickness Effect. To determine the effect of specimen thickness on the dynamic stress-strain behavior of the silicone rubber during a SHPB experiment, we conducted experiments with 12.7-mm diameter samples that had thicknesses of 12.7, 6.58, 3.06, and 1.53 mm. The striker initial velocity was adjusted according to the specimen thickness in an effort to maintain the same nominal strain rate. To minimize the effects of friction on specimen/bar interfaces, the aluminum bar end faces were lapped, and a thin layer of Vaseline nursery jelly was applied to the interfaces. Figure 6(*a*) shows the dynamic force histories on the front and back surfaces of the specimen as recorded by the two embedded quartz-crystal force transducers. The results in Fig. 6(*a*) show that the specimen with thickness 12.7 mm (aspect ratio of 1.0) was never in a state of dynamic stress equilibrium during the entire 150-μs loading period. The large spike near the beginning of front-end force history was at least an order of magnitude larger than the maximum amplitude in the back-end force history. This probably caused failure of the specimen that started at the front end. This large amplitude of front-end force never propagated to the back-end where the force was significantly smaller. A 1-wave and 2-wave analysis (Gray and Blumenthal [16]) also produced similar results. When the specimen thickness was reduced to 6.58 mm, the force history profiles were very similar (Fig. 6(*b*)), and the desired dynamic stress equilibrium was never reached. Again, the large spike in the front-end force may have initiated the failure in the specimen. When the thickness was reduced to 3.06 mm, the oscillations in the force histories were getting close to being in-phase, but the amplitude of the front-end force was still significantly higher (Fig. 6(*c*)). When the specimen thickness was further reduced to 1.53 mm, no significant improvement in stress equilibrium was observed (Fig. 6(*d*)). Since the desired stress equilibrium was never reached as the aspect ratio of the specimens was reduced from 1.0 to 0.12, we concluded that a state of dynamic stress equilibrium in the rubber specimens cannot be obtained by only reducing the specimen thickness. Next, we show that the incident pulse needs to be controlled to avoid the sudden impact that results in the large spike in the front-end force.

3.3 Pulse Shaping. As shown in Fig. 2, the shape of the incident pulse in a conventional SHPB experiment is nearly trapezoidal with a rise time of about 10 μs. The results shown in Fig. 6 indicate that the nonequilibrium stress state or the nonhomogeneous deformation in the specimen shown in Fig. 3 cannot be overcome by a reduction in specimen thickness alone due to the sharp rise of the loading pulse. The shape of the loading pulse must be controlled so that its rise time is longer than the stress equilibrating time in the specimen. In this way, significant deformation/failure in the specimen occurs only after dynamic equilibrium has been reached. Pulse-shaping techniques have recently been further developed to control the loading pulse (Ravichandran and Chen [34], Nemat-Nasser et al. [35], Togami et al. [32], and Frew et al. [15,25]). In this study, the control of rise time was achieved by attaching a combination of C11000 copper ([36]) and plastic disks, called a pulse shaper, at the impact end of the

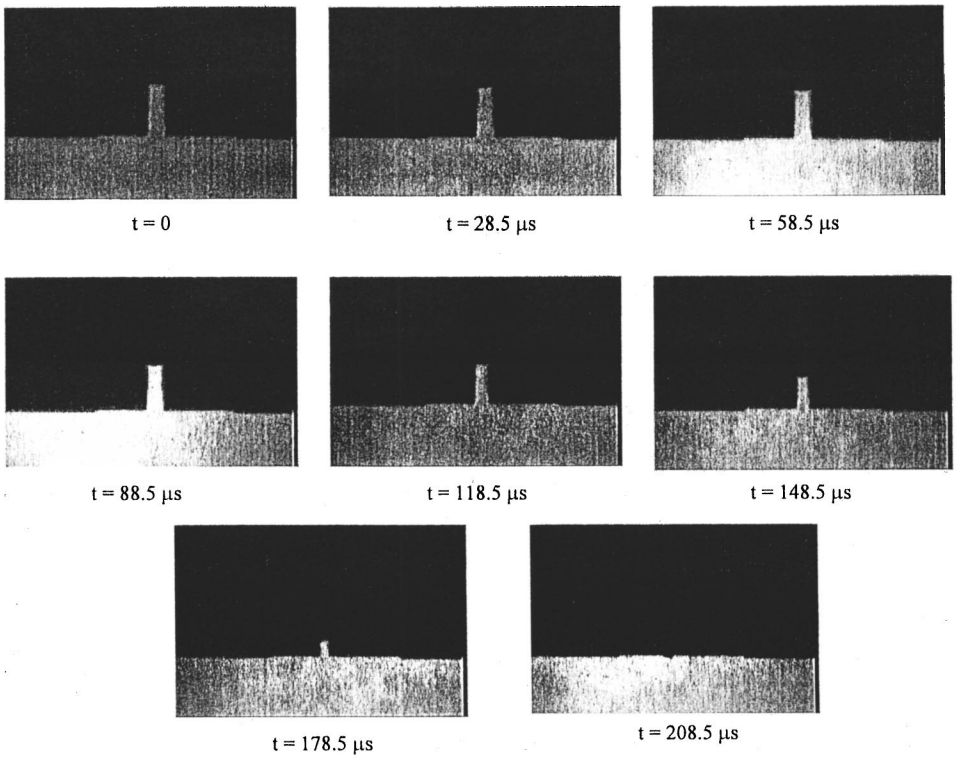


Fig. 9 Sequential images of a thin rubber specimen deforming during a SHPB test

incident bar. The plastic deformation of the pulse shaper upon impact effectively increases the rise time of the incident pulse in the bar. The amplitude and duration of an incident pulse are controlled by varying the pulse shaper and the striker bar velocity and length. An ideal incident pulse must produce dynamic sample equilibrium and a nearly constant reflected pulse for constant strain rate. If the specimen is linearly elastic, ramp incident pulses of different slopes will generate constant strain rates in the specimen (Frew et al. [15,25]). If the specimen is not linearly elastic, as in most cases with rubber and foam materials, the shape of the desired incident pulse depends on the shape of the transmitted pulse. This situation requires iterations in experiments to approach constant strain rates through precise control of the profiles of the incident pulses. Frew et al. [25] have developed an analytical model to predict the incident pulse in terms of the dimensions of a single-disk pulse-shaper, the length and the material of the striker, and the striking velocity. This model provides an effective guidance in the dynamic experimental iterations to approach valid test conditions. A typical oscilloscope record of a pulse-shaped SHPB experiment is shown in Fig. 7.

The four specimen thicknesses of 12.7, 6.58, 3.06, and 1.53 mm were used again for the pulse-shaped experiments with RTV 630 silicone rubber. Figure 8 shows the dynamic force histories on the front and back surfaces of the specimens as recorded by the two embedded quartz-crystal force transducers. The results in Fig. 8(*a*) show that, even with a pulse shaper, the specimen with an aspect ratio of 1.0 never reaches a state of dynamic stress equilibrium. This is consistent with the intrinsic behavior of rubber; i.e., a thick piece of rubber will act as a shock absorber. There is again a large spike in the front-end force history (Fig. 8(*a*)), which may have caused the failure of specimens, starting at the front end. This large amplitude of front-end force never propagated to the back end where the force was significantly smaller, which is consistent with the function of a shock absorber. When the specimen thickness was reduced to 6.35 mm, the force history profiles were very similar to the thicker sample case, although the back-end force had a much higher amplitude during the later stages of the experiment, whereas the amplitude of the spike from the front end reduces significantly (Fig. 8(*b*)), as compared to the 12.7-mm thick specimen. The specimen did not fail due to the reduced loading amplitude. When the thickness was reduced to 3.06 mm, both of the force history profiles follow the same trend qualitatively (Fig. 8(*c*)). However, significant oscillations existed in the front-end force history. These oscillations could not propagate through the thickness of the rubber specimen due to the dispersive nature of the rubber material. Therefore, the stress state and deformation in the specimen were still not uniform or homogeneous. When the specimen thickness was further reduced to 1.53 mm, the oscillations in the front-end force history were much smaller in amplitude and the front-end force history nearly agreed with the back-end force history (Fig. 8(*d*)). Therefore, at this thickness, the specimen did not absorb any significant portion of the loading pulse. Rather, the entire specimen was in a nearly stress-equilibrium state. The mechanical response averaging over the specimen volume represents the pointwise material behavior of the silicone rubber under impact. Thus, for this case we obtain valid stress-strain data.

The results shown in Figs. 6 and 8 indicate that, to achieve a dynamic homogeneous deformation in the soft specimen with SHPB testing, pulse shaping must be employed. In addition, a thin specimen must be used. However, reducing the specimen thickness alone will not facilitate a dynamic stress equilibrium and uniform deformation. On the other hand, pulse shaping alone on thick specimens will not result in an equilibrium state of stress either. It should be noted that, when the thickness of the 12.7-mm diameter specimen is reduced to 1.53 mm as required for axial stress equilibrium, the friction on the interfaces between the specimen and the bars becomes a concern. To minimize the friction effects in this study, the bar ends were lapped and a thin layer of Vaseline nursery jelly was applied between the specimen and the bar ends, as mentioned previously. Using such smooth bar end-faces and proper lubrication in the dynamic experiments, the thin specimen expands in the radial directions nearly freely as shown in the high-speed digital images in Fig. 9, which were taken using

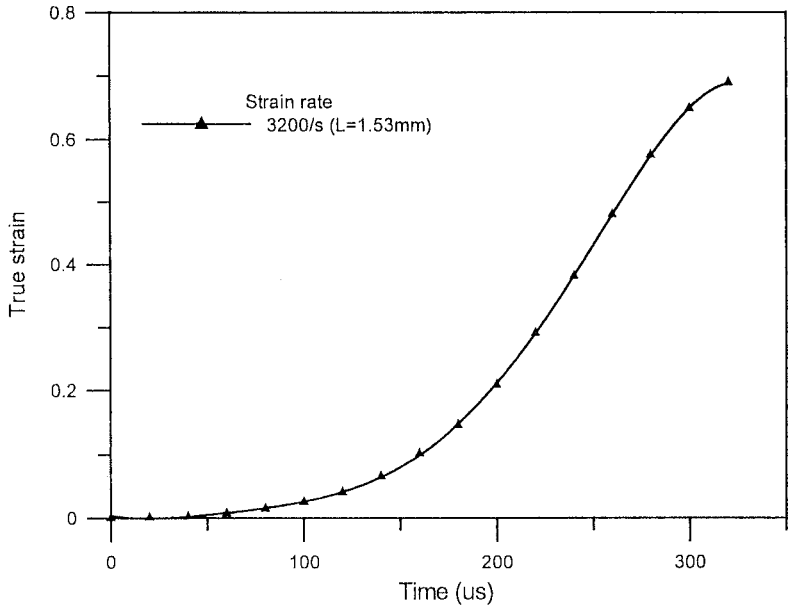


Fig. 10 A typical dynamic strain history of an RTV 630 specimen

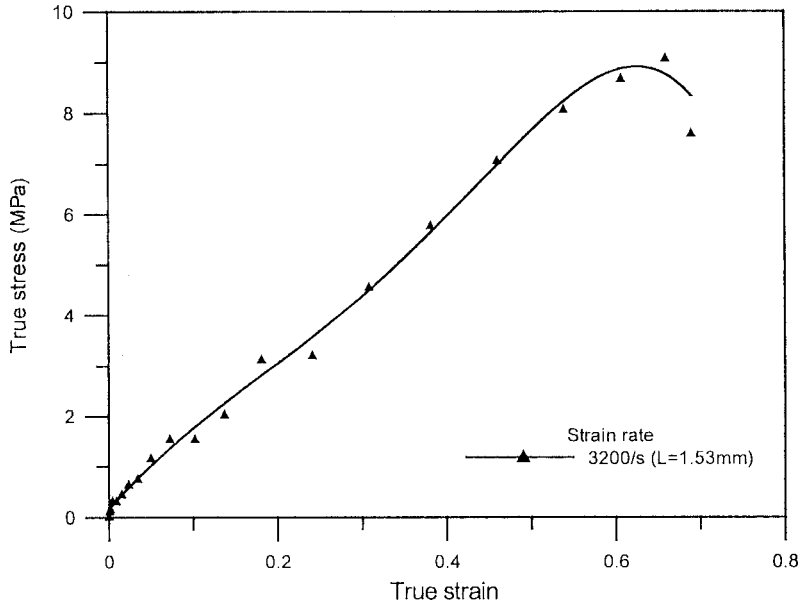


Fig. 11 Dynamic compressive stress-strain curve of an RTV 630 specimen

back-light illumination condition to show the specimen edge more clearly. Figure 9 also show that the specimen deformed uniformly.

When the axial forces are nearly the same on both end-faces of a 1.53-mm thick specimen, the axial stress in the specimen is expected to be nearly uniform over the 1.53 mm thickness. Therefore, the deformation is also expected to be uniform in the specimen as verified by the high-speed photographs shown in Fig. 9. The basic assumptions of SHPB are thus verified in soft material testing by the careful employment of a pulse shaper and a thin specimen. These verified assumptions are: the specimen undergoes homogeneous deformation; the bars remain elastic; and the specimen/bar interfaces remain flat. The last two assumptions are easily satisfied since the aluminum alloy bars are much stronger and more rigid than the elastomeric specimens.

When all the basic assumptions are satisfied, strain rate in the specimen is proportional to the reflected signal shown in Fig. 7 and is given by (1). A constant amplitude of the reflected pulse corresponds to a constant strain rate in the soft specimen. The pulse shaper should be designed such that an equilibrium stress state is reached, as shown in Fig. 8(*d*), and the specimen is deforming at a near constant strain rate over a major portion of the experiment duration, especially during the late stages of the experiment when the loading levels are high, as shown in Fig. 7. The pulse shaper for this experiment consisted of a hardened C11000 copper disk with a 5.56 mm diameter and 2.42 mm thickness, another hardened C11000 copper disk with a 1.64 mm diameter and 1.62 mm thickness, and a layer of silicone rubber with a 1.64 mm diameter and 1.51 mm thickness at the impact end.

After all the basic assumptions are validated, the experimental data can be reduced using conventional SHPB methods (Gray [15]). Figure 10 shows a typical dynamic strain history for the 1.53-mm thick RTV 630 specimen. The original oscilloscope records of the SHPB experiment on this specimen are shown in Fig. 7. The fact that the strain increases almost linearly with in-

creasing time during the majority of the experiment's duration indicates that a nearly dynamic constant strain rate has been achieved. The slope is taken as the strain rate of the experiment. On the other hand, if a pulse shaper is not used, dynamic stress equilibrium cannot be reached which results in an invalid experiment. Figure 11 shows the dynamic compressive stress-strain curve of the RTV 630 specimen at a strain rate of 3200/s. This stress-strain curve is considered to represent the realistic dynamic compressive behavior of this material because dynamic stress equilibrium and homogeneous deformation at a nearly constant strain rate have been achieved during the experiment, as shown in Figs. 7, 8(*d*), 9, and 10. It was also noted that the specimen did not fail due to the absence of an excessively large amplitude spike in the front-end force history.

4 A Modified SHPB Experiment With a Low-Density Polyurethane Foam

Besides elastomers, the SHPB technique with pulse shaping and a thin specimen can also be used to determine the dynamic compressive behavior of other soft materials. For example, Fig. 12 shows the oscilloscope record of a SHPB test on a polyurethane foam specimen[3] with a density of 290 kg/m^3. The rigid closed-cell specimen had a 12.7 mm diameter and 5.21 mm thickness. The thickness is more than ten times the average cell size in the foam. After iterations, the pulse shaper to achieve dynamic equilibrium and a nearly constant strain rate consisted of a hardened C11000 copper disk with a 6.42 mm diameter and 2.42 mm thickness, and

[3]General Plastics Manufacturing Co., Tacoma, WA 98409

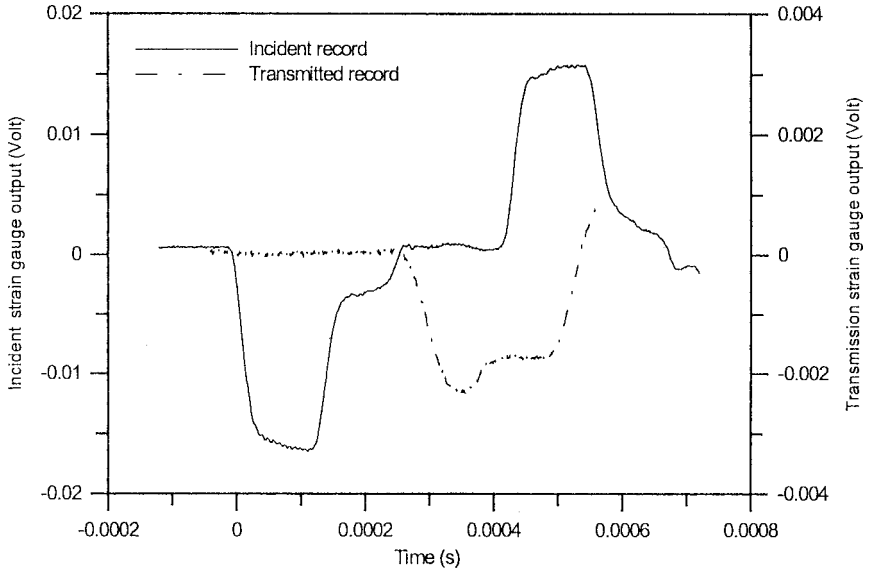


Fig. 12 A typical oscilloscope record of a dynamic compressive experiment on a polyurethane specimen with a pulse shaper on an aluminum SHPB

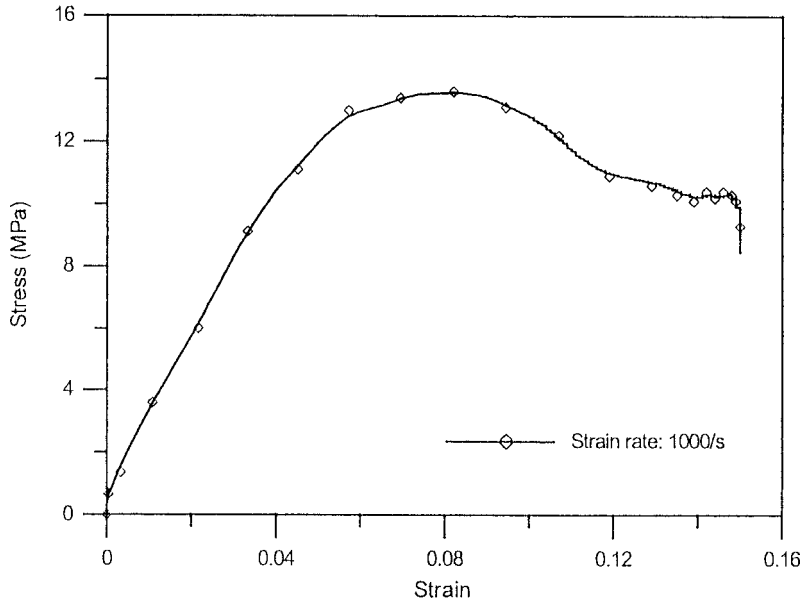


Fig. 13 Dynamic compressive stress-strain curve of a polyurethane specimen

a silicone rubber cube with a 1.51-mm side length at the impact end. Following the procedure in Fig. 8(*d*), a check on the front-end and back-end force histories indicates that dynamic stress equilibrium has been reached in the experiment. The nearly constant amplitude in the reflected signal in Fig. 12 indicates a constant strain rate in the specimen during the experiment. Figure 13 shows the dynamic compressive stress-strain curve of the foam at a strain rate of 1000/s, which is reduced from the oscilloscope records shown in Fig. 12. Unlike elastomeric materials, the volume of the foam material does not remain constant during compression. In fact, the Poisson's ratio is nearly zero during the initial stages of the dynamic compression. Therefore, the stress-strain curves in Fig. 13 are in engineering stress and strain.

5 Conclusions

Specimen thickness effects on the dynamic compressive behavior of an elastomer, RTV 630 silicone rubber, during split Hopkinson pressure bar (SHPB) testing have been experimentally investigated. To examine the dynamic stress equilibrium in the soft specimen, quartz-crystal force transducers were installed at the specimen/aluminum bar interfaces to directly record the axial force histories on the two ends of the specimen. Results show that the conventional SHPB experiment with a soft specimen results in invalid data due to nonhomogeneous deformation during dynamic compression. Reducing the thickness alone cannot achieve dynamic stress equilibrium and homogeneous deformation in the soft specimen. The loading pulse profile must also be carefully shaped using a pulse-shaping technique to ensure a homogeneous deformation at a constant strain rate in thin specimens during dynamic experiments. The techniques presented for dynamic elastomeric material testing are also applicable to other soft materials, such as polymeric foams.

Acknowledgments

This work was supported by the U.S. Army Research Office through a grant to The University of Arizona (G-DAAD19-00-1-0493) and Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-ACO4-94AL8500. D. J. Frew was employed by the U.S. Army Engineer Research and Development Center at the Waterways Experiment Station in Vicksburg, MS, during the course of this study.

References

[1] Vallee, G. E., and Shukla, A., 1996, ''A Study of the Dynamic Behavior of Elastomeric Materials Using Finite Elements,'' ASME J. Eng. Mater. Technol., **118**, pp. 503–508.
[2] Clamroth, R., 1981, ''Determination of Viscoelastic Properties by Dynamic Testing,'' Polym. Test., **2**, pp. 263–286.
[3] Harris, J. A., 1987, ''Dynamic Testing Under Nonsinusoidal Conditions and the Consequences of Nonlinearity for Service Performance,'' Rubber Chem. Technol., **60**, pp. 870–887.
[4] Kolsky, H. 1949, ''An Investigation of the Mechanical Properties of Materials at Very High Rates of Loading,'' Proc. R. Soc. London, Ser. B, **B62**, pp. 676–700.
[5] Gray, G. T., 2000, ''Classic Split-Hopkinson Pressure Bar Testing,'' *Mechanical Testing and Evaluation*, Metals Handbook, American Society for Metals, Materials Park, Ohio, **8**, pp. 462–476.
[6] Ross, C. A., Jerome, D. M., Tedesco, J. W., and Hughes, M. L., 1996, ''Moisture and Strain Rate Effects on Concrete Strength,'' ACI Mater. J., **93**, pp. 293–300.
[7] Chen, W., and Ravichandran, G., 1996, ''Static and Dynamic Compressive Behavior of Aluminum Nitride under Moderate Confinement,'' J. Am. Ceram. Soc., **79**(3), pp. 579–584.
[8] Chen, W., and Ravichandran, G., 1997, ''Dynamic Compressive Behavior of a Glass Ceramic Under Lateral Confinement,'' J. Mech. Phys. Solids, **45**, pp. 1303–1328.
[9] Chen, W., and Ravichandran, G., 2000, ''Failure Mode Transition in Ceramics Under Dynamic Multiaxial Compression,'' Int. J. Fract., **101**, pp. 141–159.
[10] Walley, S. M., Field, J. E., Pope, P. H., and Safford, N. A., 1989, ''A Study of the Rapid Deformation Behavior of a Range of Polymers,'' Philos. Trans. R. Soc. London, Ser. A, **A328**, pp. 1–33.
[11] Gamby, D., and Chaoufi, J., 1991, ''Asymptotic Analysis of Wave Propagation in a Finite Viscoplastic Bar,'' Acta Mech., **87**, pp. 163–178.
[12] Wang, L., Labibes, K., Azari, Z., and Pluvinage, G., 1994, ''Generalization of Split Hopkinson Bar Technique to Use Viscoelastic Bars,'' Int. J. Impact Eng., **15**, pp. 669–686.
[13] Zhao, H., Gary, G., and Klepaczko, J. R., 1997, ''On the Use of a Viscoelastic Split Hopkinson Pressure Bar,'' Int. J. Impact Eng., **19**, pp. 319–330.
[14] Sawas, O., Brar, N. S., and Brockman, R. A., 1998, ''Dynamic Characterization of Compliant Materials Using an All-Polymeric Split Hopkinson Bar,'' Exp. Mech., **38**, pp. 204–210.
[15] Frew, D. J., Forrestal, M. J., and Chen, W., 2001, ''A Split Hopkinson Bar Technique to Determine Compressive Stress-Strain Data for Rock Materials,'' Exp. Mech., **41**, pp. 40–46.
[16] Gray, G. T., and Blumenthal, W. R., 2000, ''Split-Hopkinson Pressure Bar Testing of Soft Materials,'' *Mechanical Testing and Evaluation*, Metals Handbook, American Society for Metals, Materials Park, Ohio, **8**, pp. 488–496.
[17] Chen, W., Zhang, B., and Forrestal, M. J., 1999, ''A Split Hopkinson Bar Technique for Low-Impedance Materials,'' Exp. Mech., **39**, pp. 81–85.
[18] Chen, W., Lu, F., and Zhou, B., 2000, ''A Quartz Crystal Embedded Split Hopkinson Bar for Soft Materials,'' Exp. Mech., **40**, pp. 1–6.
[19] Dioh, N. N., Leevers, P. S., and Williams, J. G., 1993, ''Thickness Effects in Split Hopkinson Pressure Bar Tests,'' Polymer, **34**, pp. 4230–4234.
[20] Gray, G. T., III, Blumenthal, W. R., Trujillo, C. P., and Carpenter, R. W., II, 1997, ''Influence of Temperature and Strain Rate on the Mechanical Behavior of Adiprene L-100,'' J. Phys. IV, **7**, pp. 523–528.
[21] Frantz, C. E., Follansbee, P. S., and Wright, W. J., 1984, ''New Experimental Techniques with the Split Hopkinson Pressure Bar,'' *8th International Conference on High Energy Rate Fabrication*, Pressure Vessel and Piping Division, I. Berman and J. W. Schroeder, eds., San Antonio, TX, June 17–21 ASME, New York.
[22] Follansbee, P. S., 1985, ''The Hopkinson Bar,'' *Mechanical Testing*, Metals Handbook, 9th Ed., American Society for Metals, Metals Park, OH, **8**, pp. 198–217.
[23] Follansbee, P. S., and Frantz, C., 1983, ''Wave Propagation in the SHPB,'' ASME J. Eng. Mater. Technol., **105**, pp. 61–66.
[24] Duffy, J., Campbell, J. D., and Hawley, R. H., 1971, ''On the Use of a Torsional Split Hopkinson Bar to Study Rate Effects in 1100-0 Aluminum,'' ASME J. Appl. Mech., **37**, pp. 83–91.
[25] Frew, D. J., Forrestal, M. J., and Chen, W., 2002, ''Pulse-Shaping Techniques for Testing Brittle Materials With a Split Hopkinson Pressure Bar,'' Exp. Mech., **42**, pp. 93–106.
[26] Lindholm, U. S., 1964, ''Some Experiments With the Split Hopkinson Pressure Bar,'' J. Mech. Phys. Solids, **12**, pp. 317–335.
[27] Chen, W., and Zhang, X., 1997, ''Dynamic Response of Epon 828/T-403 Under Multiaxial Loading at Various Temperatures,'' ASME J. Eng. Mater. Technol., **119**, pp. 305–308.
[28] Chen, W., Subhash, G., and Ravichandran, G., 1994, ''Evaluation of Ceramic Specimen Geometries Used in Split Hopkinson Pressure Bar,'' DYMAT J., **1**, pp. 193–210.
[29] Ravichandran, G., and Subhash, G., 1994, ''Critical Appraisal of Limiting Strain Rates for Compression Testing of Ceramics in a Split Hopkinson Pressure Bar,'' J. Am. Ceram. Soc., **77**, pp. 263–267.
[30] Karnes, C. H., and Ripperger, E. A., 1966, ''Strain Rate Effects in Cold Worked High-Purity Aluminum,'' J. Mech. Phys. Solids, **14**, pp. 75–88.
[31] Wasley, R. J., Hoge, K. G., and Cast, J. C., 1969, ''Combined Strain Gauge-Quartz Crystal Instrumented Hopkinson Split Bar,'' Rev. Sci. Instrum., **40**, pp. 889–894.
[32] Togami, T. C., Baker, W. E., and Forrestal, M. J., 1996, ''A Split Hopkinson Bar Technique to Evaluate the Performance of Accelerometers,'' ASME J. Appl. Mech., **63**, pp. 353–356.
[33] Wu, X. J., and Gorham, D. A., 1997, ''Stress Equilibrium in the Split Hopkinson Pressure Bar Test,'' J. Phys. IV, **7**, pp. 91–96.
[34] Ravichandran, G., and Chen, W., 1991, ''Dynamic Behavior of Brittle Materials Under Uniaxial Compression,'' *Experiments in Micromechanics of Fracture Resistant Materials*, K. S. Kim ed., AMD-130, ASME, New York, pp. 85–90.
[35] Nemat-Nasser, S., Isaacs, J. B., and Starrett, J. E., 1991, ''Hopkinson Techniques for Dynamic Recovery Experiments,'' Proc. R. Soc. London, Ser. A, **A435**, pp. 371–391.
[36] Lewis, C. F., 1979, ''Properties and Selection: Nonferrous Alloys and Pure Metals,'' *Metals Handbook*, 9th Ed., **2**, American Society for Metals, Materials Park, OH.

C. D. Copper
Tyco Electronics,
Harrisburg, PA 17105

W. D. Pilkey
Department of Mechanical and
Aerospace Engineering,
University of Virginia,
Charlottesville, VA 22903

Thermoelasticity Solutions for Straight Beams

This paper presents a thermoelastic solution technique for beams with arbitrary quasi-static temperature distributions that create large transverse normal and shear stresses. This technique calculates the stress resultants and centroid displacements along a beam. Then, the stress resultants and temperature distribution are used to calculate the stress distributions on a cross section of the beam. Simple examples demonstrate the numerical efficiency of the proposed technique and the inadequacy of the strength of materials theory to solve these types of problems. [DOI: 10.1115/1.1427340]

Introduction

The theory of thermoelasticity predicts the response of a solid continuum to thermal and mechanical loads and displacement boundary conditions. Unfortunately, the resulting differential equations are difficult to solve directly. Hence for particular geometries like beams, the thermoelastic equations under various assumptions have been reduced to tractable sets of equations referred to as the theory of strength of materials. For beams, the strength of materials solution is based on the assumption that the beam is long and thin and the transverse normal stresses and strains are small.

Consider the class of thermally loaded beam problems that have large transverse normal stresses and strains. For these, the strength of materials solution technique is insufficient to predict the response. Boley [1] was able to prove that a beam with a simply connected cross section would fall into this class of problems only if the temperature distribution on the cross section was not plane-harmonic ($\nabla^2 T \neq 0$), i.e., only if there are thermal sources on the cross section. He implied that this is also true for beams with multiply connected cross sections. As will be demonstrated, this contention is not the case for beams with multiply connected cross sections.

This paper intends to propose a method to solve the thermoelastic equations for straight beams which can be applied in a systematic manner to large complicated structures. Since the proposed method is based on the thermoelastic equations, it can be used accurately on thermally loaded beam structures with large transverse normal stresses. Hence it provides a new tool for engineers.

Previously, methods to solve the thermoelastic equations for straight beams have been developed. In 1960, Boley [1] presented a thermoelastic formulation for a free-free prismatic beam. He used an Airy stress function to calculate the stress distributions on the cross section of a prismatic beam with a temperature distribution that is arbitrary except that it varies linearly along the length of the beam. If the beam lies along the x-axis, the Airy stress function provides a solution for the three normal stresses σ_{xx}, σ_{yy}, σ_{zz} and the shear stress σ_{yz}. The Airy stress function is defined by the biharmonic equation. The shear stresses σ_{xy} and σ_{xz} are determined by a second stress function.

Various individuals have tried to solve the biharmonic equation in Boley's formulation. In 1960, Barrekette [2] applied Boley's theory to a prismatic beam with an elliptical cross section and solved the biharmonic equation using a series solution. In 1989, Pilkey and Liu [3] used a finite element code to solve the biharmonic equation. By using a plate element they were able to obtain an analogical solution for the Airy stress function for a beam with any simply connected cross section. Although this method was successful, they did not attempt to find the second stress function required for the remaining two shear stresses. All of these attempts to use Boley's theory for calculating the stress distribution on a thermally loaded beam have produced accurate results. But they have all been limited to beams with simply connected cross sections.

To develop an improved thermoelastic solution technique for straight beams, this paper derives a solution for relatively simple geometry and loading and then assumes that the relationships between stresses and stress resultants and between stress resultants and centroid displacements hold for more general cases. The thermoelastic solution is developed for a prismatic beam with an arbitrary cross-sectional shape, without in-span mechanical loads or supports, and with a temperature distribution that varies linearly with respect to the span of the beam. Hence, if x is the coordinate along the length of the beam, then the temperature distribution can be written as

$$T(x,y,z) = T_1(y,z) + xT_2(y,z) \tag{1}$$

where T_1 and T_2 are arbitrary functions of y and z.

To obtain the thermoelastic solution for the simple beam, components of the stresses and displacements are defined to satisfy differential equations that coincidentally are the same as the relationships that represent the state of plane strain. This does not mean that the beam is in a state of plane strain, but that components of the stresses and displacements satisfy the equations that define plane strain. This approach permits commercial plane-strain programs to be used to solve the straight beam problem.

The thermoelastic solution of the simple beam defines a solution technique which is uncoupled. First, the stress resultants and centroid displacements are calculated along a beam, as in strength of materials. By identifying an alternative definition of the thermal load, traditional strength of materials can be used to get accurate stress resultants and centroid displacements. Once they are known, the stress resultants and temperature distribution are used to calculate the stress distributions on the cross section of the beam.

Basic Equations of Thermoelasticity

Strain-Displacement Relationship. For linear elasticity, the strains $\epsilon_{ij}(i,j=x,y,z)$ are defined by

$$\epsilon_{ij} = \tfrac{1}{2}(u_{i,j} + u_{j,i}); \quad (i,j=x,y,z) \tag{2}$$

where u_i represents the displacements in the x, y, and z direction.

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. 22, 1993; final revision, Aug. 9, 2001. Associate Editor: X. Markenscoff. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

Material Law. For a homogeneous, isotropic solid, the stresses and strains are related by

$$\sigma_{ij}=\delta_{ij}[\lambda\epsilon_{kk}-(3\lambda+2\mu)\alpha T]+2\mu\epsilon_{ij} \tag{3a}$$

or

$$\epsilon_{ij}=\frac{1}{E}[(1+\nu)\sigma_{ij}-\delta_{ij}(\nu\sigma_{kk}-\alpha ET)] \tag{3b}$$

where $i,j,k=x,y,z$. The constants ν, μ, λ and E satisfy $2\mu(1+\nu)=E$ and $\lambda(1+\nu)(1-2\nu)=\nu E$, α is the coefficient of thermal expansion, $\delta_{ij}=1$ if $i=j$, and $\delta_{ij}=0$ if $i\neq j$.

Equations of Equilibrium (No Body Forces). For a body to be in equilibrium, the stresses must satisfy

$$\sigma_{ij,j}=0; \quad (i,j=x,y,z) \tag{4}$$

at every point in the body.

Plane-Strain Formulation

In plane-strain problems, the out-of-plane normal strain and shear strains are assumed to be zero. This section presents the displacement formulation for a thermally loaded plane-strain problem where the x-axis is normal to the plane. Thus, $\epsilon_{xx}=\epsilon_{xy}=\epsilon_{xz}=0$. A region on which the problem could be defined is shown in Fig. 1. The temperature distribution is $T=T(y,z)$ and the unknown displacements in the y and z directions are defined as $u_y=u_y(y,z)$ and $u_z=u_z(y,z)$. Substituting the displacements and temperature distribution into the strain-displacement relationship and material law in Eqs. (2) and (3a) yields

$$\sigma_{ij}=\delta_{ij}[\lambda u_{k,k}-(3\lambda+2\mu)\alpha T]+\mu(u_{i,j}+u_{j,i}); \quad (i,j,k=y,z)$$
$$\sigma_{xx}=\lambda(u_{y,y}+u_{z,z})-(3\lambda+2\mu)\alpha T; \quad \sigma_{xy}=\sigma_{xz}=0. \tag{5}$$

Substitution of the stresses of Eq. (5) into the equilibrium Eq. (4) leads to the two differential equations

$$\delta_{ij}[\lambda u_{k,kj}-(3\lambda+2\mu)\alpha T_{,j}]+\mu(u_{i,jj}+u_{j,ij})=0; \quad (i,j,k=y,z) \tag{6}$$

which the displacement functions must satisfy in the region of the problem. The third equation of equilibrium is satisfied automatically.

The stresses in Eq. (5) must satisfy the traction-free boundary conditions. The boundary conditions are defined by the outward pointing normal stress σ_{nn} and the shear stress σ_{nm} which are displayed on an exterior and interior boundary in Fig. 1. They are defined by

$$\sigma_{nn}=\sigma_{yy}\cos^2\psi+\sigma_{zz}\sin^2\psi+2\sigma_{yz}\cos\psi\sin\psi$$
$$\sigma_{nm}=(\cos^2\psi-\sin^2\psi)\sigma_{yz}+\sin\psi\cos\psi(\sigma_{zz}-\sigma_{yy}), \tag{7}$$

where ψ is measured from the positive y-axis to the outward pointing normal vector on the boundary. Since the boundary is traction free, $\sigma_{nn}=\sigma_{nm}=0$ on the boundary. Substituting the stresses in Eq. (5) into Eq. (7) and setting σ_{nn} and σ_{nm} to zero defines the boundary conditions as

$$[\lambda u_{z,z}+(\lambda+2\mu)u_{y,y}-(3\lambda+2\mu)\alpha T]\cos^2\psi$$
$$+[\lambda u_{y,y}+(\lambda+2\mu)u_{z,z}-(3\lambda+2\mu)\alpha T]\sin^2\psi$$
$$+2\mu(u_{y,z}+u_{z,y})\cos\psi\sin\psi=0$$
$$\mu(u_{y,z}+u_{z,y})(\cos^2\psi-\sin^2\psi)+2\mu\sin\psi\cos\psi(u_{z,z}-u_{y,y})=0. \tag{8}$$

It follows from the uniqueness theorem presented by Boley [1] that if u_y and u_z are smooth and satisfy (6) across the area and (8) on the boundary of the problem, the displacements constitute a unique solution to the plane-strain problem, ignoring rigid-body motion.

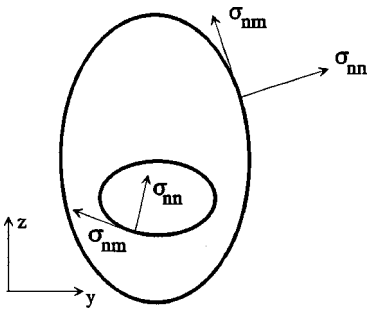


Fig. 1 Multiply connected plane-strain region with traction forces displayed on the inner and outer boundaries

Straight Beam Formulation

This section derives a solution of the thermoelastic equations for the prismatic beam in Fig. 2 without in-span loads or supports and with a temperature distribution defined in Eq. (1). The end conditions of the beam include mechanical loads and displacements. The stress resultants on the cross section of the beam, which are displayed in Fig. 2, include a normal resultant, N_x, two shear resultants, V_y and V_z, and two bending couples, M_y and M_z. If the origin of the y-z plane is at the centroid of cross section, N_x, M_y, and M_z are defined as

$$N_x=\int_A\sigma_{xx}dA; \quad M_y=\int_A z\sigma_{xx}dA; \quad M_z=-\int_A y\sigma_{xx}dA \tag{9}$$

where A is the area of the cross section of the beam. By equilibrium, V_y and V_z are expressed as

$$V_z=M_{y,x}; \quad -V_y=M_{z,x}. \tag{10}$$

The solution of the thermoelastic equations is obtained by the semi-inverse method in which assumptions are made about the stresses at the outset. Then, the thermoelastic equations are solved based on the assumptions. Since the beam lacks in-span loads and supports, $V_{y,x}=V_{z,x}=0$. This inspires the first assumption that

$$\sigma_{xz,x}=\sigma_{xy,x}=0. \tag{11a}$$

Since $V_{y,x}=V_{z,x}=0$, Eq. (10) leads to $M_{y,xx}=M_{z,xx}=0$. This is the basis for the second assumption that

$$\sigma_{xx,xx}=\sigma_{yy,xx}=\sigma_{zz,xx}=\sigma_{yz,xx}=0. \tag{11b}$$

Integrate the four expressions in Eq. (11b) with respect to x, giving

$$\sigma_{yy}=\sigma_{yy1}+x\sigma_{yy2}; \quad \sigma_{zz}=\sigma_{zz1}+x\sigma_{zz2}; \quad \sigma_{yz}=\sigma_{yz1}+x\sigma_{yz2}$$
$$\sigma_{xx}=\sigma_{xx1}+x\sigma_{xx2}+C_1(x)+yC_2(x)+zC_3(x) \tag{12}$$

where σ_{xxk}, σ_{yyk}, σ_{zzk}, and σ_{yzk} $(k=1,2)$ are functions of y and z, and $C_i(x)$ is defined as

$$C_i(x)=C_{i1}+xC_{i2}; \quad (i=1,2,3) \tag{13}$$

where C_{i1} and C_{i2} are constants. The form of σ_{xx} in (12) is different from that of the other stresses. This was done to accommodate the evaluation of stress resultants later.

At this point, the values of σ_{xxk}, σ_{yyk}, σ_{zzk}, and σ_{yzk} $(k=1,2)$ in Eq. (12) are defined to satisfy differential equations that are identical in form to those that represent a traction-free plane-strain problem on a region identical to the cross section of the beam with a temperature distribution T_k $(k=1,2)$ from Eq. (1). The beam in Fig. 2 is not in a state of plane strain, but that does not prevent components of the stresses from satisfying the same differential equations that define plane strain. Let $u_{yk}(y,z)$ and $u_{zk}(y,z)$ $(k=1,2)$ be the displacements in the y and z directions that produce the stress components σ_{xxk}, σ_{yyk}, σ_{zzk}, and σ_{yzk} for the temperature distribution T_k. Hence the displacements u_{yk} and

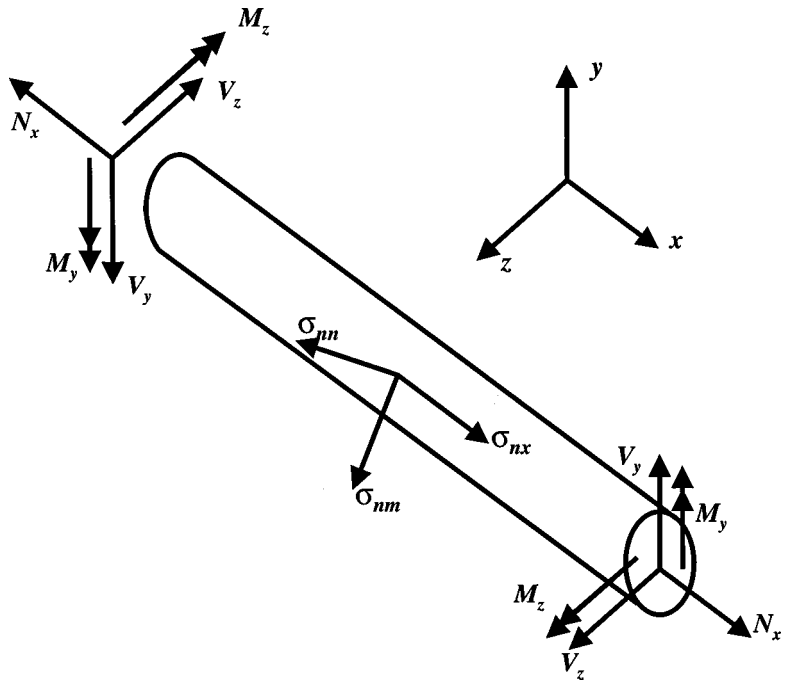


Fig. 2 Prismatic beam displaying the traction forces on the cross section boundary and the positive stress resultants at the ends

u_{zk} $(k=1,2)$ satisfy the plane-strain differential Eq. (6) and boundary conditions (8). Using Eq. (5), the stresses σ_{xxk}, σ_{yyk}, σ_{zzk}, and σ_{yzk} can be expressed as

$$\sigma_{ijk}=\delta_{ij}[\lambda u_{nk,n}-(3\lambda+2\mu)\alpha T_k]+\mu(u_{ik,j}+u_{jk,i});$$

$$(i,j,n=y,z:\ k=1,2)$$

$$\sigma_{xxk}=\lambda(u_{yk,y}+u_{zk,z})-(3\lambda+2\mu)\alpha T_k;\quad (k=1,2). \tag{14}$$

These representations of the stresses are functions of y and z. Hence, they can be used in Eq. (12) to define the stresses of the beam without violating the assumptions in Eq. (11*b*).

Once the values in Eq. (14) are substituted into Eq. (12), the material law in Eq. (3*b*) will define four of the strains of the beam. Use of these four strains and the strain-displacement relationship in Eq. (2) produces the displacements

$$u_x(x,y,z)=w(y,z)+\frac{1}{E}[D_{1,x}(x)+yD_{2,x}(x)+zD_{3,x}(x)]$$

$$u_y(x,y,z)=u_{y1}(y,z)+xu_{y2}(y,z)-\frac{1}{E}\left[D_2(x)+\nu\left(yC_1+\frac{(y^2-z^2)}{2}C_2+yzC_3\right)\right] \tag{15}$$

$$u_z(x,y,z)=u_{z1}(y,z)+xu_{z2}(y,z)-\frac{1}{E}\left[D_3(x)+\nu\left(zC_1+yzC_2+\frac{(z^2-y^2)}{2}C_3\right)\right]$$

where the functions $D_i(x)$ are

$$D_i(x)=C_{i1}\frac{x^2}{2}+C_{i2}\frac{x^3}{6};\quad (i=1,2,3) \tag{16}$$

and C_{i1} and C_{i2} are the same constants as in Eq. (13). When the strains are integrated to get the displacements in Eq. (15), arbitrary integration functions are produced. Most of these have to be defined to satisfy Eq. (11), but one remains arbitrary and appears in Eq. (15) as a warping term, $w(y,z)$.

Substituting Eq. (15) into the strain-displacement relationship and the material law in Eqs. (2) and (3*a*) leads to the two remaining shear stresses

$$\sigma_{xy}=\mu\left[w_{,y}+u_{y2}-\frac{\nu}{E}\left(yC_{12}+\frac{(y^2-z^2)}{2}C_{22}+yzC_{32}\right)\right]$$

$$\sigma_{xz}=\mu\left[w_{,z}+u_{z2}-\frac{\nu}{E}\left(zC_{12}+yzC_{22}+\frac{(z^2-y^2)}{2}C_{32}\right)\right] \tag{17}$$

which satisfy the assumption in Eq. (11*a*).

Substitution of the stresses of Eqs. (12), (14), and (17) into the equilibrium Eq. (4) provides three differential equations that the displacements in Eq. (15) must satisfy. Since u_{yk} and u_{zk} $(k=1,2)$ satisfy Eq. (6) for a temperature distribution T_k $(k=1,2)$, two of the equilibrium equations are satisfied automatically. The third equilibrium equation leads to the differential equation

$$w_{,yy}+w_{,zz}+\frac{\sigma_{xx2}}{\mu}+u_{y2,y}+u_{z2,z}+\left(\frac{1}{\mu}-\frac{2\nu}{E}\right)(C_{12}+yC_{22}+zC_{32})=0. \tag{18}$$

The stresses must satisfy the traction-free conditions on the boundary of the cross section of the beam. The conditions are on the outward pointing normal stress σ_{nn} and the two shear stresses σ_{nm} and σ_{xn} shown in Fig. 2. The σ_{nn} and σ_{nm} stresses are defined in Eq. (7) and the σ_{xn} stress is defined by

$$\sigma_{xn}=\sigma_{xz}\sin\psi+\sigma_{xy}\cos\psi. \tag{19}$$

Since the beam lacks in-span loads and supports, $\sigma_{nn}=\sigma_{nm}=\sigma_{xn}=0$ on the boundary of the cross section. Since u_{yk} and u_{zk} $(k=1,2)$ satisfy Eq. (8) for T_k $(k=1,2)$, the requirement that $\sigma_{nn}=\sigma_{nm}=0$ on the boundary of the cross section is satisfied automatically. This leaves $\sigma_{xn}=0$ to be considered. Substituting Eq. (17) into Eq. (19) and setting σ_{xn} to zero defines the boundary condition

$$\begin{aligned} w_{,y}\cos\psi+w_{,z}\sin\psi &= \left[\frac{\nu}{E}\left(yC_{12}+\frac{(y^2-z^2)}{2}C_{22}+yzC_{32}\right)-u_{y2}\right]\cos\psi \\ &\quad+\left[\frac{\nu}{E}\left(zC_{12}+yzC_{22}+\frac{(z^2-y^2)}{2}C_{32}\right)-u_{z2}\right]\sin\psi. \end{aligned} \tag{20}$$

Now consider the stress resultants and centroid displacements of the beam. Substitute σ_{xx} from Eq. (12) into Eq. (9) and solve for $C_i(x)$ to obtain the familiar relations

$$\begin{aligned} C_1(x)&=\frac{1}{A}\left[N_x-\int_A(\sigma_{xx1}+x\sigma_{xx2})dA\right] \\ C_2(x)&=\frac{I_y\left(-M_z-\int_A y(\sigma_{xx1}+x\sigma_{xx2})dA\right)-I_{yz}\left(M_y-\int_A z(\sigma_{xx1}+x\sigma_{xx2})dA\right)}{I_yI_z-I_{yz}^2} \\ C_3(x)&=\frac{I_z\left(M_y-\int_A z(\sigma_{xx1}+x\sigma_{xx2})dA\right)-I_{yz}\left(-M_z-\int_A y(\sigma_{xx1}+x\sigma_{xx2})dA\right)}{I_zI_y-I_{yz}^2} \end{aligned} \tag{21}$$

where I_y, I_z, and I_{yz} are the cross-sectional bending moments of inertia. Since the beam lacks in-span supports and loads, N_x does not vary with respect to x and M_y and M_z vary linearly with respect to x. Hence the definitions of $C_i(x)$ in Eq. (21) are linear functions of x and do not violate Eq. (13). The displacement end conditions are applied to the centroid of the beam, which is at the origin of the y-z plane. Differentiating Eq. (15) with respect to x and evaluating at $y=z=0$ leads to

$$\begin{aligned} &u_{x,x}(x,0,0)=\frac{C_1(x)}{E};\quad u_{y,xx}(x,0,0)=-\frac{C_2(x)}{E}; \\ &u_{z,xx}(x,0,0)=-\frac{C_3(x)}{E}. \end{aligned} \tag{22}$$

Note that the differential equations produced by substituting Eq. (21) into (22) are identical to those defined by strength of materials if σ_{xxk} is replaced by $-\alpha ET_k$ $(k=1,2)$.

If u_{yk} and u_{zk} $(k=1,2)$ satisfy the plane-strain problem for the temperature T_k, w satisfies Eq. (18) and (20), and C_i satisfies Eq. (21) and (22), then the displacements in Eq. (15) satisfy equilibrium, material law, strain displacement, the traction-free boundary conditions on the cross section, and the end conditions of the beam. It should be recognized that no effort is made to match the stresses σ_{xx}, σ_{xy}, and σ_{xz} with the boundary conditions at the ends of the beam. This is justified by Saint Venant's principle as discussed by Boley [1]. Saint Venant's principle maintains that these effects on the stresses that are away from the ends of the beam are small. The minimum distance from the ends of the beam where the stresses are accurate is equal to the maximum width of the cross section of the beam.

Fortunately, the unknown components of the displacements in Eq. (15) do not have to be found simultaneously. The equations that define the unknown components are uncoupled. First, obtain solutions for the two traction-free plane-strain problems for the temperature distribution T_k that define u_{yk}, u_{zk}, σ_{xxk}, σ_{yyk}, σ_{zzk}, and σ_{yzk} $(k=1,2)$. Second, calculate the integrals of σ_{xxk} $(k=1,2)$ in Eq. (21). Third, solve the coupled ordinary differential equations produced by substituting Eq. (21) into (22) to get the stress resultants and centroid displacements along the length of the beam. Fourth, use Eq. (12) and (21) to get the stress distributions for σ_{xx}, σ_{yy}, σ_{zz}, and σ_{yz}. Fifth, to calculate w, solve the Poisson equation in (18) with the boundary condition in (20). Finally, use Eq. (17) to obtain the stress distributions for σ_{xy} and σ_{xz}. This demonstrates how plane-strain theory can be used as an integral part of the thermoelastic solution of a straight beam.

Comparison With Strength of Materials

By Eq. (3*b*) $\sigma_{xxi}=\nu(\sigma_{yyi}+\sigma_{zzi})-\alpha ET_i$ $(i=1,2)$. As $\nu(\sigma_{yyi}+\sigma_{zzi})$ approaches 0, σ_{xxi} approaches $-\alpha ET_i$. By replacing σ_{xxi} with $-\alpha ET_i$ in Eqs. (12), (21), and (22), they become identical to those for strength of materials. Hence the strength of materials solution for the stress resultants, centroid displacements and normal bending stress is a special case of the thermoelastic solution for small values of $\nu(\sigma_{yyi}+\sigma_{zzi})$. Note that if ν is small then strength of materials produces accurate results, but may not produce all the necessary results because σ_{yyi} and σ_{zzi}, which are ignored by strength of materials, may still be significant.

Application to Beam Structures

As stated in the Introduction, the purpose of this thermoelastic derivation is to establish relationships on the cross section of a simple beam between stress resultants and centroid displacements and between stress resultants and stress distributions, and then assume they hold for a more general case. Consider a beam structure or frame with thermal loading. By assuming that the equations derived in the previous section hold at every cross section of every member of the structure, a solution can be obtained. What follows is the resulting step-by-step technique to obtain the solution for the beam structure.

Step 1: Consider a member of the structure. Define a coordinate system such that x lies along the length of the beam and the origin of the y-z plane is at the centroid of the cross section. The thermal load of the beam expressed in this coordinate system is $T(x,y,z)$. At $x=x_o$, the cross section of the member is A_o. Calculate the solution of a traction-free plane-strain problem with a temperature distribution $T(x_o,y,z)$ on a region identical to A_o using the material properties of the beam. Denote the stresses of the plane-strain problem as σ_{xxo}, σ_{yyo}, σ_{zzo}, and σ_{yzo}. This should be done at every cross section along the length of the beam. As x_o varies, the shape of A_o and the plane-strain stresses vary. This step

should be repeated for every member of the structure. These calculations can be done by a general purpose finite element code.
Step 2: Evaluate the integrals that define M_{Tz}, M_{Ty}, and P_T as

$$P_T = -\int_{A_0} \sigma_{xxo} dA; \quad M_{Ty} = -\int_{A_o} z\sigma_{xxo} dA;$$
$$M_{Tz} = -\int_{A_o} y\sigma_{xxo} dA \tag{23}$$

which vary with respect to x. This should be repeated for every member of the structure. A numerical method can be used to perform the integration.
Step 3: Define a fictitious temperature distribution, $T_f = a_1(x) + ya_2(x) + za_3(x)$, that satisfies

$$-\int_A k\alpha E T_f dA = \int_A k\sigma_{xxo} dA; \quad (k = 1, y, z) \tag{24}$$

for each member of the structure. If the y and z-axes correspond to the principal axes, then $a_1 = P_T/\alpha E A_o$, $a_2 = M_{Tz}/\alpha E I_z$, and $a_3 = M_{Ty}/\alpha E I_y$ where M_{Tz}, M_{Ty}, and P_T are defined in Step 2.
Step 4: Use a general purpose finite element program to analyze the structure. Apply the boundary conditions and force loads to the model. Instead of using the true temperature distributions on the beam elements, use the fictitious ones defined in Step 3. The solution will give the correct thermoelastic values for the stress resultants and centroid displacements along the length of each member of the structure.
Step 5: Once the stress resultants have been calculated in Step 4, the distributions of σ_{xx}, σ_{yy}, σ_{zz}, and σ_{yz} on the cross section A_o of any member of the structure can be computed by

$$\sigma_{yy} = \sigma_{yyo}; \quad \sigma_{zz} = \sigma_{zzo}; \quad \sigma_{yz} = \sigma_{yzo} \tag{25}$$
$$\sigma_{xx} = \sigma_{xxo} + \frac{1}{A_o}(N_x + P_T) + y\frac{I_y(-M_z + M_{Tz}) - I_{yz}(M_y + M_{Ty})}{I_y I_z - I_{yz}^2} + z\frac{I_z(M_y + M_{Ty}) - I_{yz}(-M_z + M_{Tz})}{I_z I_y - I_{yz}^2}.$$

Step 6: The two shear stresses σ_{xy} and σ_{xz} remain to be considered. In order to get these shear stresses, repeat Steps 1 and 2 for a particular cross section of interest, but use $T_{,x}(x_o, y, z)$ instead of $T(x_o, y, z)$ as the thermal load. Denote the stresses of the plane-strain problem as σ_{xxp}, σ_{yyp}, σ_{zzp}, and σ_{yzp}, the displacements as u_{yp} and u_{zp}, and the values defined in Step 2 as M_{Tzp}, M_{Typ} and P_{Tp}.
Step 7: Solve for the warping, $w(y,z)$, on the cross section of interest by using the differential equation

$$w_{,yy} + w_{,zz} + \frac{\sigma_{xxp}}{\mu} + u_{yp,y} + u_{zp,z} + \left(\frac{1}{\mu} - \frac{2\nu}{E}\right)\left(\frac{P_{Tp} + N_{x,x}}{A_o} + yB_2 + zB_3\right) = 0 \tag{26}$$

and boundary condition

$$w_{,y}\cos\psi + w_{,z}\sin\psi = \left[\frac{\nu}{E}\left(y\frac{P_{Tp} + N_{x,x}}{A_o} + \frac{(y^2 - z^2)}{2}B_2 + yzB_3\right) - u_{yp}\right]\cos\psi + \left[\frac{\nu}{E}\left(z\frac{P_{Tp} + N_{x,x}}{A_o} + yzB_2 + \frac{(z^2 - y^2)}{2}B_3\right) - u_{zp}\right]\sin\psi \tag{27}$$

where the B_2 and B_3 are defined as

$$B_2 = \frac{I_y(-M_{z,x} + M_{Tzp}) - I_{yz}(M_{y,x} + M_{Typ})}{I_y I_z - I_{yz}^2}$$
$$B_3 = \frac{I_z(M_{y,x} + M_{Typ}) - I_{yz}(-M_{z,x} + M_{Tzp})}{I_y I_z - I_{yz}^2}. \tag{28}$$

and $N_{x,x}$, $M_{y,x}$, and $M_{z,x}$ are the derivatives of the stress resultants found in Step 4. A two-dimensional steady-state thermal conduction problem is defined by the differential equation $k(T_{,xx} + T_{,yy}) + Q = 0$ and the boundary condition $k(T_{,x}\cos\psi + T_{,y}\sin\psi) = q$ where k is the thermal resistance, Q is the thermal source, and q is the thermal flow. By comparing these equations to (26) and (27), it is clear that an analogical solution for w can be calculated by a two-dimensional general purpose finite element code for thermal conduction if $k = 1$, q is equal to the right-hand side of Eq. (27) and

$$Q = \frac{\sigma_{xxp}}{\mu} + u_{yp,y} + u_{zp,z} + \left(\frac{1}{\mu} - \frac{2\nu}{E}\right)\left(\frac{P_{Tp} + N_{x,x}}{A_o} + yB_2 + zB_3\right). \tag{29}$$

Step 8: Calculate the remaining shear stresses on the cross section of interest by using

$$\sigma_{xy} = \mu\left[w_{,y} + u_{yp} - \frac{\nu}{E}\left(y\frac{P_{Tp} + N_{x,x}}{A_o} + \frac{(y^2 - z^2)}{2}B_2 + yzB_3\right)\right]$$
$$\sigma_{xz} = \mu\left[w_{,z} + u_{zp} - \frac{\nu}{E}\left(z\frac{P_{Tp} + N_{x,x}}{A_o} + yzB_2 + \frac{(z^2 - y^2)}{2}B_3\right)\right]. \tag{30}$$

Simply Connected Example

Consider a free-free beam that lies along the x-axis with a square cross section. The height and width of the cross section is $H = 0.1$ meters and the length is $L = 10.0$ meters. The thermal load is defined as

$$T(x,y,z) = T_c\frac{x}{L}\left(\frac{z}{H}\right)^3$$

where $T_c = 1$ K and the material constants are $E = 70$ GPa, $\nu = 0.35$, and $\alpha = 23.6$ E-6/K.

Strength of Materials Solution. The strength of materials solution, presented by Boley [1] and Ugural and Fenster [4], yields

$$\sigma_{xx} = T_c\alpha E\frac{x}{L}\left[\frac{3}{5}\left(\frac{z}{H}\right) - \left(\frac{z}{H}\right)^3\right]$$
$$\sigma_{xz} = \alpha T_c\frac{E}{L}\left[\frac{z^2}{H}\left(\frac{3}{10} - \frac{z^2}{4H^2}\right) - \frac{H}{20}\right] \quad \sigma_{xy} = 0.$$

The remaining stresses are ignored. The maximum stress values on the cross section at $x = 0.1$ calculated by this solution are listed in Table 1.

Proposed Thermoelastic Solution. The plane-strain solutions in Steps 1 and 5 use uniform 10×10 meshes of ANSYS stif42 (plane-strain option) elements. The Newton-Cotes method is used to evaluate the integrals in Steps 2 and 6. To solve for $\omega(y,z)$ in Step 7, a uniform 50×50 mesh of ANSYS stif55 elements is used. The maximum stress values on the cross section at $x = 0.1$ calculated by this solution are listed in Table 1.

Direct Three-Dimensional Thermoelastic Solution. As a check on the accuracy of the method presented in this paper, a three-dimensional finite element solution is performed. A 0.4-meter length of the beam is modeled using ANSYS stif45 elements. The element mesh is $10\times10\times20$ and spans the distance between $x = -0.1$ and $x = 0.3$. This model will serve as a check

Table 1 Largest Stress Values Calculated by Each Method (Pa) at $x=0.1$

Stress	3-D Thermoelastic Solution	Proposed Thermoelastic Solution	Strength of Materials Solution
σ_{xx}	9386	9433	6608 ($z=-$H)
σ_{yy}	8977	9054	ignored
σ_{zz}	4430	4525	ignored
σ_{xy}	1938	2066	0
σ_{xz}	2981	3073	826 ($z=0$)
σ_{yz}	1531	1582	ignored

for the stresses at $x=0.1$. The maximum stress values on the cross section at $x=0.1$ calculated by this solution are listed in Table 1.

Summary. As can be seen by Table 1, the proposed thermoelastic solution technique produced accurate maximum stress values. Due to the large transverse normal stress σ_{yy}, this problem is outside of the usual range of problems that elementary strength of materials is intended to solve. Hence, it is not surprising that strength of materials calculates a maximum normal stress that is in error by 30 percent and a maximum shear stress that is in error by 72 percent.

The computational requirements of the proposed thermoelastic solution technique to solve this problem were considerably less than the three-dimensional thermoelastic solution. The ratio of computer memory required to run both solutions is 90/1. The ratio of computer time is 82/1.

Tubular Example

Consider a free-free tubular beam that lies along the x-axis between $x=-L$ and $x=L$. The inner radius is a and the outer radius is b where $a<b\ll L$. The material properties are E, ν and α, and the temperature distribution is $T=T(x,r)$ where $r^2=y^2+z^2$. Define θ as the circumference coordinate.

Thermoelastic Solution. Step 1 becomes the solution of an axisymmetric plane-strain problem presented by Timoshenko [5]. By starting with Timoshenko's closed-form solution, the remaining steps of the solution technique can be calculated without numerical techniques. All the differential equations are ordinary and can be solved directly. This solution gives stress distributions of

$$\sigma_{rr}=\frac{\alpha E}{(1-\nu)}\frac{1}{r^2}\left[\left(\frac{r^2-a^2}{b^2-a^2}\right)\int_a^b Trdr-\int_a^r Trdr\right]$$

$$\sigma_{\theta\theta}=\frac{\alpha E}{(1-\nu)}\frac{1}{r^2}\left[\left(\frac{r^2+a^2}{b^2-a^2}\right)\int_a^b Trdr+\int_a^r Trdr-r^2T\right]$$

$$\sigma_{xx}=\frac{\alpha E}{\langle 1-\nu\rangle}\left[\frac{2}{(b^2-a^2)}\int_a^b Trdr-T\right]$$

$$\sigma_{rx}=\frac{\alpha E}{(1-\nu)r}\left[\left(\frac{a^2-r^2}{b^2-a^2}\right)\int_a^b T_{,x}rdr+\int_a^r T_{,x}rdr\right]$$

$$\sigma_{r\theta}=\sigma_{\theta x}=0.$$

Strength of Materials. By using the strength of materials solution, the stresses are found to be

$$\sigma_{xx}=\alpha E\left[\frac{2}{(b^2-a^2)}\int_a^b Trdr-T\right]$$

$$\sigma_{rx}=\frac{\alpha E}{r}\left[\left(\frac{a^2-r^2}{b^2-a^2}\right)\int_a^b T_{,x}rdr+\int_a^r T_{,x}rdr\right]$$

$$\sigma_{\theta x}=0.$$

The remaining stresses are ignored.

Comparison. The ratios of σ_{xx} and σ_{rx} produced by the thermoelastic technique and strength of materials technique are

$$\frac{\sigma_{xx}(\text{Thermoelastic})}{\sigma_{xx}(\text{Strength of Materials})}=\frac{\sigma_{rx}(\text{Thermoelastic})}{\sigma_{rx}(\text{Strength of Materials})}=\frac{1}{1-\nu}.$$

Thus, if the stress is nonzero, the error of the strength of materials solution for σ_{xx} and σ_{rx} is

$$\text{Percent Error}=\nu 100.$$

Hence, as $\nu\rightarrow 1/2$ the error approaches 50 percent, and as $\nu\rightarrow 0$ the error approaches 0 percent.

Summary

1. Strength of materials is inadequate to solve beam structures with thermal loads that produce large transverse normal stresses. These types of problems require thermoelastic solutions.
2. The technique presented in this paper solves the thermoelastic equations in an efficient manner using numerical techniques available on the commercial market. This is due to the fact that the thermoelastic equations have been reduced to a set of uncoupled problems that have been extensively studied analytically and numerically.

References

[1] Boley, B. A., and Weiner, J. H., 1960, *Theory of Thermal Stresses*, John Wiley and Sons, New York, pp. 76–77, 307–314, 328–332.

[2] Barrekette, E. S., 1960, "Thermoelastic Stresses in Beam," ASME J. Appl. Mech., **27**, pp. 465–473.

[3] Pilkey, W. D., and Liu, Y., 1949, "Thermal Bending Stresses on Beam Cross Section," Finite Elem. Anal. Design, **6**, No. 1, pp. 23–31.

[4] Ugural, A. C., and Fenster, S. K., 1975, *Advanced Strength and Applied Elasticity*, Elsevier, New York, pp. 141–144.

[5] Timoshenko, S., and Goodier, J., 1951, *Theory of Elasticity*, McGraw-Hill, New York, pp. 408–409.

Stress Intensity Factors for Finite Interface Cracks Between a Special Pair of Transversely Isotropic Materials

V. Boniface
Post Doctoral Fellow

L. Banks-Sills
Professor,
Mem. ASME

The Dreszer Fracture Mechanics Laboratory,
Department of Solid Mechanics,
Materials and Structures,
The Fleischman Faculty of Engineering,
Tel Aviv University,
69978 Ramat Aviv, Israel

An infinite bimaterial system made of two dissimilar, transversely isotropic materials bonded together (with the lower material being mathematically degenerate) and subject to remote loads is considered. An analytical expression for the complex stress intensity factor of a finite crack along the interface between these two materials is obtained. This result is extended to the case of an infinite array of collinear cracks along a similar interface. Next, the finite element method is employed to analyze these geometries for specific material properties. An area M-integral is used to extract stress intensity factors from the finite element results, which compare well with those obtained from the analytic expressions. Different types of loads are considered. [DOI: 10.1115/1.1459067]

1 Introduction

In recent years, much research has been directed toward understanding interface fracture. This is largely stimulated by the increasing use of composite materials, which often fail by delamination. Though work on interface fracture began more than three decades ago, it received a new thrust from the seminal papers of Rice [1] and Hutchinson [2]. The issue of crack face contact was sufficiently resolved by Rice [1] when he obtained bounds on the phase angle ψ (mode mixity). It was found that there is a large range of ψ for which the contact region is sufficiently small.

Interface cracks between anisotropic materials were studied by Bassani and Qu [3,4] who derived a condition for nonoscillatory behavior. Later Suo [5] analyzed an interface crack for both nonoscillatory and oscillatory singularities. He found that the near-tip field for the oscillatory singularity is similar to that of a crack between two isotropic materials. Working independently, Ting [6,7] determined the singularity for both the oscillatory and nonoscillatory cases. In addition, he developed the asymptotic stress and displacement fields ([8]). Recently, Banks-Sills and Boniface [9] addressed the specific problem of a semi-infinite crack between two transversely isotropic materials one of which was mathematically degenerate. They obtained the asymptotic stress and displacement fields at the crack tip using the Stroh formalism ([10]). An excellent treatise on these issues is provided by Ting [11].

Stress intensity factors which characterize a given crack are important parameters in evolving an effective fracture design methodology. The stress intensity factors along with their critical values (which are measured experimentally) can be incorporated into an appropriate fracture criterion. This criterion is useful in predicting crack extension and hence assists in mapping a safe operating envelope within which fracture is largely avoided. The present work is directed towards obtaining stress intensity factors for specific crack geometries along a bimaterial interface.

In this investigation, an infinite bimaterial system consisting of two transversely isotropic materials, with the lower half mathematically degenerate, is considered. This configuration closely approximates an interface in a laminated composite with a 0 deg/90 deg ply layup. So the results presented herein find ready application in understanding fracture of laminated composites. Analytic expressions for the stress intensity factors are obtained for two crack configurations in this bimaterial system, namely (a) a finite crack along the interface between the two materials and (b) an infinite array of collinear cracks along such an interface.

The M-integral method is used to determine stress intensity factors from finite element results. This method requires expressions for the asymptotic stress field in this bimaterial system which are available in [9]. Numerical results are obtained and compared to values computed from the analytic expressions. The M-integral, developed by Wang and Yau [12], allows for the determination and separation of the stress intensity factors K_1 and K_2 for interface cracks between two isotropic materials. Since the structure of the near-tip fields for cracks between two anisotropic materials is similar, the same integral formulation may be used. This method has been successfully employed in many problems including isotropic bimaterials [12–14], orthotropic bimaterials [15], and the present bimaterial system ([9]). Additionally, nonplanar problems were addressed by Nakamura [16], Nahta and Moran [17], and Gosz et al. [18].

It should be emphasized here that the goal of this paper is to obtain analytic expressions for stress intensity factors of finite cracks in two geometries. The finite element method and M-integral are used only for comparison, as this numerical approach has been developed and validated earlier in [9].

In Section 2, the bimaterial system is described, and important parameters like singularity, stress intensity factors, interface energy release rate, and crack-face displacements are defined. The problem of a single finite interface crack is addressed in Section 3 and the expression for stress intensity factors derived. Section 4 focuses on stress intensity factors for an infinite array of collinear interface cracks. Finite element results are compared with these analytical expressions in Section 5.

2 The Bimaterial System

An infinite bimaterial system consisting of two transversely isotropic materials bonded together under plane-strain conditions is considered (Fig. 1). In one of the materials (assumed here as the upper material), the axis of symmetry is the x_1-axis and in the other (i.e., the lower material) it is the x_3-axis. The latter results in a mathematical degeneracy. Two crack problems in this bimaterial

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. 10, 2000; final revision, Nov. 14, 2000. Associate Editor: K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

system are studied here: (1) a finite crack along the interface as illustrated in Fig. 2(*a*), and (2) an infinite array of collinear cracks along the interface as in Fig. 3(*a*). Relevant mechanical properties (Young's moduli, shear moduli, and Poisson's ratios) in the axial and transverse directions are E_A, E_T, G_A, G_T, ν_A, and ν_T. Since the material is transversely isotropic, $G_T = E_T/2(1+\nu_T)$. For simplicity, the two materials are assumed identical although their axes of symmetry differ. This assumption is also largely the result of the emphasis of this paper on the 0 deg/90 deg laminate composite. It should, however, be noted that dissimilar materials can be analyzed by using the appropriate material properties.

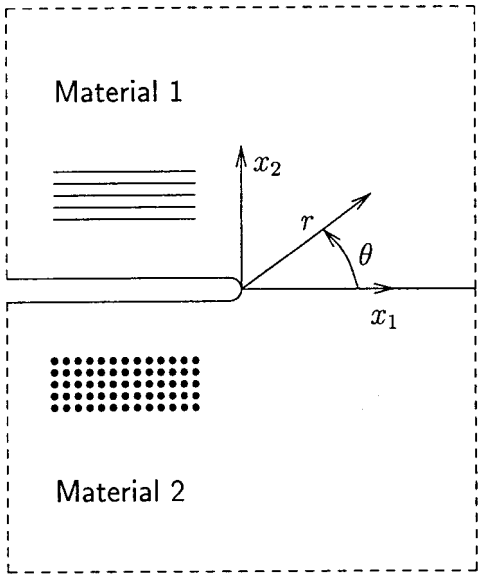


Fig. 1 Crack in a transversely isotropic bimaterial system

Assuming that singular stresses at the crack tip are proportional to r^{δ} where r is the radial distance from the crack tip (Fig. 1) and following [11]

$$\delta = -\frac{1}{2} \quad \text{and} \quad -\frac{1}{2} \mp i\epsilon \tag{1}$$

with the bimaterial constant

$$\epsilon = \frac{1}{2\pi}\ln\left(\frac{1+\beta}{1-\beta}\right) \tag{2}$$

where

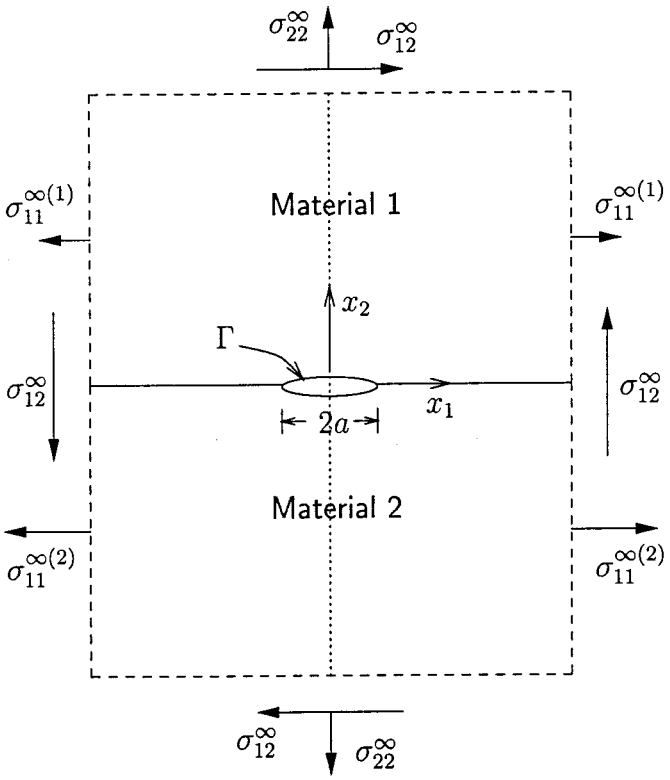
$$\beta = \left\{-\frac{1}{2}\mathrm{tr}(\check{\mathbf{S}})^2\right\}^{1/2}. \tag{3}$$

The first value of δ in (1) leads to the nonoscillatory solution while the second results in the oscillatory solution. Nonoscillatory behavior is possible if there exists a nonzero vector $\mathbf{t}_0$ satisfying the condition ([11], p. 428)

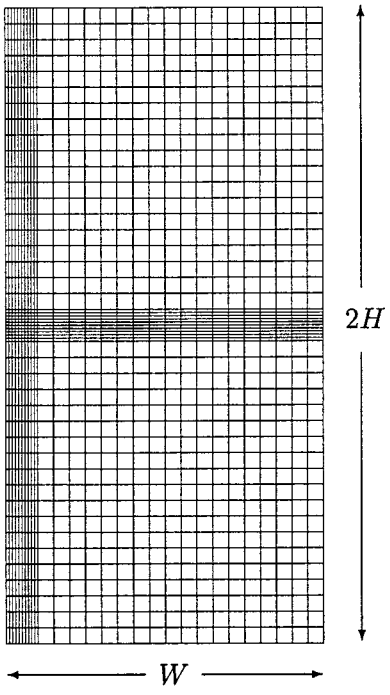
$$\check{\mathbf{S}}\mathbf{t}_0 = \mathbf{0}.$$

Evaluating $\check{\mathbf{S}}$ for this bimaterial system, as explained subsequently, it is seen that $\check{\mathbf{S}} \neq \mathbf{0}$. Further this particular $\check{\mathbf{S}}$ leads to $\mathbf{t}_0 = \mathbf{0}$ for the plane-strain condition considered here. So in this analysis, oscillatory behavior is addressed and

$$\delta = -\frac{1}{2} + i\epsilon. \tag{4}$$



(*a*)



(*b*)

Fig. 2 A finite interface crack (*a*) schematic diagram, (*b*) finite element mesh

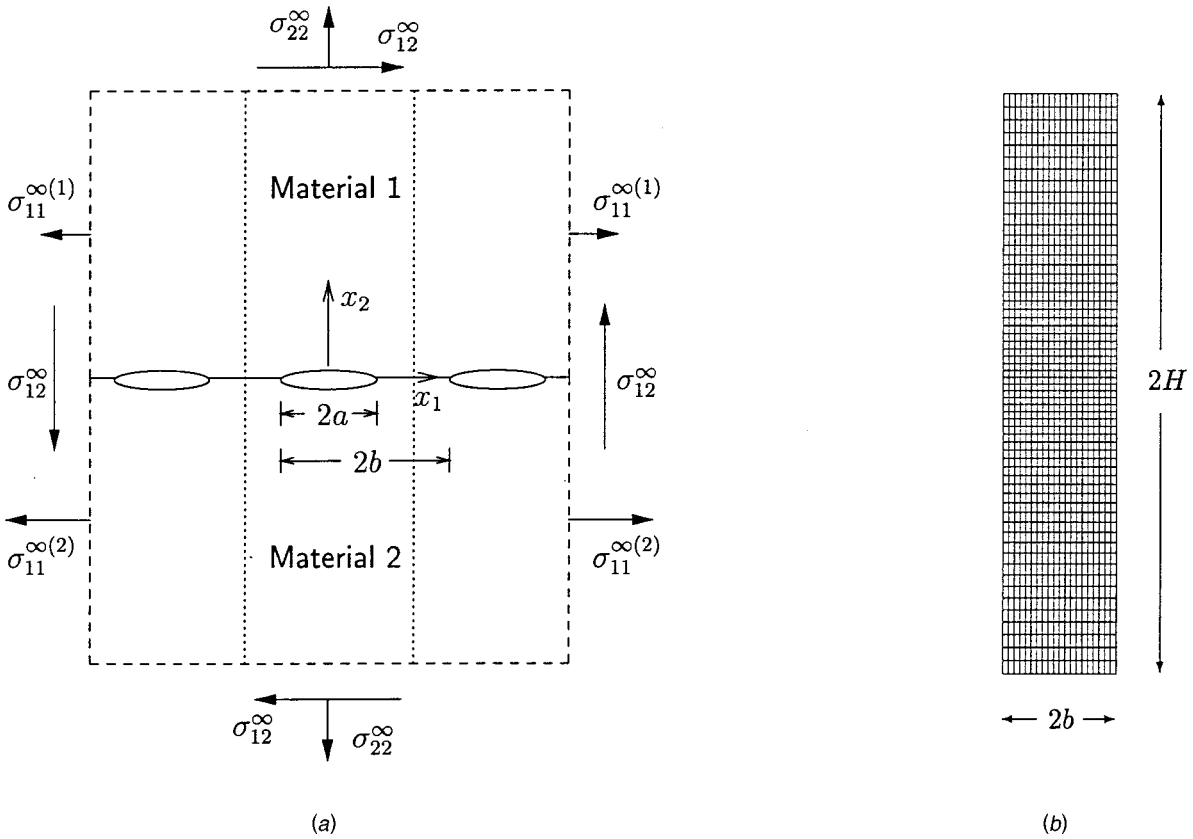


Fig. 3 Infinite array of collinear interface cracks (*a*) schematic diagram, (*b*) finite element mesh

It is seen that it is immaterial if the conjugate of the above value is used, provided one is consistent throughout the analysis.

The 3×3 matrix $\check{\mathbf{S}}$ is given by

$$\check{\mathbf{S}}=\mathbf{D}^{-1}\mathbf{W} \tag{5}$$

where

$$\mathbf{D}=\mathbf{L}_1^{-1}+\mathbf{L}_2^{-1}, \tag{6}$$

$$\mathbf{W}=\mathbf{S}_1\mathbf{L}_1^{-1}-\mathbf{S}_2\mathbf{L}_2^{-1}. \tag{7}$$

The subscripts 1 and 2 in (6) and (7) represent, respectively, the upper and lower material. $\mathbf{S}_j$ and $\mathbf{L}_j$ are Barnett-Lothe tensors. Since these tensors are real and

$$-\mathbf{A}_j\mathbf{B}_j^{-1}=\mathbf{S}_j\mathbf{L}_j^{-1}+i\mathbf{L}_j^{-1}, \tag{8}$$

knowledge of the left-hand side of (8) is sufficient to determine (6) and (7). Instead of presenting more general expressions for the matrices $\mathbf{A}_j$ and $\mathbf{B}_j$ which may be found in ([11], pp. 170–172), the specific matrices for the upper and lower materials are given in Appendix A.

Using these matrices, ϵ is obtained as

$$\epsilon=w/\sqrt{D_{11}D_{22}} \tag{9}$$

where

$$w=-\frac{(1+\nu_T)\nu_A}{E_A}+\beta_1\beta_2\left(1-\nu_A^2\frac{E_T}{E_A}\right)\frac{1}{E_A}+\frac{(1-2\kappa)}{4G_T}, \tag{10}$$

$$D_{11}=\frac{(\beta_1+\beta_2)}{E_A}\left(1-\nu_A^2\frac{E_T}{E_A}\right)+\frac{1+2\kappa}{4G_T}, \tag{11}$$

$$D_{22}=\frac{\beta_1\beta_2(\beta_1+\beta_2)}{E_A}\left(1-\nu_A^2\frac{E_T}{E_A}\right)+\frac{1+2\kappa}{4G_T}. \tag{12}$$

The material parameters D_{11} and D_{22} are components of the matrix $\mathbf{D}$ in (6). The parameter κ is given in Appendix A. Substituting κ in the above relations, it is seen that D_{11} and D_{22} always have the same sign. The constants β_j, $j=1,2,3$ are related to $p_j^{(1)}$, the three complex eigenvalues for the upper material ([11], pp. 121–128), where $p_j^{(1)}=i\beta_j$ for a transversely isotropic material with this material symmetry.

If the upper and lower materials are exchanged, then $\mathbf{D}$ in (6) remains the same but $\mathbf{W}$ in (7) changes sign. In carrying out the analysis, this reverses the sign of w in (10), and hence ϵ in (9).

Asymptotic stress and displacement fields in the neighborhood of the crack tip in Fig. 1 are given in [9]. The stress intensity factors K_1 and K_2 may be combined as a complex stress intensity factor K given by

$$K=K_1+iK_2 \tag{13}$$

such that the stress components along the interface ahead of the crack tip are related as

$$\left(\sqrt{\frac{D_{22}}{D_{11}}}\sigma_{22}+i\sigma_{12}\right)\Bigg|_{\theta=0}=\frac{Kr^{i\epsilon}}{\sqrt{2\pi r}}. \tag{14}$$

Further, the complex stress intensity factor in (13) may be written in nondimensional form as

$$\tilde{K}=\frac{KL^{i\epsilon}}{\sigma\sqrt{\pi L}} \tag{15}$$

where L is an arbitrary length parameter and σ is the applied stress. The nondimensional complex stress intensity factor may also be expressed as

$$\tilde{K}=|\tilde{K}|e^{i\psi} \tag{16}$$

so that the phase angle

$$\psi=\arctan\left[\frac{\mathrm{Im}\{KL^{i\epsilon}\}}{\mathrm{Re}\{KL^{i\epsilon}\}}\right]=\arctan\left[\sqrt{\frac{D_{11}}{D_{22}}}\left(\frac{\sigma_{12}}{\sigma_{22}}\right)\right]\Bigg|_{\theta=0,r=L}. \quad (17)$$

Crack-face displacements in the vicinity of the crack tip in terms of K are found to be

$$\sqrt{\frac{D_{11}}{D_{22}}}\,\Delta u_2+i\Delta u_1=\frac{2D_{11}}{(1+2i\epsilon)\cosh\pi\epsilon}\sqrt{\frac{r}{2\pi}}\,Kr^{i\epsilon} \quad (18)$$

where $\Delta u_j=u_j^{(1)}(r,\pi)-u_j^{(2)}(r,-\pi)$. The superscripts 1 and 2 on u_j refer to material 1 and 2, respectively.

The interface energy release rate $\mathcal{G}_i$ is related to the stress intensity factors by

$$\mathcal{G}_i=\frac{1}{H}(K_1^2+K_2^2) \quad (19)$$

where

$$\frac{1}{H}=\frac{D_{11}}{4\cosh^2\pi\epsilon}. \quad (20)$$

Note that the subscript i in (19) represents interface and $\mathcal{G}_i$ has units of force per length.

It should be noted that inherently for any interface both K_1 and K_2 must be prescribed or equivalently $\mathcal{G}_i$ and ψ. In describing an interface fracture criterion, one may prescribe either a relation between K_1 and K_2 or express the critical energy release rate $\mathcal{G}_{ic}$ as a function of the phase angle ψ.

3 A Finite Interface Crack

The bimaterial system described above is considered. Let Γ denote a crack of length $2a$ which is located at

$$x_2=\pm 0, \quad -a<x_1<+a$$

as in Fig. 2(*a*). Following [11] (p. 140), a stress function vector $\mathbf{\Phi}$ may be defined with components Φ_j such that

$$\sigma_{j1}=-\Phi_{j,2} \quad \text{and} \quad \sigma_{j2}=\Phi_{j,1} \quad (21)$$

where the comma denotes differentiation with respect to x_1 or x_2. In terms of the stress function and displacement vectors ($\mathbf{\Phi}$ and $\mathbf{u}$), boundary and continuity conditions for this problem are

$$\frac{\partial\mathbf{\Phi}^{(1)}}{\partial x_1}=\frac{\partial\mathbf{\Phi}^{(2)}}{\partial x_1}=0 \quad \text{for } x_2=\pm 0, \quad |x_1|<a \quad (22)$$

$$\frac{\partial\mathbf{\Phi}^{(1)}}{\partial x_1}=\frac{\partial\mathbf{\Phi}^{(2)}}{\partial x_1}, \quad \mathbf{u}^{(1)}=\mathbf{u}^{(2)} \quad \text{for } x_2=0, \quad |x_1|>a. \quad (23)$$

Here the superscripts 1 and 2 refer to materials 1 and 2, respectively. Equation (22) represents zero traction on the crack faces and (23) is continuity of tractions and displacements across the interface.

Further, stresses $\sigma_{11}^{\infty(1)}$, $\sigma_{11}^{\infty(2)}$, σ_{22}^{∞}, and σ_{12}^{∞} are prescribed at infinity. From force and moment equilibrium considerations, stresses σ_{22}^{∞} and σ_{12}^{∞} are the same in both materials. The stress component σ_{11}^{∞} is, however, different in the two materials ($\sigma_{11}^{\infty(1)}$ and $\sigma_{11}^{\infty(2)}$); these are related by equating the strain ϵ_{11} across the interface as shown later in this section.

Displacement and stress function vectors for the upper material can be written using remote stresses and strains ([11], pp. 378–380) as

$$\mathbf{u}^{(1)}=\mathrm{Re}\{e^{\pi\epsilon}\mathbf{A}_1\langle f_0(z_*^{(1)},\epsilon)\rangle\mathbf{B}_1^{-1}\mathbf{h}+e^{-\pi\epsilon}\mathbf{A}_1\langle\overline{f}_0(z_*^{(1)},\epsilon)\rangle\mathbf{B}_1^{-1}\overline{\mathbf{h}}\}+x_1\boldsymbol{\epsilon}_1^{\infty(1)}+x_2\boldsymbol{\epsilon}_2^{\infty(1)} \quad (24)$$

$$\mathbf{\Phi}^{(1)}=\mathrm{Re}\{e^{\pi\epsilon}\mathbf{B}_1\langle f_0(z_*^{(1)},\epsilon)\rangle\mathbf{B}_1^{-1}\mathbf{h}+e^{-\pi\epsilon}\mathbf{B}_1\langle\overline{f}_0(z_*^{(1)},\epsilon)\rangle\mathbf{B}_1^{-1}\overline{\mathbf{h}}\}+x_1\mathbf{t}_2^{\infty(1)}-x_2\mathbf{t}_1^{\infty(1)} \quad (25)$$

respectively. Vectors $\mathbf{t}_1^{\infty(1)}$ and $\mathbf{t}_2^{\infty(1)}$ contain information about stresses at infinity in material 1, while strains at infinity are incorporated in $\boldsymbol{\epsilon}_1^{\infty(1)}$ and $\boldsymbol{\epsilon}_2^{\infty(1)}$. They are evaluated as explained in [11] (pp. 378–380) and are given later in this section. The bar over a quantity represents its complex conjugate. The function $f_0(z,\epsilon)$ is continuous everywhere except at Γ and vanishes at infinity. It is obtained by solving a Hilbert problem. For the mathematically degenerate lower material [9],

$$\mathbf{u}^{(2)}=\mathrm{Re}\{e^{-\pi\epsilon}\mathbf{A}_2'\mathbf{F}_0(z^{(2)},\epsilon)\mathbf{B}_2'^{-1}\mathbf{h}+e^{\pi\epsilon}\mathbf{A}_2'\overline{\mathbf{F}}_0(z^{(2)},\epsilon)\mathbf{B}_2'^{-1}\overline{\mathbf{h}}\}+x_1\boldsymbol{\epsilon}_1^{\infty(2)}+x_2\boldsymbol{\epsilon}_2^{\infty(2)} \quad (26)$$

$$\mathbf{\Phi}^{(2)}=\mathrm{Re}\{e^{-\pi\epsilon}\mathbf{B}_2'\mathbf{F}_0(z^{(2)},\epsilon)\mathbf{B}_2'^{-1}\mathbf{h}+e^{\pi\epsilon}\mathbf{B}_2'\overline{\mathbf{F}}_0(z^{(2)},\epsilon)\mathbf{B}_2'^{-1}\overline{\mathbf{h}}\}+x_1\mathbf{t}_2^{\infty(2)}-x_2\mathbf{t}_1^{\infty(2)} \quad (27)$$

where

$$\mathbf{F}_0(z^{(2)},\epsilon)=\begin{bmatrix} f_0(z^{(2)},\epsilon) & x_2f_0'(z^{(2)},\epsilon) & 0\\ 0 & f_0(z^{(2)},\epsilon) & 0\\ 0 & 0 & f_0(z^{(2)},\epsilon)\end{bmatrix}$$

and $(\,)'$ represents differentiation with respect to z. As before $\mathbf{t}_1^{\infty(2)}$ and $\mathbf{t}_2^{\infty(2)}$ are related to the stresses and $\boldsymbol{\epsilon}_1^{\infty(2)}$ and $\boldsymbol{\epsilon}_2^{\infty(2)}$ are related to the strains at infinity in material 2, and are presented later in this section.

Using the traction continuity condition in (23) as $|x_1|\to\infty$,

$$\mathbf{t}_2^{\infty(1)}=\mathbf{t}_2^{\infty(2)}.$$

Defining these vectors using complex conjugate scalars t_γ and $\overline{t}_\gamma$ as

$$\mathbf{t}_2^{\infty(1)}=\mathbf{t}_2^{\infty(2)}=t_\gamma\mathbf{h}+\overline{t}_\gamma\overline{\mathbf{h}} \quad (28)$$

and satisfying (22) leads to

$$e^{\pi\epsilon}g_0^+(x_1,\epsilon)+e^{-\pi\epsilon}g_0^-(x_1,\epsilon)=-2t_\gamma \quad (29)$$

$$e^{-\pi\epsilon}\overline{g}_0^+(x_1,\epsilon)+e^{\pi\epsilon}\overline{g}_0^-(x_1,\epsilon)=-2\overline{t}_\gamma \quad (30)$$

where

$$g_0(z,\epsilon)=\frac{\partial}{\partial z}f_0(z,\epsilon). \quad (31)$$

Solving the Hilbert problem in (29) and (30) as explained by Muskhelishvili [19],

$$g_0(z,\epsilon)=\frac{-\mathcal{X}(z,\epsilon)}{\pi i}\int_\Gamma\frac{e^{-\pi\epsilon}t_\gamma d\lambda}{\mathcal{X}^+(\lambda,\epsilon)(\lambda-z)}+P(z,\epsilon)\mathcal{X}(z,\epsilon) \quad (32)$$

$$\overline{g}_0(z,\epsilon)=\frac{-\overline{\mathcal{X}}(z,\epsilon)}{\pi i}\int_\Gamma\frac{e^{\pi\epsilon}\overline{t}_\gamma d\lambda}{\overline{\mathcal{X}}^+(\lambda,\epsilon)(\lambda-z)}+\overline{P}(z,\epsilon)\overline{\mathcal{X}}(z,\epsilon) \quad (33)$$

where the Plemelj function

$$\mathcal{X}(z,\epsilon)=(z+a)^{-1/2-i\epsilon}(z-a)^{-1/2+i\epsilon}. \quad (34)$$

One condition for $g_0(z,\epsilon)$ to be holomorphic at infinity is for $P(z,\epsilon)$ to be a polynomial in z of order less than 1, i.e., a constant. The first term in (32) is defined as

$$I_1\equiv\frac{-\mathcal{X}(z,\epsilon)}{\pi i}\int_\Gamma\frac{e^{-\pi\epsilon}t_\gamma d\lambda}{\mathcal{X}^+(\lambda,\epsilon)(\lambda-z)}.$$

Considering vanishingly small contours around the crack,

$$I_1 \equiv \frac{-\mathcal{X}(z,\epsilon)}{\pi i}\frac{1}{(1+e^{-2\pi\epsilon})}\oint_\Gamma \frac{e^{-\pi\epsilon}t_\gamma d\lambda}{\mathcal{X}(\lambda,\epsilon)(\lambda-z)}.$$

Now for large values of $|z|$

$$\frac{1}{\mathcal{X}(z,\epsilon)} = z + i2a\epsilon + \frac{\alpha_{-1}}{z} + \frac{\alpha_{-2}}{z^2} + \ldots$$

where α_{-1}, $\alpha_{-2}, \ldots$ are appropriate constants. Applying Cauchy's formula and considering the pole at $|z|=\infty$,

$$\oint_\Gamma \frac{d\lambda}{\mathcal{X}(\lambda,\epsilon)(\lambda-z)} = 2\pi i\left[\frac{1}{\mathcal{X}(z,\epsilon)} - (z+i2a\epsilon)\right]$$

so that

$$I_1 = \frac{-\mathcal{X}(z,\epsilon)}{\pi i}\frac{t_\gamma}{2\cosh\pi\epsilon}2\pi i\left[\frac{1}{\mathcal{X}(z,\epsilon)} - (z+i2a\epsilon)\right]$$
$$= \frac{-t_\gamma}{\cosh\pi\epsilon}[1-\mathcal{X}(z,\epsilon)(z+i2a\epsilon)].$$

Substituting in (32),

$$g_0(z,\epsilon) = \frac{-t_\gamma}{\cosh\pi\epsilon}[1-\mathcal{X}(z,\epsilon)(z+i2a\epsilon)] + P(z,\epsilon)\mathcal{X}(z,\epsilon). \tag{35}$$

Integrating the above expression

$$f_0(z,\epsilon) = \int g_0(z,\epsilon)dz$$
$$= \frac{-t_\gamma}{\cosh\pi\epsilon}[z-\mathcal{X}_p(z,\epsilon)] + \int P(z,\epsilon)\mathcal{X}(z,\epsilon)dz \tag{36}$$

where

$$\mathcal{X}_p(z,\epsilon) = (z+a)^{1/2-i\epsilon}(z-a)^{1/2+i\epsilon}. \tag{37}$$

Similarly

$$\overline{g}_0(z,\epsilon) = \frac{-\overline{t}_\gamma}{\cosh\pi\epsilon}[1-\overline{\mathcal{X}}(z,\epsilon)(z-i2a\epsilon)] + \overline{P}(z,\epsilon)\overline{\mathcal{X}}(z,\epsilon), \tag{38}$$

$$\overline{f}_0(z,\epsilon) = \frac{-\overline{t}_\gamma}{\cosh\pi\epsilon}[z-\overline{\mathcal{X}}_p(z,\epsilon)] + \int \overline{P}(z,\epsilon)\overline{\mathcal{X}}(z,\epsilon)dz, \tag{39}$$

$$\overline{\mathcal{X}}_p(z,\epsilon) = (z+a)^{1/2+i\epsilon}(z-a)^{1/2-i\epsilon}. \tag{40}$$

Imposing the condition of zero crack opening at the crack tips, $P(z,\epsilon)=0=\overline{P}(z,\epsilon)$, so that

$$f_0(z,\epsilon) = \frac{-t_\gamma}{\cosh\pi\epsilon}[z-\mathcal{X}_p(z,\epsilon)] \tag{41}$$

and

$$\overline{f}_0(z,\epsilon) = \frac{-\overline{t}_\gamma}{\cosh\pi\epsilon}[z-\overline{\mathcal{X}}_p(z,\epsilon)]. \tag{42}$$

Defining

$$\mathbf{d} = \frac{t_\gamma}{\cosh\pi\epsilon}\mathbf{h} \tag{43}$$

and using strategies explained in [11] (pp. 378–380) to incorporate uniform stresses at infinity, the displacement and stress function vectors for the two materials become

$$\mathbf{u}^{(1)} = \mathrm{Re}\{e^{\pi\epsilon}\mathbf{A}_1\langle f(z_*^{(1)},\epsilon)\rangle\mathbf{B}_1^{-1}\mathbf{d} + e^{-\pi\epsilon}\mathbf{A}_1\langle\overline{f}(z_*^{(1)},\epsilon)\rangle\mathbf{B}_1^{-1}\overline{\mathbf{d}}\}$$
$$+x_1\begin{bmatrix}\left(1-\nu_A^2\frac{E_T}{E_A}\right)\frac{1}{E_A}\sigma_{11}^{\infty(1)} - (1+\nu_T)\frac{\nu_A}{E_A}\sigma_{22}^{\infty}\\ 0\\ 0\end{bmatrix}$$
$$+x_2\begin{bmatrix}\frac{1}{G_A}\sigma_{12}^{\infty}\\ (1+\nu_T)\frac{\nu_A}{E_A}\sigma_{11}^{\infty(1)} + (1-\nu_T^2)\frac{1}{E_T}\sigma_{22}^{\infty}\\ 0\end{bmatrix} \tag{44}$$

and

$$\mathbf{\Phi}^{(1)} = \mathrm{Re}\{e^{\pi\epsilon}\mathbf{B}_1\langle f(z_*^{(1)},\epsilon)\rangle\mathbf{B}_1^{-1}\mathbf{d} + e^{-\pi\epsilon}\mathbf{B}_1\langle\overline{f}(z_*^{(1)},\epsilon)\rangle\mathbf{B}_1^{-1}\overline{\mathbf{d}}\}$$
$$+x_1\begin{bmatrix}\sigma_{12}^{\infty}\\ \sigma_{22}^{\infty}\\ 0\end{bmatrix} - x_2\begin{bmatrix}\sigma_{11}^{\infty(1)}\\ \sigma_{12}^{\infty}\\ 0\end{bmatrix} \tag{45}$$

where

$$f(z,\epsilon) = (z-a)^{1/2+i\epsilon}(z+a)^{1/2-i\epsilon} - z. \tag{46}$$

Along the x_1-axis,

$$f(x_1,\epsilon) = \begin{cases}\pm\sqrt{x_1^2-a^2}e^{i\epsilon X} - x_1 & \text{for } x_2=0, \pm x_1 > a\\ \pm i\sqrt{a^2-x_1^2}e^{\mp\epsilon X} - x_1 & \text{for } x_2=\pm 0, |x_1| < a\end{cases} \tag{47}$$

where

$$X = \ln\left|\frac{x_1-a}{x_1+a}\right|.$$

The same function in (46) was obtained by Ting [11] (p. 428), for a finite length crack with tractions applied to it. The diagonal matrix

$$\langle f(z_*^{(1)},\epsilon)\rangle = \mathrm{diag}[f(z_1^{(1)},\epsilon),\quad f(z_2^{(1)},\epsilon),\quad f(z_3^{(1)},\epsilon)]$$

where $z_j^{(1)} = x_1 + p_j^{(1)}x_2$ and $\mathbf{d}$ is a complex 3×1 vector to be determined. For the specific case at hand, transversely isotropic material with the axial direction coinciding with the x_1-axis, $p_j^{(1)} = i\beta_j$ with $j=1,2,3$. The matrices $\mathbf{A}_1$ and $\mathbf{B}_1^{-1}$ are given in Appendix A.

For the mathematically degenerate lower material,

$$\mathbf{u}^{(2)} = \mathrm{Re}\{e^{-\epsilon\pi}\mathbf{A}_2'\mathbf{F}(z^{(2)},\epsilon)\mathbf{B}_2'^{-1}\mathbf{d} + e^{\epsilon\pi}\mathbf{A}_2'\overline{\mathbf{F}}(z^{(2)},\epsilon)\mathbf{B}_2'^{-1}\overline{\mathbf{d}}\}$$
$$+x_1\begin{bmatrix}\left(\frac{1}{E_T} - \frac{\nu_A^2}{E_A}\right)\sigma_{11}^{\infty(2)} - \left(\frac{\nu_T}{E_T} + \frac{\nu_A^2}{E_A}\right)\sigma_{22}^{\infty}\\ 0\\ 0\end{bmatrix}$$
$$+x_2\begin{bmatrix}\frac{1}{G_T}\sigma_{12}^{\infty}\\ \left(\frac{\nu_T}{E_T} + \frac{\nu_A^2}{E_A}\right)\sigma_{11}^{\infty(2)} + \left(\frac{1}{E_T} - \frac{\nu_A^2}{E_A}\right)\sigma_{22}^{\infty}\\ 0\end{bmatrix} \tag{48}$$

and

$$\mathbf{\Phi}^{(2)}=\mathrm{Re}\{e^{-\epsilon\pi}\mathbf{B}_2'\mathbf{F}(z^{(2)},\epsilon)\mathbf{B}_2'^{-1}\mathbf{d}+e^{\epsilon\pi}\mathbf{B}_2'\overline{\mathbf{F}}(z^{(2)},\epsilon)\mathbf{B}_2'^{-1}\overline{\mathbf{d}}\}+x_1\begin{bmatrix}\sigma_{12}^{\infty}\\ \sigma_{22}^{\infty}\\ 0\end{bmatrix}-x_2\begin{bmatrix}\sigma_{11}^{\infty(2)}\\ \sigma_{12}^{\infty}\\ 0\end{bmatrix} \tag{49}$$

where

$$\mathbf{F}(z^{(2)},\epsilon)=\begin{bmatrix}f(z^{(2)},\epsilon) & x_2 f'(z^{(2)},\epsilon) & 0\\ 0 & f(z^{(2)},\epsilon) & 0\\ 0 & 0 & f(z^{(2)},\epsilon)\end{bmatrix}. \tag{50}$$

The function $f(z^{(2)},\epsilon)$ is defined in (46) and $(\,)'$ represents differentiation with respect to z. For a transversely isotropic material with the axial direction coinciding with the x_3-axis, $p_j^{(2)}=i$; hence $z^{(2)}=z=x_1+ix_2$.

Using (28) and (43) along with the following from [11] (p. 427),

$$\check{\mathbf{S}}\mathbf{d}=-i\beta\mathbf{d} \quad \text{and} \quad \check{\mathbf{S}}\overline{\mathbf{d}}=i\beta\overline{\mathbf{d}} \tag{51}$$

so that the vector $\mathbf{d}$ is obtained as

$$\mathbf{d}=\frac{\sqrt{1-\beta^2}}{2}\begin{bmatrix}\sigma_{12}^{\infty}-i\sqrt{\dfrac{D_{22}}{D_{11}}}\sigma_{22}^{\infty}\\ i\sqrt{\dfrac{D_{11}}{D_{22}}}\sigma_{12}^{\infty}+\sigma_{22}^{\infty}\\ 0\end{bmatrix}. \tag{52}$$

To determine K as defined in (14), stresses σ_{22} and σ_{12} on the interface ahead of the crack tip are required. These stresses, obtained from (21), are equal in the two materials as a result of the first equation in (23). Using one of the materials, say material 1, these stresses are given by

$$\begin{bmatrix}\sigma_{12}\\ \sigma_{22}\\ \sigma_{32}\end{bmatrix}_{\theta=0}=\left.\frac{\partial\mathbf{\Phi}^{(1)}}{\partial x_1}\right|_{\theta=0}=2\cosh\pi\epsilon\,\mathrm{Re}\{g(x_1,\epsilon)\mathbf{d}\}+\begin{bmatrix}\sigma_{12}^{\infty}\\ \sigma_{22}^{\infty}\\ 0\end{bmatrix}.$$

Differentiating (46) to obtain $g(x_1,\epsilon)$ and substituting in the above equation,

$$\begin{bmatrix}\sigma_{12}\\ \sigma_{22}\\ \sigma_{32}\end{bmatrix}_{\theta=0}=2\cosh\pi\epsilon\,\mathrm{Re}\{[(x_1+i2a\epsilon)(x_1+a)^{-1/2-i\epsilon}\times(x_1-a)^{-1/2+i\epsilon}-1]\mathbf{d}\}+\begin{bmatrix}\sigma_{12}^{\infty}\\ \sigma_{22}^{\infty}\\ 0\end{bmatrix}.$$

Moving the coordinate system to the right crack tip by relating $x_1=r+a$ and considering only singular terms for $r\ll a$,

$$\begin{bmatrix}\sigma_{12}\\ \sigma_{22}\\ \sigma_{32}\end{bmatrix}_{\theta=0}=2\cosh\pi\epsilon\,\mathrm{Re}\{r^{-1/2+i\epsilon}(2a)^{1/2-i\epsilon}(1/2+i\epsilon)\mathbf{d}\}.$$

Substituting for $\mathbf{d}$ from (52) in the above equation, singular components of σ_{22} and σ_{12} are obtained which along with (14) relate K to stresses at infinity as

$$K=(1+2i\epsilon)\sqrt{\pi a}(2a)^{-i\epsilon}\left[\sqrt{\frac{D_{22}}{D_{11}}}\sigma_{22}^{\infty}+i\sigma_{12}^{\infty}\right]. \tag{53}$$

If the two materials are isotropic (and not transversely isotropic), $D_{11}=D_{22}$ as $\beta_1=\beta_2=1$ in (11) and (12). Then the above expression for K agrees with Rice [1] for a finite crack along the interface of isotropic bimaterials.

Following Rice and Sih [20] for isotropic bimaterials, displacement continuity is imposed across the interface as in (23) for $|x_1|\to\infty$, and stresses $\sigma_{11}^{\infty(1)}$ and $\sigma_{11}^{\infty(2)}$ are related as

$$\sigma_{11}^{\infty(2)}=\frac{E_T}{E_A}\sigma_{11}^{\infty(1)}+\left[\frac{E_A\nu_T+E_T(\nu_A^2-\nu_A-\nu_A\nu_T)}{E_A-\nu_A^2E_T}\right]\sigma_{22}^{\infty}. \tag{54}$$

4 Infinite Array of Collinear Interface Cracks

Next, an infinite array of cracks of equal length $2a$ along the interface of the above bimaterial system is considered. They are spaced at constant intervals $2b$ $(>2a)$ as shown in Fig. 3(a). Let Γ denote the cracks which are located at

$$x_2=\pm 0, \quad 2nb-a<x_1<2nb+a, \quad n=-\infty,\ldots,0,\ldots,+\infty.$$

As before, stresses $\sigma_{11}^{\infty(1)}$, $\sigma_{11}^{\infty(2)}$, σ_{22}^{∞}, and σ_{12}^{∞} are prescribed at infinity. Again $\sigma_{11}^{\infty(1)}$ and $\sigma_{11}^{\infty(2)}$ are related using displacement continuity across the interface as $|x_1|\to\infty$. Boundary and continuity conditions in terms of stress function and displacement vectors ($\mathbf{\Phi}$ and $\mathbf{u}$) are

$$\frac{\partial\mathbf{\Phi}^{(1)}}{\partial x_1}=\frac{\partial\mathbf{\Phi}^{(2)}}{\partial x_1}=0 \quad \text{for} \quad x_2=\pm 0, \quad 2nb-a<x_1<2nb+a, \quad n=-\infty,\ldots,0,\ldots,+\infty \tag{55}$$

$$\frac{\partial\mathbf{\Phi}^{(1)}}{\partial x_1}=\frac{\partial\mathbf{\Phi}^{(2)}}{\partial x_1}, \quad \mathbf{u}^{(1)}=\mathbf{u}^{(2)} \quad \text{for} \quad x_2=0, \quad 2nb+a<x_1<2(n+1)b-a, \quad n=-\infty,\ldots,0,\ldots,+\infty \tag{56}$$

where superscripts 1 and 2 refer to materials 1 and 2, respectively.

Since the bimaterial system is the same as in Section 3, stress function and displacement vectors given in (24), (25), (26), and (27) are valid here though the function $f_0(z,\epsilon)$ is different. Proceeding as before and formulating the Hilbert problem in (29) and (30), the solution is obtained as (32) and (33) where the Plemelj function

$$\mathcal{X}(z,\epsilon)=\prod_{n=-\infty}^{\infty}[z-(2nb-a)]^{-1/2-i\epsilon}[z-(2nb+a)]^{-1/2+i\epsilon}=\prod_{n=-\infty}^{\infty}[(z-2nb)+a]^{-1/2-i\epsilon}[(z-2nb)-a]^{-1/2+i\epsilon} \tag{57}$$

and $P(z,\epsilon)$ is a polynomial in z of order less than n. Evaluating the first term in (32) by considering vanishingly small contours around the cracks leads to

$$g_0(z,\epsilon)=\frac{-t_\gamma}{\cosh\pi\epsilon}\left\{1-\mathcal{X}(z,\epsilon)\prod_{n=-\infty}^{\infty}[(z-2nb)+i2a\epsilon]\right\}+P(z,\epsilon)\mathcal{X}(z,\epsilon). \tag{58}$$

Integrating the above expression

$$f_0(z,\epsilon)=\int g_0(z,\epsilon)dz=\frac{-t_\gamma}{\cosh\pi\epsilon}[z-\mathcal{X}_p(z,\epsilon)]+\int P(z,\epsilon)\mathcal{X}(z,\epsilon)dz \tag{59}$$

where

$$\mathcal{X}_p(z,\epsilon)=\prod_{n=-\infty}^{\infty}[(z-2nb)+a]^{1/2-i\epsilon}[(z-2nb)-a]^{1/2+i\epsilon}. \tag{60}$$

Imposing the condition of zero crack opening at the crack tips,

$$f_0(z,\epsilon)=\frac{-t_\gamma}{\cosh \pi\epsilon}[z-\mathcal{X}_p(z,\epsilon)]. \tag{61}$$

Similarly

$$\overline{f_0}(z,\epsilon)=\frac{-\overline{t}_\gamma}{\cosh \pi\epsilon}[z-\overline{\mathcal{X}}_p(z,\epsilon)]. \tag{62}$$

Since the remote loads are identical to those in the previous section, the last two terms in (24), (25), (26), and (27) are the same. As before if

$$\mathbf{d}=\frac{t_\gamma}{\cosh \pi\epsilon}\mathbf{h}$$

the displacement and stress function vectors in the two materials are given by (44), (45), (48), (49), and (50) with

$$f(z,\epsilon)=\prod_{n=-\infty}^{\infty}[(z-2nb)+a]^{1/2-i\epsilon}[(z-2nb)-a]^{1/2+i\epsilon}-z. \tag{63}$$

Also the vector $\mathbf{d}$ is the same as that in (52).

Stresses σ_{22} and σ_{12} on the interface ahead of the crack tip are used to obtain K. Proceeding as in the previous section,

$$\begin{bmatrix}\sigma_{12}\\ \sigma_{22}\\ \sigma_{32}\end{bmatrix}_{\theta=0}=2\cosh \pi\epsilon\,\mathrm{Re}\{g(x_1,\epsilon)\mathbf{d}\}+\begin{bmatrix}\sigma_{12}^{\infty}\\ \sigma_{22}^{\infty}\\ 0\end{bmatrix}.$$

Substituting for $g(x_1,\epsilon)$ by differentiating (63) and neglecting the uniform stress terms,

$$\begin{bmatrix}\sigma_{12}\\ \sigma_{22}\\ \sigma_{32}\end{bmatrix}_{\theta=0}=2\cosh \pi\epsilon\,\mathrm{Re}\left\{\prod_{n=-\infty}^{\infty}[(x_1-2nb)+i2a\epsilon]\times[(x_1-2nb)+a]^{-1/2-i\epsilon}[(x_1-2nb)-a]^{-1/2+i\epsilon}\mathbf{d}\right\}. \tag{64}$$

Following Rice and Sih [20] and using

$$\sin \pi t=\pi t\prod_{n=1}^{\infty}(1-t^2/n^2),$$

$$\prod_{n=-\infty}^{\infty}[(x_1-2nb)+i2a\epsilon][(x_1-2nb)+a]^{-1/2-i\epsilon}\times[(x_1-2nb)-a]^{-1/2+i\epsilon}=\sin\left[\frac{\pi(x_1+i2a\epsilon)}{2b}\right]\left\{\sin\left[\frac{\pi(x_1+a)}{2b}\right]\right\}^{-1/2-i\epsilon}\times\left\{\sin\left[\frac{\pi(x_1-a)}{2b}\right]\right\}^{-1/2+i\epsilon}.$$

This result is substituted into (64) and the coordinate system moved to the right crack tip by relating $x_1=r+a$. Considering only singular terms for $r\ll a,b$

$$\begin{bmatrix}\sigma_{12}\\ \sigma_{22}\\ \sigma_{32}\end{bmatrix}_{\theta=0}=2\cosh \pi\epsilon\,\mathrm{Re}\left\{\sin\left[\frac{\pi a(1+i2\epsilon)}{2b}\right]\times\left[\sin\left(\frac{\pi a}{b}\right)\right]^{-1/2-i\epsilon}\left(\frac{\pi r}{b}\right)^{-1/2+i\epsilon}\mathbf{d}\right\}.$$

Using $\mathbf{d}$ from (52) in the above equation, singular components of σ_{22} and σ_{12} are obtained. They are substituted in (14) to determine K as

$$K=\sqrt{2b}\frac{1}{\sqrt{\sin(2\pi a/b)}}[\sin(\pi a/b)\cosh(2\pi a\epsilon/b)+i\cos(\pi a/b)\sinh(2\pi a\epsilon/b)]\times\exp\left\{-i\epsilon\ln\left[\frac{b}{\pi}\sin(2\pi a/b)\right]\right\}\times\left[\sqrt{\frac{D_{22}}{D_{11}}}\;\sigma_{22}^{\infty}+i\sigma_{12}^{\infty}\right]. \tag{65}$$

Again, if the two materials are isotropic (and not transversely isotropic), $D_{11}=D_{22}$ from (11) and (12). The above expression then matches that given by Murakami [21] which has been derived from relations given in [20].

Stresses $\sigma_{11}^{\infty(1)}$ and $\sigma_{11}^{\infty(2)}$ are related by the earlier expression in (54) since conditions at infinity remain the same.

5 Finite Element Calculations

The finite element method is exploited to numerically determine stress intensity factors for the geometries considered in Sections 3 and 4 and a specific material combination. To extract the K values from the numerical results a conservative area integral—the M-integral—is employed. It was presented by Banks-Sills and Boniface [9] and is described for completeness in Appendix B.

5.1 Finite Length Crack in an Infinite Body. The geometry and loading exhibited in Fig. 2(*a*) are considered. The infinite body is approximated by considering one with width and height equal to 20 times the crack length [i.e., $H=W=20a$, see Fig. 2(*b*)]. Symmetry is exploited to model half the body. The finite element mesh containing 1344 eight-noded isoparametric elements and 4195 nodal points is illustrated in Fig. 2(*b*). The mesh is finer near the crack tip and quarter-point elements are employed at the crack tip. A uniform remote tensile stress $\sigma_{22}^{\infty}=1$ Pa is applied on the upper and lower surfaces as shown. Assuming $\sigma_{11}^{\infty(1)}=1$ Pa and using (54), $\sigma_{11}^{\infty(2)}=0.601$ Pa. Finite element analyses are carried out with ADINA [22]. Since this package does not include transversely isotropic material behavior explicitly, the orthotropic scheme is used with material properties in two perpendicular directions assumed equal. While G_A is the controlling shear modulus in the upper material, it is G_T in the lower material. It is seen that a very accurate value of $G_T=E_T/2(1+\nu_T)$ must be input to ensure transerve isotropy, failing which errors creep into the subsequent processing of finite element results. Although the singularity at the crack tip is a combination of square root and oscillatory, it was found ([14]) that better results are obtained for interface cracks in bimaterial isotropic bodies when quarter-point elements are employed instead of regular eight-noded isoparametric elements.

The analyses presented in Sections 3 and 4 are for homogeneous transversely isotropic materials. In this study, homogenized effective mechanical properties of a graphite/epoxy (AS4/3501-6) fiber-reinforced material are employed. Some of these properties are presented in Table 1 ([23]). The volume fraction of the fibers is about 65 percent. In the upper material, the graphite fibers are aligned with the x_1-direction, while in the lower material they are along the x_3-direction resulting in a mathematical degeneracy

(shown schematically in Fig. 1). Parameters necessary for calculation of the stress intensity factors and energy release rate are given in Table 2.

For this configuration, K_1 and K_2 are obtained by applying the area M-integral method to the numerical results. These are compared with analytical values from (53) in Table 3. Excellent agreement is observed validating the expression in (53), as well as the M-integral calculations. These M-integral results are obtained from the third ring outside the crack tip. This ring was chosen on the basis of error estimates from the superconvergent patch recovery technique ([24,25]). While the maximum error in energy norm in the first ring is approximately 35 percent, it reduced to about 7 percent in the third ring. However, on evaluating the area M-integral in the outer rings (greater than 3), it is seen that this value does not vary by more than 0.2 percent. Since it is advantageous to minimize the number of elements in the M-integral computation, the third ring is considered optimum.

Table 1 Effective mechanical properties of a graphite/epoxy (AS4/3501-6) fiber-reinforced composite [23]

property	value
E_A	138.2 GPa
E_T	10.4 GPa
G_A	5.5 GPa
ν_A	0.3
ν_T	0.55

Table 2 Material parameters required for stress intensity factor and energy release rate calculations

parameter	value
β_1	0.624
β_2	4.896
β_3	1.280
ϵ	−0.028
D_{11} (GPa)$^{-1}$	0.231
D_{22} (GPa)$^{-1}$	0.312

Table 3 Stress intensity factors for a finite interface crack (units of K are Nm$^{-(3/2+i\epsilon)}$)

	K_1	K_2
Analytic	2.0638	−0.0749
M-integral	2.0683	−0.0754
Error	−0.2%	−0.7%

5.2 Collinear Cracks Along the Interface. Next, the geometry and loading in Fig. 3(*a*) are considered. The finite element mesh of this geometry is shown in Fig. 3(*b*). It has 1200 eight-noded isoparametric elements and 3780 nodes, and corresponds to the strip enclosed by dotted lines in Fig. 3(*a*). Values of $b=2a$ and $H=10a$ are chosen. Three different cases of remote loads are considered: (a) tension only, (b) tension and shear, and (c) shear only. The bimaterial system is made of graphite/epoxy (AS4/3501-6) fiber-reinforced composite as in the previous subsection. Effective mechanical properties and material parameters are the same as in Tables 1 and 2. For this configuration, K_1 and K_2 obtained using the area M-integral method are shown along with analytical values from (65) in Table 4. It is seen that they compare well confirming both the analytic expression given in (65), as well as the numerical analysis. Again, the area M-integral values are from the third ring from the crack tip. To study the effect of mesh refinement, the number of elements was quadrupled by halving the dimensions of each element in the coarse mesh shown in Fig. 3(*b*). Results from the fine mesh are also shown in Table 4. These values show a closer match with the analytical result compared to the earlier coarse mesh. However, the relatively coarse mesh used initially also produces reasonably accurate results.

6 Conclusion

This paper presents analytical expressions of stress intensity factors for two different crack geometries in a particular bimaterial system. This bimaterial system is made of two dissimilar transversely isotropic materials (with the lower material math-

Table 4 Stress intensity factors for an infinite array of collinear interface cracks (units of K are Nm$^{-(3/2+i\epsilon)}$ and percentage error is given in brackets)

	K_1	K_2
Case (a) : $\sigma_{22}^{\infty}=1.0$ Pa, $\sigma_{12}^{\infty}=0.0$		
Analytic	5.2124	0.0409
M-integral (coarse mesh)	5.2069 (0.1%)	0.0404 (1.2%)
M-integral (fine mesh)	5.2094 (0.1%)	0.0412 (0.7%)
Case (b) : $\sigma_{22}^{\infty}=1.0$ Pa, $\sigma_{12}^{\infty}=1.0$ Pa		
Analytic	5.1773	4.5214
M-integral (coarse mesh)	5.1707 (0.1%)	4.5138 (0.2%)
M-integral (fine mesh)	5.1737 (0.1%)	4.5187 (0.1%)
Case (c) : $\sigma_{22}^{\infty}=0.0$, $\sigma_{12}^{\infty}=1.0$ Pa		
Analytic	−0.0351	4.4805
M-integral (coarse mesh)	−0.0362 (−3.1%)	4.4734 (0.2%)
M-integral (fine mesh)	−0.0357 (−1.7%)	4.4777 (0.1%)

ematically degenerate) bonded together. Uniform loads at infinity were applied. First, stress intensity factors for a finite crack along the interface between these two materials were obtained by solution of a Hilbert problem. This result was then extended to the problem of an infinite array of collinear cracks along a similar interface.

In addition, finite element analyses for a specific fiber reinforced material were carried out on these bodies. Stress intensity factors were calculated by means of an area M-integral. These compared well with results obtained from the analytic expressions. These analytic expressions provide a simple means to understand *any* problem pertaining to this crack configuration and bimaterial system, without resorting to numerical analysis of *specific* ones. Also they can be used as benchmarks for more complicated problems.

The accuracy of the finite element computations, together with the M-integral for determining the stress intensity factors has been demonstrated earlier in [9]. This approach will next be applied to a Brazilian disk specimen composed of fiber-reinforced material in order to determine calibration equations for testing.

Acknowledgment

We would like to thank Mr. Rami Eliasi for his assistance with the finite element calculations.

Appendix A

Some Important Matrices. In this Appendix, the matrices $\mathbf{A}_j$, $\mathbf{B}_j$, and $\mathbf{B}_j^{-1}$ which appear in the paper are presented for the specific transversely isotropic materials studied here. They are related to the material properties. The subscript j represents the upper and lower materials, 1 and 2, respectively.

The x_1-direction is the axial direction of the upper material. The matrix $\mathbf{A}_1$ is given by

$$\mathbf{A}_1=-\begin{bmatrix} k_1^{(1)}Q_1 & k_2^{(1)}Q_2 & 0 \\ ik_1^{(1)}Q_3/\beta_1 & ik_2^{(1)}Q_4/\beta_2 & 0 \\ 0 & 0 & -ik_3^{(1)}/\beta_3 G_T \end{bmatrix} \tag{66}$$

where $k_j^{(1)}$, $j=1,2,3$, are normalization factors for the upper material and are given by

$$k_1^{(1)}=\frac{(1+i)}{2}\left[\frac{(1+\nu_T)^2}{\beta_1 E_T}-\frac{\beta_1^3}{E_A}\left(1-\nu_A^2\frac{E_T}{E_A}\right)\right]^{-1/2}, \tag{67}$$

$$k_2^{(1)}=\frac{(1+i)}{2}\left[\frac{(1+\nu_T)^2}{\beta_2 E_T}-\frac{\beta_2^3}{E_A}\left(1-\nu_A^2\frac{E_T}{E_A}\right)\right]^{-1/2}, \tag{68}$$

$$k_3^{(1)}=\frac{(1+i)}{2}\left[\frac{2(1+\nu_T)}{\beta_3 E_T}\right]^{-1/2}. \tag{69}$$

The constants β_j, $j=1,2,3$ are related to the three complex eigenvalues of the elastic constants $p_j^{(1)}$ ([11], pp. 121–128), where $p_j^{(1)}=i\beta_j$ for a transversely isotropic material with this material symmetry. The constants Q_j are related to the material properties as

$$Q_1=\frac{1}{E_A}[\beta_1^2(1-\nu_A^2E_T/E_A)+\nu_A(1+\nu_T)], \tag{70}$$

$$Q_2=\frac{1}{E_A}[\beta_2^2(1-\nu_A^2E_T/E_A)+\nu_A(1+\nu_T)], \tag{71}$$

$$Q_3=(1+\nu_T)[\beta_1^2\nu_A/E_A+(1-\nu_T)/E_T], \tag{72}$$

$$Q_4=(1+\nu_T)[\beta_2^2\nu_A/E_A+(1-\nu_T)/E_T]. \tag{73}$$

The material parameters E_A, E_T, G_A, G_T, ν_A, and ν_T are the usual material properties in the axial and transverse directions (namely, Young's moduli, shear moduli, and Poisson's ratios); since the material is transversely isotropic, $G_T=E_T/2(1+\nu_T)$. The matrix $\mathbf{B}_1$ is given by

$$\mathbf{B}_1=\begin{bmatrix} -ik_1^{(1)}\beta_1 & -ik_2^{(1)}\beta_2 & 0 \\ k_1^{(1)} & k_2^{(1)} & 0 \\ 0 & 0 & -k_3^{(1)} \end{bmatrix}. \tag{74}$$

Its inverse is given by

$$\mathbf{B}_1^{-1}=\frac{1}{\beta_2-\beta_1}\begin{bmatrix} -i/k_1^{(1)} & \beta_2/k_1^{(1)} & 0 \\ i/k_2^{(1)} & -\beta_1/k_2^{(1)} & 0 \\ 0 & 0 & -(\beta_2-\beta_1)/k_3^{(1)} \end{bmatrix}. \tag{75}$$

In the lower material, the axial direction coincides with the x_3-direction. The mechanical properties E_A, E_T, G_A, G_T, ν_A, and ν_T are taken to be the same as for the upper material; but they are in different coordinate directions. It turns out that this material is mathematically degenerate. It has three identical complex eigenvalues $p_j^{(2)}=i$ where the subscript $j=1,2,3$. To determine the stress and displacement fields, matrices alternative to $\mathbf{A}_2$ and $\mathbf{B}_2$ are required; these are $\mathbf{A}_2'$ and $\mathbf{B}_2'$. Since

$$\mathbf{AB}^{-1}=\mathbf{A}'\mathbf{B}'^{-1} \tag{76}$$

it is possible to calculate $\boldsymbol{\epsilon}$ with the aid of (8). On the other hand, one may determine $\mathbf{A}_2\mathbf{B}_2^{-1}$ without calculating the individual matrices ([11], p. 173). For brevity, only the primed matrices are presented as

$$\mathbf{A}_2'=\begin{bmatrix} k_1^{(2)} & -ik_1^{(2)}\kappa & 0 \\ ik_1^{(2)} & -k_1^{(2)}\kappa & 0 \\ 0 & 0 & k_3^{(2)} \end{bmatrix} \tag{77}$$

$$\mathbf{B}_2'=\begin{bmatrix} 2iG_Tk_1^{(2)} & G_Tk_1^{(2)} & 0 \\ -2G_Tk_1^{(2)} & -iG_Tk_1^{(2)} & 0 \\ 0 & 0 & iG_Ak_3^{(2)} \end{bmatrix} \tag{78}$$

and

$$\mathbf{B}_2'^{-1}=\begin{bmatrix} -i/(4G_Tk_1^{(2)}) & -1/(4G_Tk_1^{(2)}) & 0 \\ 1/(2G_Tk_1^{(2)}) & i/(2G_Tk_1^{(2)}) & 0 \\ 0 & 0 & -i/(G_Ak_3^{(2)}) \end{bmatrix} \tag{79}$$

where

$$\kappa=\frac{3-\nu_T-4\nu_A^2E_T/E_A}{2(1+\nu_T)}. \tag{80}$$

The orthogonalization employed to obtain the normalization factors $k_1^{(2)}$ and $k_3^{(2)}$ may be found in [11] (pp. 489–492) and [26]; they are given by

$$k_1^{(2)}=\frac{1}{4\sqrt{E_T(1-\nu_A^2E_T/E_A)}}, \tag{81}$$

$$k_3^{(2)}=\frac{(1-i)}{2\sqrt{G_A}}. \tag{82}$$

Appendix B

Area M-Integral. In this Appendix, the path-independent M-integral is described. The M-integral was introduced by Yau et al. [27] for separation of mixed modes in homogeneous bodies. It was converted to an area integral for isotropic bimaterials by Shih and Asaro [28]. It may be written as

$$M^{(1,2)}=\int_A\left[\sigma_{ij}^{(1)}\frac{\partial u_i^{(2)}}{\partial x_1}+\sigma_{ij}^{(2)}\frac{\partial u_i^{(1)}}{\partial x_1}-W^{(1,2)}\delta_{1j}\right]\frac{\partial q_1}{\partial x_j}dA. \tag{83}$$

In (83), indicial notation is employed, the superscripts (1) and (2) represent two solutions and δ is the Kronecker delta. The mutual strain energy density $W^{(1,2)}$ of the two solutions is given by

$$W^{(1,2)}=\sigma_{ij}^{(1)}\epsilon_{ij}^{(2)}=\sigma_{ij}^{(2)}\epsilon_{ij}^{(1)}. \tag{84}$$

The function q_1 is defined for finite element analysis as

$$q_1=\sum_{m=1}^{8}N_m(\xi,\eta)q_{1m} \tag{85}$$

where N_m are the finite element shape functions of an eight-noded isoparametric element and ξ and η are the coordinates in the parent element ([29]). The calculation of the *M*-integral is carried out in a ring of elements surrounding the crack tip (the area *A* in (83)). The elements within the ring move as a rigid body. For each of these elements q_1 is unity; so that, the derivative of q_1 with respect to x_j is zero. For all elements outside the ring, q_1 is zero; so that, again the derivative of q_1 is zero. For elements belonging to the ring, the vector q_{1m} in (85) is chosen so that the virtual crack extension does not disturb the relative nodal point positions in their new locations; for example, a regular element with nodes at the mid-sides contains only midside nodes after distortion. The relationship

$$M^{(1,2)}=\frac{2}{H}[K_1^{(1)}K_1^{(2)}+K_2^{(1)}K_2^{(2)}] \tag{86}$$

may also be obtained.

In (83) and (86), problem (1) is that for which a solution is sought. Two auxiliary solutions are required in order to determine both $K_1^{(1)}$ and $K_2^{(1)}$ for this problem. These are denoted as $(2a)$ and $(2b)$.

For solution $(2a)$, choose $K_1^{(2a)}=1$ and $K_2^{(2a)}=0$. Such a solution does exist for some special loading. Equation (86) becomes

$$M^{(1,2a)}=\frac{2}{H}K_1^{(1)} \tag{87}$$

and from (83)

$$M^{(1,2a)}=\int_A\left[\sigma_{ij}^{(1)}\frac{\partial u_i^{(2a)}}{\partial x_1}+\sigma_{ij}^{(2a)}\frac{\partial u_i^{(1)}}{\partial x_1}-W^{(1,2a)}\delta_{1j}\right]\frac{\partial q_1}{\partial x_j}dA. \tag{88}$$

The displacements required for solution (1) are taken from a finite element analysis of the problem to be solved; the stresses and strains are calculated from these. Asymptotic expressions for the stresses, strains, and displacements for solution $(2a)$ are employed. These stresses and displacements are presented in [9].

For solution $(2b)$, $K_1^{(2b)}=0$ and $K_2^{(2b)}=1$. Equation (86) becomes

$$M^{(1,2b)}=\frac{2}{H}K_2^{(1)} \tag{89}$$

and from (83)

$$M^{(1,2b)}=\int_A\left[\sigma_{ij}^{(1)}\frac{\partial u_i^{(2b)}}{\partial x_1}+\sigma_{ij}^{(2b)}\frac{\partial u_i^{(1)}}{\partial x_1}-W^{(1,2b)}\delta_{1j}\right]\frac{\partial q_1}{\partial x_j}dA. \tag{90}$$

After calculating these integrals, $K_1^{(1)}$ and $K_2^{(1)}$ are found by equating (87) and (88), (89), and (90).

References

[1] Rice, J. R., 1988, "Elastic Fracture Mechanics Concepts for Interfacial Cracks," ASME J. Appl. Mech., **55**, pp. 98–103.

[2] Hutchinson, J. W., 1990, "Mixed-Mode Fracture Mechanics of Interfaces," *Metal-Ceramic Interface*, M. Rühle, A. G. Evans, M. F. Ashby, and J. P. Hirth, eds., Pergamon Press, Oxford, pp. 295–301.

[3] Bassani, J. L., and Qu, J., 1989, "Finite Crack on Bimaterial and Bicrystal Interfaces," J. Mech. Phys. Solids, **37**, pp. 434–453.

[4] Qu, J., and Bassani, J. L., 1989, "Cracks on Bimaterial and Bicrystal Interfaces," J. Mech. Phys. Solids, **37**, pp. 417–434.

[5] Suo, Z., 1990, "Singularities, Interfaces and Cracks in Dissimilar Anisotropic Media," Proc. R. Soc. London, Ser. A, **427**, 331–358.

[6] Ting, T. C. T., 1986, "Explicit Solution and Invariance of the Singularities at an Interface Crack in Anisotropic Composites," *International Journal of Solids and Structures*, **22**, pp. 965–983.

[7] Ting, T. C. T., 1990, "Interface Cracks in Anisotropic Bimaterials," J. Mech. Phys. Solids **38**, pp. 505–513.

[8] Ting, T. C. T., 1992, "Interface Cracks on Anisotropic Elastic Bimaterials— A Decomposition Principle," Int. J. Solids Struct., **29**, pp. 1989–2003.

[9] Banks-Sills, L., and Boniface, V., 2000, "Fracture Mechanics of an Interface Crack Between a Special Pair of Transversely Isotropic Materials," *Multi-Scale Deformation and Fracture in Materials and Structures*, The James R. Rice 60th Anniversary Volume, to appear.

[10] Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," Philos. Mag., **7**, pp. 625–646.

[11] Ting, T. C. T., 1996, *Anisotropic Elasticity–Theory and Applications*, Oxford University Press, Oxford.

[12] Wang, S. S., and Yau, J. F., 1981, "Interfacial Cracks in Adhesively Bonded Scarf Joints," AIAA J., **19**, pp. 1350–1356.

[13] Matos, P. P. L., McMeeking, R. M., Charalambides, P. G., and Drory, M. D., 1989, "A method for Calculating Stress Intensities in Bimaterial Fracture," Int. J. Fract., **40**, pp. 235–254.

[14] Banks-Sills, L., Travitzky, N., Ashkenazi, D., and Eliasi, R., 1999, "A Methodology for Measuring Interface Fracture Properties of Composite Materials," Int. J. Fract., **99**, pp. 143–161.

[15] Charalambides, P. G., and Zhang, W., 1996, "An Energy Method for Calculating the Stress Intensities in Orthotropic Bimaterial Fracture," Int. J. Fract., **76**, pp. 97–120.

[16] Nakamura, T., 1991, "Three-Dimensional Stress Fields of Elastic Interface Cracks," ASME J. Appl. Mech., **58**, pp. 939–946.

[17] Nahta, R., and Moran, B., 1993, "Domain Integrals for Axisymmetric Interface Crack Problems," Int. J. Solids Struct., **30**, pp. 2027–2040.

[18] Gosz, M., Dolbow, J., and Moran, B., 1998, "Domain Integral Formulation for Stress Intensity Factor Computation Along Curved Three-Dimensional Interface Cracks," *International Journal of Solids and Structures*, **35**, pp. 1763–1783.

[19] Muskhelishvili, N. I., 1953, *Some Basic Problems of the Mathematical Theory of Elasticity*, P. Noordhoff Ltd., Groningen, Holland, pp. 427–450.

[20] Rice, J. R., and Sih, G. C., 1965, "Plane Problems of Cracks in Dissimilar Media," ASME J. Appl. Mech., **32**, pp. 418–423.

[21] Murakami, Y., 1988, *Stress Intensity Factors Handbook*, Vol. 1, Pergamon Press, Oxford, p. 440.

[22] Bathe, K. J., 1999, *ADINA–Automatic Dynamic Incremental Nonlinear Analysis System*, Version 7.3, Adina Engineering, Inc., USA.

[23] Pagano, N., 1999, private communication.

[24] Zienkiewicz, O. C., and Zhu, J. Z., 1992, "The Superconvergent Patch Recovery and a posteriori Error Estimates. Part 1: The Recovery Technique," Int. J. Numer. Methods Eng., **33**, pp. 1331–1364.

[25] Thurston, M. E., and Zehnder, A. T., 1996, "Nickel-Alumina Interfacial Fracture Toughness: Experiments and Analysis of Residual Stress Effects," Int. J. Fract., **76**, pp. 221–241.

[26] Ting, T. C. T., and Hwu, C., 1988, "Sextic Formalism in Anisotropic Elasticity for Almost Non-semisimple Matrix **N**," Int. J. Solids Struct., **24**, pp. 65–76.

[27] Yau, J. F., Wang, S. S., and Corten, H. T., 1980, "A Mixed-Mode Crack Analysis of Isotropic Solids Using Conservation Laws of Elasticity," ASME J. Appl. Mech., **47**, pp. 335–341.

[28] Shih, C. F., and Asaro, R. J., 1988, "Elastic-Plastic Analysis of Cracks on Bimaterial Interfaces: Part I—Small Scale Yielding," ASME J. Appl. Mech., **55**, pp. 299–316.

[29] Banks-Sills, L., and Sherman, D., 1992, "On the Computation of Stress Intensity Factors for Three-Dimensional Geometries by Means of the Stiffness Derivative and J-Integral Methods," Int. J. Fract., **53**, pp. 1–20.

V. Parameswaran
Assistant Professor,
Department of Ship Technology,
Cochin University of Science and Technology,
Kochi 682016, India
Mem. ASME

A. Shukla[1]
Simon Ostrach Professor,
Department of Mechanical Engineering and
Applied Mechanics,
University of Rhode Island,
Kingston, RI 02881
Fellow ASME

Asymptotic Stress Fields for Stationary Cracks Along the Gradient in Functionally Graded Materials

Stress field for stationary cracks, aligned along the gradient, in functionally graded materials is obtained through an asymptotic analysis coupled with Westergaard's stress function approach. The first six terms of the stress field are obtained for both opening mode and shear mode loading. It is observed that the structure of the terms other than $r^{-1/2}$ and r^0 are influenced by the nonhomogeneity. Using this stress field, contours of constant maximum shear stress are generated and the effect of nonhomogeneity on these contours is discussed. [DOI: 10.1115/1.1459072]

1 Introduction

Several researchers have established that the near-tip stress fields around cracks in functionally graded materials (FGMs) retain the inverse square root behavior ([1,2]). Without assuming any specific form of property variation, Eischen [3] has shown that the angular functions associated with the first two terms ($r^{-1/2}$ and r^0) of the stress field in nonhomogeneous materials are not affected by the material property variation and the effect of nonhomogeneity reflects only in higher order terms. However, explicit form of the higher order terms is not yet established for FGMs, hence, one has to use the expansions available for homogeneous materials in order to extract fracture parameters from full field experimental data. This could lead to serious errors, as the stress field in FGMs is identical to that of homogeneous materials only very close to the crack tip.

In the present work, stress field for stationary cracks aligned along the direction of property variation in an FGM with exponentially varying elastic modulus is developed through an asymptotic analyses coupled with Westergaard's stress function approach ([4]). The first six terms of the expansion for opening mode and shear mode loading are obtained. Using this stress field, contours of constant maximum shear stress are generated for two different levels of nonhomogeniety and the effect of nonhomogeneity on these contours is discussed.

2 Theoretical Formulation

The elastic modulus of the FGM ($E(x)$) is assumed to vary exponentially along the line of the crack as given in (1) and the Poisson's ratio (ν) is assumed to be constant.

$$E(x)=E_c \exp(\lambda x) \tag{1}$$

E_c is the modulus at the crack tip ($x=0$) and λ is the nonhomogeneity parameter having dimension (length)$^{-1}$. Defining the in plane stress components ($\sigma_{ij}, i,j \in \{x,y\}$) in terms of the Airy's stress function $F(x,y)$ and using Hooke's law, the compatibility equation takes the following form:

$$\nabla^2(\nabla^2 F)-2\lambda\frac{\partial}{\partial x}(\nabla^2 F)+\lambda^2\nabla^2 F-\lambda^2(1+\nu)\frac{\partial^2 F}{\partial y^2}=0,$$

$$\text{where } \nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}. \tag{2}$$

The solution to (2) is obtained in a series form through an asymptotic analysis discussed in the next section.

3 Asymptotic Analysis

The crack-tip coordinates are scaled to fill the entire field of observation using the transformation

$$\eta_1=\frac{x}{\varepsilon}, \quad \text{and} \quad \eta_2=\frac{y}{\varepsilon}, \quad 0<\varepsilon\leqslant 1. \tag{3}$$

In the scaled coordinates (η_1, η_2), (2) takes the form

$$\nabla^2(\nabla^2 F)-2\lambda\varepsilon\frac{\partial}{\partial \eta_1}(\nabla^2 F)+\lambda^2\varepsilon^2\nabla^2 F-\lambda^2\varepsilon^2(1+\nu)\frac{\partial^2 F}{\partial \eta_2^2}=0,$$

$$\text{where } \nabla^2=\frac{\partial^2}{\partial \eta_1^2}+\frac{\partial^2}{\partial \eta_2^2}. \tag{4}$$

It is assumed at this stage that the stress function $F(\varepsilon\eta_1, \varepsilon\eta_2)$ can be expanded in powers of the parameter ε as follows:

$$F(\varepsilon\eta_1,\varepsilon\eta_2)=\sum_{m=0}^{\infty}\varepsilon^{(m+3/2)}\phi_m(\eta_1,\eta_2)+\sum_{n=0}^{\infty}\varepsilon^{(n+2)}\psi_n(\eta_1,\eta_2). \tag{5}$$

Of the two series the first one corresponding to ϕ_m is the singular series and the second one containing ψ_n corresponds to that for finiteness of the domain ([5]). Substitution of this solution into (4) leads to a series of differential equations each associated with a specific power of the parameter ε as given in (6).

$$\sum_{m=0}^{\infty}\left\{\varepsilon^{(m+3/2)}\nabla^2(\nabla^2\phi_m)-\varepsilon^{(m+5/2)}2\lambda\frac{\partial}{\partial\eta_1}(\nabla^2\phi_m)\right.$$
$$\left.+\varepsilon^{(m+7/2)}\left(\lambda^2\nabla^2\phi_m-\lambda^2(1+\nu)\frac{\partial^2\phi_m}{\partial\eta_2^2}\right)\right\}$$
$$+\sum_{n=0}^{\infty}\left\{\varepsilon^{(n+2)}\nabla^2(\nabla^2\psi_n)-\varepsilon^{(n+3)}2\lambda\frac{\partial}{\partial\eta_1}(\nabla^2\psi_n)\right.$$
$$\left.+\varepsilon^{(n+4)}\left(\lambda^2\nabla^2\psi_n-\lambda^2(1+\nu)\frac{\partial^2\psi_n}{\partial\eta_2^2}\right)\right\}=0 \tag{6}$$

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, February 1, 2001; final revision, November 16, 2001. Associate Editor: K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

For (6) to be valid the differential equations corresponding to each power of $\varepsilon(\varepsilon^{3/2},\varepsilon^2,\varepsilon^{5/2}\ldots)$ should vanish independently. This leads to the following set of differential equations:

$$m=n=0,\quad \nabla^2(\nabla^2\phi_m)=0,\quad \nabla^2(\nabla^2\psi_n)=0 \tag{7}$$

$$m=n=1,\quad \nabla^2(\nabla^2\phi_m)-2\lambda\frac{\partial}{\partial\eta_1}(\nabla^2\phi_{m-1})=0,$$
$$\nabla^2(\nabla^2\psi_n)-2\lambda\frac{\partial}{\partial\eta_1}(\nabla^2\psi_{n-1})=0 \tag{8}$$

$$m,n>1,\qquad \nabla^2(\nabla^2\phi_m)-2\lambda\frac{\partial}{\partial\eta_1}(\nabla^2\phi_{m-1})+\lambda^2\nabla^2\phi_{m-2}-\lambda^2(1+\nu)\frac{\partial^2\phi_{m-2}}{\partial\eta_2^2}=0,$$
$$\nabla^2(\nabla^2\psi_n)-2\lambda\frac{\partial}{\partial\eta_1}(\nabla^2\psi_{n-1})+\lambda^2\nabla^2\psi_{n-2}-\lambda^2(1+\nu)\frac{\partial^2\psi_{n-2}}{\partial\eta_2^2}=0.$$

It should be noticed at this stage that (7) is identical to that for homogeneous materials for which the solutions exist ([4]). The differential Eqs. (8) and (9), corresponding to the higher powers of ε, are coupled to the lower-order functions. These are solved in a recursive manner and the solutions for opening mode and shear mode are provided in the following sections.

4 Opening Mode Loading

Considering the symmetry of the normal stress components about the line of the crack and the traction-free crack-face boundary conditions, the first three terms in the expansion ($m=0,1,2$) are

$$\phi_0=\mathrm{Re}\{\bar{\bar{G}}_0\}+\eta_2\,\mathrm{Im}\{\bar{G}_0\},$$
$$\phi_1=\mathrm{Re}\{\bar{\bar{G}}_1\}+\eta_2\,\mathrm{Im}\{\bar{G}_1\}-\frac{\lambda}{2}\eta_2^2\,\mathrm{Re}\{\bar{G}_0\}$$
$$\phi_2=\mathrm{Re}\{\bar{\bar{G}}_2\}+\eta_2\,\mathrm{Im}\{\bar{G}_2\}-\frac{\lambda}{2}\eta_2^2\,\mathrm{Re}\{\bar{G}_1\}-\frac{(3-\nu)\lambda^2}{24}\eta_2^3\,\mathrm{Im}\{\bar{G}_0\} \tag{10}$$
$$\bar{\bar{G}}_0=\frac{4}{3}A_0\zeta^{3/2},\quad \bar{\bar{G}}_1=\frac{4}{15}A_1\zeta^{5/2}\quad\text{and}\quad \bar{\bar{G}}_2=\frac{4}{35}A_2\zeta^{7/2},$$
$$\zeta=\eta_1+i\eta_2,\quad \bar{G}_0=\frac{\partial\bar{\bar{G}}_0}{\partial\zeta},\quad i=\sqrt{-1}.$$

$\bar{\bar{G}}_0$, $\bar{\bar{G}}_1$, and $\bar{\bar{G}}_2$ are the first, second, and third terms in the series solution of the stress function for homogeneous materials ([4]). For $n=0$, 1, and 2 the solutions can be written in terms of the functions $\bar{H}_0$, $\bar{H}_1$, and $\bar{H}_2$, which are the first three terms of the series solution ([5]) for homogeneous materials, as

$$\psi_0=\eta_2\,\mathrm{Im}\{\bar{H}_0\}$$
$$\psi_1=\eta_2\,\mathrm{Im}\{\bar{H}_1\}-\frac{\lambda}{2}\eta_2^2\,\mathrm{Re}\{\bar{H}_0\}$$
$$\psi_2=\eta_2\,\mathrm{Im}\{\bar{H}_2\}-\frac{\lambda}{2}\eta_2^2\,\mathrm{Re}\{\bar{H}_1\}-\frac{(3-\nu)\lambda^2}{24}\eta_2^3\,\mathrm{Im}\{\bar{H}_0\}-\frac{(1+\nu)\lambda^2}{8}\eta_2^2\,\mathrm{Re}\{\bar{\bar{H}}_0\} \tag{11}$$
$$\text{and}\quad \bar{\bar{H}}_0=\frac{1}{2}B_0\zeta^2,\quad \bar{H}_0=B_0\zeta,\quad \bar{H}_1=\frac{1}{2}B_1\zeta^2$$
$$\text{and}\quad \bar{H}_2=\frac{1}{3}B_3\zeta^3.$$

In (10) and (11), A_n and B_n are real constants. Switching back to the (x, y) coordinates through (3) and (5), a six-term expansion for the stresses can be obtained as

$$\sigma_{xx}=\sum_{n=0}^{2}\{\mathrm{Re}\{Z_n\}-y\,\mathrm{Im}\{Z_n'\}+2\,\mathrm{Re}\{Y_n\}-y\,\mathrm{Im}\{Y_n'\}\}-\sum_{n=0}^{1}\left\{\frac{\lambda}{2}\{2\,\mathrm{Re}\{\bar{Z}_n\}-4y\,\mathrm{Im}\{Z_n\}-y^2\,\mathrm{Re}\{Z_n'\}+2\,\mathrm{Re}\{\bar{Y}_n\}-4y\,\mathrm{Im}\{Y_n\}-y^2\,\mathrm{Re}\{Y_n'\}\}\right\}-\frac{(3-\nu)}{24}\lambda^2\{6y\,\mathrm{Im}\{\bar{Z}_0\}+6y^2\,\mathrm{Re}\{Z_0\}-y^3\,\mathrm{Im}\{Z_0'\}+6y\,\mathrm{Im}\{\bar{Y}_0\}+6y^2\,\mathrm{Re}\{Y_0\}-y^3\times\mathrm{Im}\{Y_0'\}\}-\frac{(1+\nu)}{8}\lambda^2\{2\,\mathrm{Re}\{\bar{\bar{Y}}_0\}-4y\,\mathrm{Im}\{\bar{Y}_0\}-y^2\,\mathrm{Re}\{Y_0\}\}$$

$$\sigma_{yy}=\sum_{n=0}^{2}\{\mathrm{Re}\{Z_n\}+y\,\mathrm{Im}\{Z_n'\}+y\,\mathrm{Im}\{Y_n'\}\}-\sum_{n=0}^{1}\left\{\frac{\lambda}{2}\{y^2\,\mathrm{Re}\{Z_n'\}+y^2\,\mathrm{Re}\{Y_n'\}\}\right\}-\frac{(3-\nu)}{24}\lambda^2\{y^3\,\mathrm{Im}\{Z_0'\}+y^3\,\mathrm{Im}\{Y_0'\}\}-\frac{(1+\nu)}{8}\lambda^2y^2\,\mathrm{Re}\{Y_0\}$$

$$\sigma_{xy}=\sum_{n=0}^{2}-\{y\,\mathrm{Re}\{Z_n'\}+y\,\mathrm{Re}\{Y_n'\}+\mathrm{Im}\{Y_n\}\}+\sum_{n=0}^{1}\left\{\frac{\lambda}{2}\{2y\,\mathrm{Re}\{Z_n\}-y^2\,\mathrm{Im}\{Z_n'\}+2y\,\mathrm{Re}\{Y_n\}-y^2\,\mathrm{Im}\{Y_n'\}\}\right\}+\frac{(3-\nu)}{24}\lambda^2\{3y^2\,\mathrm{Im}\{Z_0\}+y^3\,\mathrm{Re}\{Z_0'\}+3y^2\,\mathrm{Im}\{Y_0\}+y^3\,\mathrm{Re}\{Y_0'\}\}+\frac{(1+\nu)}{8}\lambda^2\{2y\,\mathrm{Re}\{\bar{Y}_0\}-y^2\,\mathrm{Im}\{Y_0\}\}. \tag{12}$$

The complex functions Z_n and Y_n, which are counterparts of the functions G_n and H_n in the (x, y) coordinate system, are listed in the Appendix. Unlike homogeneous materials, Eqs. (12) for the stress field contain functions of the Poisson's ratio. This is due to the presence of Poisson's ratio-dependent coefficients in the governing Eq. (2) as opposed to the bi-harmonic equation for homogeneous materials. Note that the stresses collapse to their homogeneous counterparts on setting the nonhomogeneity parameter λ to zero.

5 Shear Mode Loading

Following the same procedure, outlined in the previous section and keeping in view the inherent nature of the shear mode problem, the first six terms of the expansion for the stress field are obtained as

$$\sigma_{xx}=\sum_{n=0}^{2}\{2\,\mathrm{Im}\{Z_n\}+y\,\mathrm{Re}\{Z_n'\}+\mathrm{Im}\{Y_n\}+y\,\mathrm{Re}\{Y_n'\}\}$$
$$-\sum_{n=0}^{1}\left\{\frac{\lambda}{2}\{2\,\mathrm{Im}\{\overline{Z}_n\}+4y\,\mathrm{Re}\{Z_n\}-y^2\,\mathrm{Im}\{Z_0'\}+2\,\mathrm{Im}\{\overline{Y}_n\}\right.$$
$$\left.+4y\,\mathrm{Re}\{Y_n\}-y^2\,\mathrm{Im}\{Y_0'\}\}\right\}+\frac{(3-\nu)}{24}\lambda^2\{6y\,\mathrm{Re}\{\overline{Z}_0\}-6y^2$$
$$\times\mathrm{Im}\{Z_0\}-y^3\,\mathrm{Re}\{Z_0'\}+6y\,\mathrm{Re}\{\overline{Y}_0\}-6y^2\,\mathrm{Im}\{Y_0\}-y^3$$
$$\times\mathrm{Re}\{Y_0'\}\}-\frac{(1+\nu)}{8}\lambda^2\{2\,\mathrm{Im}\{\bar{\bar{Z}}_0\}+4y\,\mathrm{Re}\{\overline{Z}_0\}-y^2\mathrm{Im}\{Z_0\}\}$$

$$\sigma_{yy}=\sum_{n=0}^{2}\{-y\,\mathrm{Re}\{Z_n'\}+\mathrm{Im}\{Y_n\}-y\,\mathrm{Re}\{Y_n'\}\}$$
$$-\sum_{n=0}^{1}\left\{\frac{\lambda}{2}\{y^2\,\mathrm{Im}\{Z_n'\}-y^2\,\mathrm{Im}\{Y_n'\}\}\right\}$$
$$+\frac{(3-\nu)}{24}\lambda^2\{y^3\,\mathrm{Re}\{Z_0'\}+y^3\,\mathrm{Re}\{Y_0'\}\}$$
$$-\frac{(1+\nu)}{8}\lambda^2\{y^2\,\mathrm{Im}\{Z_0\}\}$$

$$\sigma_{xy}=\sum_{n=0}^{2}\{\mathrm{Re}\{Z_n\}-y\,\mathrm{Im}\{Z_n'\}-y\,\mathrm{Im}\{Y_n'\}\}+\sum_{n=0}^{1}\left\{\frac{\lambda}{2}\{2y\,\mathrm{Im}\{Z_n\}\right.$$
$$\left.+y^2\,\mathrm{Re}\{Z_n'\}+2y\,\mathrm{Im}\{Y_n\}+y^2\,\mathrm{Re}\{Y_n'\}\}\right\}$$
$$-\frac{(3-\nu)}{24}\lambda^2\{3y^2\,\mathrm{Re}\{Z_0\}-y^3\,\mathrm{Im}\{Z_0'\}+3y^2\,\mathrm{Re}\{Y_0\}$$
$$-y^3\,\mathrm{Im}\{Y_0'\}\}+\frac{(1+\nu)}{8}\lambda^2\{2y\,\mathrm{Im}\{\overline{Z}_0\}+y^2\,\mathrm{Re}\{Z_0\}\}. \quad (13)$$

6 Discussion on Solutions

In order to visualize the influence of nonhomogeneity on the structure of crack-tip stress fields, contours of constant maximum shear stress (isochromatics) were generated for different levels of nonhomogeneity using (12) for opening mode. The contours are generated using two values of λ ($\lambda=0.7$ and $\lambda=-0.7$), for which the elastic modulus varies by a factor of 2 and 1/2, respectively, over a distance of $h=1$ m. The constant A_0 is expressed in terms of the stress intensity factor, K_I, as $A_0=K_I/\sqrt{2\pi}$. The constants A_1, A_2, B_0, B_1, and B_2 are usually obtained by fitting the stress field to experimental data. For generating the contours the coefficients A_1, A_2, B_1, and B_2 are set to zero, however, the nonhomogeneity specific parts of the high-order terms corresponding to A_0 are retained.

Figure 1 shows the opening mode isochromatics corresponding to a K_I of 2 MPa$\sqrt{}$m for the two values of λ and for homogeneous material ($\lambda=0$). The isochromatics very near the crack tip are identical to those in homogeneous materials. However, as we move away from the crack tip the shape of the contours deviate from their homogeneous counterparts substantially. The deviations are in the fringe tilt angle and the fringe apogee point (point of maximum radius). The fringes tilt towards the stiffer side, i.e., a positive value of λ results in a forward (away from the crack face) tilt where as a negative λ leads to a backward (towards the crack face) tilt. The apogee is seen to increase (note the second order fringe) as the magnitude of the nonhomogeneity parameter $|\lambda|$ increases.

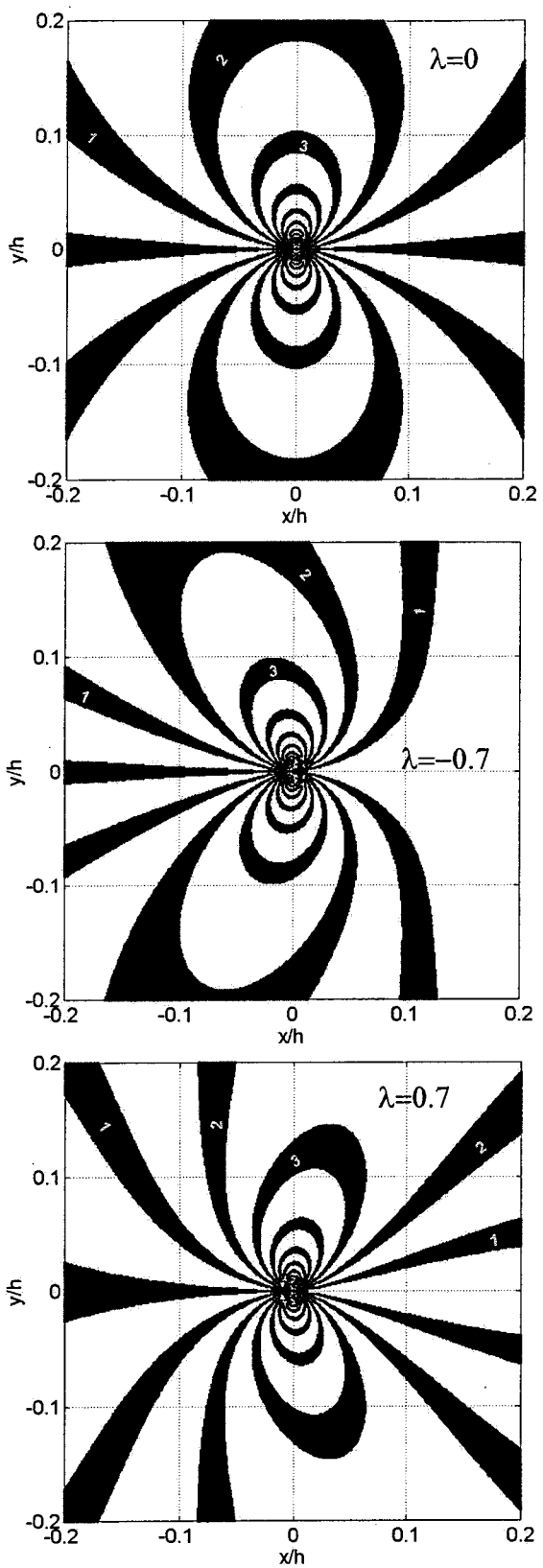


Fig. 1 Opening mode isochromatics for a functionally graded material (FGM) with exponentially varying elastic modulus

7 Closure

Higher-order terms in the expansion for crack-tip stress field in FGMs are derived for Mode I and Mode II cracks. Using this stress field, contours of constant maximum shear stress are gener-

ated for two different levels of nonhomogeneity. The contours are observed to deviate from those for homogeneous material and the deviation increases as the distance from crack tip increases.

Appendix

$$\bar{\bar{Z}}_0=\frac{4}{3}A_0z^{3/2},\quad \overline{Z}_0=2A_0z^{1/2},\quad Z_0=A_0z^{-1/2},$$

$$Z_0'=-\frac{1}{2}A_0z^{-3/2}$$

$$\bar{\bar{Z}}_1=\frac{4}{15}A_1z^{5/2},\quad \overline{Z}_1=\frac{2}{3}A_1z^{3/2},\quad Z_1=A_1z^{1/2},\quad Z_1'=\frac{1}{2}A_1z^{-1/2}$$

$$\bar{\bar{Z}}_2=\frac{4}{35}A_2z^{7/2},\quad \overline{Z}_2=\frac{2}{5}A_2z^{5/2},\quad Z_2=A_2z^{3/2},\quad Z_2'=\frac{3}{2}A_2z^{1/2}$$

$$\bar{\bar{Y}}_0=\frac{1}{2}B_0z^2,\quad \overline{Y}_0=B_0z,\quad Y_0=B_0$$

$$\overline{Y}_1=\frac{1}{2}B_1z^2,\quad Y_1=B_1z,\quad Y_1'=B_1$$

$$\overline{Y}_2=\frac{1}{3}B_2z^3,\quad Y_2=B_2z^2,\quad Y_2'=2B_2z$$

$$z=x+iy,\quad i=\sqrt{-1}$$

References

[1] Delale, F., and Erdogan, F., 1983, ''The Crack Problem for a Nonhomogeneous Plane,'' ASME J. Appl. Mech., **50**, pp. 609–614.

[2] Schovanec, L., 1986, ''A Griffith Crack Problem for an Inhomogeneous Elastic Material,'' Acta Mech., **58**, pp. 67–80.

[3] Eischen, J. W., 1987, ''Fracture of Nonhomogeneous Materials,'' Int. J. Fract., **34(3)**, pp. 3–22.

[4] Westergaard, H. M., 1939, ''Bearing Pressures and Cracks,'' ASME J. Appl. Mech., **6**, pp. A49–A53.

[5] Sanford, R. J., 1979, ''A Critical Re-examination of the Westergaard Method for Solving Opening Mode Problem,'' Mech. Res. Commun., **6(5)**, pp. 289–294.

The Interface Crack Problem of Bonded Piezoelectric and Elastic Half-Space Under Transient Electromechanical Loads

S. A. Meguid

X. Zhao

Engineering Mechanics and Design Laboratory,
Department of Mechanical and Industrial Engineering,
University of Toronto,
5 King's College Road,
Toronto, ON M5S 3G8, Canada

The interface crack problem of bonded piezoelectric and elastic half-space under transient electromechanical loads is considered. Both the permeable and impermeable boundary conditions are examined and discussed. Based on the use of integral trans form techniques, the problem is reduced either to a singular integral equation for the permeable boundary condition or to two coupled singular integral equations for the impermeable boundary condition, which can be solved using Chebyshev polynomial expansions. Numerical results are provided to show the effect of the applied electric fields, the electric boundary conditions along the crack faces and a free surface on the resulting dynamic stress intensity factor and electric displacement intensity factor. [DOI: 10.1115/1.1460910]

1 Introduction

With the increasing usage of piezoelectric materials and composites as actuating and sensing devices in smart structures, where dynamic loading is dominant, much attention has been paid to their dynamic fracture behavior. Shindo and Ozawa [1] first investigated the steady-state dynamic response of cracked piezoelectric materials under the action of incident plane harmonic waves. The dynamic Green's functions for anisotropic piezoelectric materials were derived by Norris [2]. Khutoryansky and Sosa [3] proposed dynamic representation formulas and fundamental solutions for piezoelectricity. Shindo et al. [4] studied the dynamic response of a cracked dielectric medium under the action of harmonic waves in a uniform electric field. In their recent works, Narita and Shindo [5,6] investigated the dynamic antiplane shear of a cracked piezoelectric ceramic and the scattering of Love waves by a surface-breaking crack in a piezoelectric layer over an elastic half plane. Meguid and Wang [7] studied the dynamic anti-plane interaction of two cracks in a piezoelectric medium under incident shear wave loading using the conducting crack assumption. Wang [8] further investigated the interaction of multiple interface cracks between two piezoelectric mediums.

All of the above-mentioned references are concerned with the steady-state dynamic responses of cracked piezoelectric materials and composites. However, piezoelectric materials and composites are often subjected to the action of transient dynamic loads as well in engineering applications. It is, therefore, of great importance to investigate the transient response of cracked piezoelectric materials and composites. Li and Mataga [9,10] studied the problem of a semi-infinite crack propagating in an infinite piezoelectric medium. They investigated the effect of the propagating velocity of the crack on the crack-tip fields. By the use of integral transforms and Copson-Sih's method, Chen and Yu [11], Chen and Karihaloo [12] investigated the transient response of a finite crack in an infinite piezoelectric medium under the action of antiplane mechanical loads and in-plane electric displacements. Chen and Meguid [13], Wang and Yu [14], and Shin et al. [15] considered the dynamic crack problem in a piezoelectric strip under electromechanical impact. Wang et al. [16] analyzed a cracked piezoelectric laminae subjected to electromechanical impact loads.

From the analyses concerning the transient response of cracked piezoelectric materials and composites, it can be seen that there at least exist three concerns that require further study. The first is concerned with the effect of the applied electromechanical loads on the crack tip fields. Analogous to the analyses in elastic theory, Chen and Yu [11], Chen and Karihaloo [12], Chen and Meguid [13], Wang and Yu [14], and Wang et al. [16] assumed that both mechanical loads and electric displacements are applied to the crack faces, and that the loads are *independent* of the externally applied electric fields. However, this assumption is inappropriate for piezoelectric materials. In fact, the application of electric fields will induce stresses in a piezoelectric material. When a crack is present, the stresses will be applied as external loads to the crack faces. That is to say, the loads on the crack faces are dependent on the applied electric fields. Therefore, the results in those references may be misleading. The second is concerned with the effect of the electric boundary conditions on the crack-tip fields. The modeling of electric boundary conditions along the crack faces is still an open problem. Generally, there are two well-accepted electric boundary conditions, namely: the permeable and impermeable boundary conditions. From the physical viewpoint, those two electric boundary conditions are the two extreme cases, with the permeable boundary condition representing the case where the crack faces are in complete contact and the impermeable boundary condition representing the case where the crack is open and filled with vacuum. Most of the existing works address the impermeable boundary condition. Although Wang and Yu [14] have also discussed the permeable boundary condition, and concluded that the two boundary conditions lead to the same results, the current authors have doubt about the validity of this finding. The third is concerned with the effect of the presence of a free surface on the crack-tip fields. Crack-tip fields will change due to the arrival of the reflecting waves from a free surface, but one cannot see this basic requirement from the existing studies. It is for those reasons that we offer the present study.

In this paper, we consider the problem of bonded piezoelectric and elastic half-space with an interface crack subjected to transient electromechanical loads. Both the permeable and impermeable boundary conditions are discussed. The main purpose is to

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. 1, 2001; final revision, Nov. 26, 2001. Associate Editor: K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

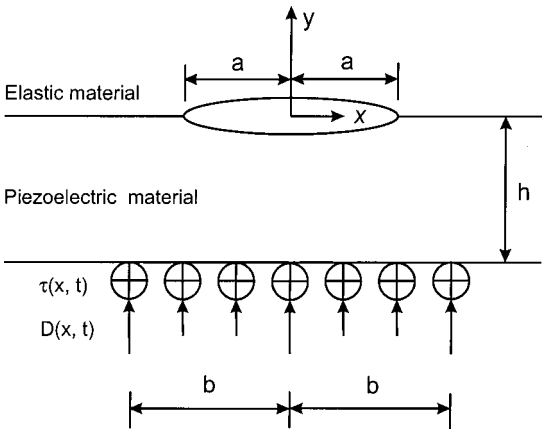


Fig. 1 Geometric configuration of the problem

provide an analytical treatment to investigate the transient behavior of piezoelectric composites under different boundary conditions, which is important for the design and numerical simulation of smart structures. Based on the use of integral transforms, the problem is reduced either to a singular integral equation for the permeable boundary condition or to two coupled singular integral equations for the impermeable boundary condition, which can be solved using Chebyshev polynomial expansions. Numerical results are provided to show the effect of applied electric fields, electric boundary conditions along the crack faces, and a free surface on the resulting dynamic stress intensity factor and electric displacement intensity factor.

2 Formulation of the Problem

Suppose that a crack of length $2a$ is present along the interface of a piezoelectric layer and the elastic half space, as shown in Fig. 1. A set of Cartesian coordinates (x,y,z) is attached to the center of the crack. The x-axis is directed along the crack line and y-axis is perpendicular to it. The poled piezoelectric strip, with the z-axis being the poling direction, occupies the region $(-h<y<0, -\infty<x<+\infty)$. At the time $t=0$, both the antiplane load and in-plane electric displacement suddenly begin to act on the lower surface of the piezoelectric strip, resulting in a coupled electric and stress wave field.

In this configuration, the piezoelectric boundary value problem is simplified considerably because only the out-of-plane displacement and the in-plane electric fields exist. The constitutive relation for the piezoelectric material can be expressed as

$$\tau_{xz}=c_{44}\frac{\partial w}{\partial x}+e_{15}\frac{\partial \phi}{\partial x},\quad \tau_{yz}=c_{44}\frac{\partial w}{\partial y}+e_{15}\frac{\partial \phi}{\partial y} \tag{1}$$

and

$$D_x=e_{15}\frac{\partial w}{\partial x}-\kappa_{11}\frac{\partial \phi}{\partial x},\quad D_y=e_{15}\frac{\partial w}{\partial y}-\kappa_{11}\frac{\partial \phi}{\partial y} \tag{2}$$

where τ_{xz} and τ_{yz} are the shear stress components, D_x and D_y are the electric displacements, w and ϕ are the mechanical displacement and electric potential, while c_{44}, e_{15}, and κ_{11} are the elastic modulus, the piezoelectric constant, and the dielectric constant of the piezoelectric material, respectively.

The equilibrium equation and the Maxwell equation for the piezoelectric material under antiplane loading are given by

$$\frac{\partial \tau_{xz}}{\partial x}+\frac{\partial \tau_{yz}}{\partial y}=\rho\frac{\partial^2 w}{\partial t^2} \tag{3}$$

$$\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}=0 \tag{4}$$

where ρ is the density of the piezoelectric material.

Substituting Eqs. (1) and (2) into (3) and (4) results in the following governing equations:

$$\nabla^2 w=c_2^{-2}\partial^2 w/\partial t^2,\quad \kappa_{11}\nabla^2\phi=e_{15}\nabla^2 w \tag{5}$$

in which

$$c_2=\sqrt{\mu/\rho},\quad \mu=c_{44}+e_{15}^2/\kappa_{11}. \tag{6}$$

The constitutive relation for the elastic material can be written as

$$\tau_{xz1}=c_{441}\frac{\partial w_1}{\partial x},\quad \tau_{yz1}=c_{441}\frac{\partial w_1}{\partial y} \tag{7}$$

where τ_{xz1} and τ_{yz1} are the shear stress components, w_1 and c_{441} are the displacement and the elastic modulus, respectively. The governing equation is given by

$$\nabla^2 w_1=c_{21}^{-2}\partial^2 w_1/\partial t^2 \tag{8}$$

in which $c_{21}=\sqrt{c_{441}/\rho_1}$, and ρ_1 is the density of the elastic material.

In the theoretical studies of crack problems, the permeable and impermeable conditions are extensively used. For the present case, it is assumed that the surface of the elastic material is grounded, so that the boundary conditions for the permeable crack problem can be written as

$$\tau_{yz}(x,-h,t)=\tau(x,t),\quad -\infty<x<\infty \tag{9a}$$

$$D_y(x,-h,t)=D(x,t),\quad -\infty<x<\infty \tag{9b}$$

$$\tau_{yz}(x,0,t)=\tau_{yz1}(x,0,t),\quad -\infty<x<\infty \tag{9c}$$

$$\tau_{yz}(x,0,t)=0,\quad |x|<a \tag{9d}$$

$$w(x,0,t)=w_1(x,0,t),\quad |x|>a \tag{9e}$$

$$\phi(x,0,t)=0,\quad -\infty<x<\infty. \tag{9f}$$

The boundary conditions for the impermeable crack problem can be expressed as

$$\tau_{yz}(x,-h,t)=\tau(x,t),\quad -\infty<x<\infty \tag{10a}$$

$$D_y(x,-h,t)=D(x,t),\quad -\infty<x<\infty \tag{10b}$$

$$\tau_{yz}(x,0,t)=\tau_{yz1}(x,0,t),\quad -\infty<x<\infty \tag{10c}$$

$$\tau_{yz}(x,0,t)=0,\quad |x|<a \tag{10d}$$

$$w(x,0,t)=w_1(x,0,t),\quad |x|>a \tag{10e}$$

$$D_y(x,0,t)=0,\quad |x|<a \tag{10f}$$

$$\phi(x,0,t)=0,\quad |x|>a. \tag{10g}$$

The analysis is performed using Laplace and Fourier transforms over time and space, respectively. The Laplace transform over time, t, and its inverse are defined by

$$f^*(p)=\int_0^\infty f(t)\exp(-pt)dt,\quad f(t)=\frac{1}{2\pi i}\int_{\mathrm{Br}} f^*(p)\exp(pt)dp \tag{11}$$

in which Br stands for the Bromwich path of integration and p is the transform variable. The time-dependency in (5) and (8) are eliminated by the application of Eq. (11). The Fourier transforms are then applied, resulting in

$$w^*(x,y,p)=\frac{1}{2\pi}\int_{-\infty}^{\infty}[A_1(\xi,p)\exp(-\gamma y)+A_2(\xi,p)\exp(\gamma y)]\exp(-i\xi x)d\xi \tag{12}$$

$$\phi^*(x,y,p)=\frac{e_{15}}{\kappa_{11}}w^*(x,y,p)+\psi^*(x,y,p) \tag{13}$$

$$\psi^*(x,y,p)=\frac{1}{2\pi}\int_{-\infty}^{\infty}[A_3(\xi,p)\exp(-|\xi|y)+A_4(\xi,p)\exp(|\xi|y)]\exp(-i\xi x)d\xi \tag{14}$$

$$w_1^*(x,y,p)=\frac{1}{2\pi}\int_{-\infty}^{\infty}A_5(\xi,p)\exp(-\gamma_1 y)\exp(-i\xi x)d\xi. \tag{15}$$

The stresses and electric displacements in Laplace transform are found to be

$$\tau_{xz}^*(x,y,p)=-\frac{\mu i}{2\pi}\int_{-\infty}^{\infty}\xi[A_1(\xi,p)\exp(-\gamma y)+A_2(\xi,p)\exp(\gamma y)]\exp(-i\xi x)d\xi - \frac{e_{15}i}{2\pi}\int_{-\infty}^{\infty}\xi[A_3(\xi,p)\exp(-|\xi|y)+A_4(\xi,p)\exp(|\xi|y)]\exp(-i\xi x)d\xi \tag{16}$$

$$\tau_{yz}^*(x,y,p)=\frac{\mu}{2\pi}\int_{-\infty}^{\infty}\gamma[-A_1(\xi,p)\exp(-\gamma y)+A_2(\xi,p)\exp(\gamma y)]\exp(-i\xi x)d\xi + \frac{e_{15}}{2\pi}\int_{-\infty}^{\infty}|\xi|[-A_3(\xi,p)\exp(-|\xi|y)+A_4(\xi,p)\exp(|\xi|y)]\exp(-i\xi x)d\xi \tag{17}$$

$$D_x^*(x,y,p)=\frac{\kappa_{11}i}{2\pi}\int_{-\infty}^{\infty}\xi[A_3(\xi,p)\exp(-|\xi|y)+A_4(\xi,p)\exp(|\xi|y)]\exp(-i\xi x)d\xi \tag{18}$$

$$D_y^*(x,y,p)=-\frac{\kappa_{11}}{2\pi}\int_{-\infty}^{\infty}|\xi|[-A_3(\xi,p)\exp(-|\xi|y)+A_4(\xi,p)\exp(|\xi|y)]\exp(-i\xi x)d\xi \tag{19}$$

$$\tau_{xz1}^*(x,y,p)=-\frac{c_{441}i}{2\pi}\int_{-\infty}^{\infty}A_5(\xi,p)\xi\exp(-\gamma_1 y)\exp(-i\xi x)d\xi \tag{20}$$

$$\tau_{yz1}^*(x,y,p)=-\frac{c_{441}}{2\pi}\int_{-\infty}^{\infty}A_5(\xi,p)\gamma_1\exp(-\gamma_1 y)\exp(-i\xi x)d\xi. \tag{21}$$

In the above formulation, $A_j(\xi,p)(j=1,2,3,4,5)$ are unknown functions, which will be determined from boundary conditions (9) and (10), and

$$\gamma=\sqrt{\xi^2+p^2c_2^{-2}},\quad \gamma_1=\sqrt{\xi^2+p^2c_{21}^{-2}}. \tag{22}$$

3 Solution of the Permeable Crack Problem

In this section, we consider the permeable crack problem. The Laplace transform of boundary conditions (9*a*)–(9*f*) can be expressed as

$$\tau_{yz}^*(x,-h,p)=\tau^*(x,p),\quad -\infty<x<\infty \tag{23a}$$

$$D_y^*(x,-h,p)=D^*(x,p),\quad -\infty<x<\infty \tag{23b}$$

$$\tau_{yz}^*(x,0,p)=\tau_{yz1}^*(x,0,p),\quad -\infty<x<\infty \tag{23c}$$

$$\tau_{yz}^*(x,0,p)=0,\quad |x|<a \tag{23d}$$

$$w^*(x,0,p)=w_1^*(x,0,p),\quad |x|>a \tag{23e}$$

$$\phi^*(x,0,p)=0,\quad -\infty<x<\infty. \tag{23f}$$

Substituting (12)–(15), (17), (19), and (21) into the boundary conditions (23*a*)–(23*c*) and (23*f*), we find

$$A_2(\xi,p)=\exp(2\gamma h)A_1(\xi,p)+\frac{\exp(\gamma h)}{\mu\gamma}\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right) \tag{24}$$

$$A_3(\xi,p)=-\frac{e_{15}[1+\exp(2\gamma h)]}{\kappa_{11}[1+\exp(2|\xi|h)]}A_1(\xi,p)+\frac{\exp(|\xi|h)\overline{D}}{\kappa_{11}|\xi|[1+\exp(2|\xi|h)]}-\frac{e_{15}\exp(\gamma h)}{\kappa_{11}\mu\gamma[1+\exp(2|\xi|h)]}\times\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right) \tag{25}$$

$$A_4(\xi,p)=-\frac{e_{15}[1+\exp(2\gamma h)]}{\kappa_{11}[1+\exp(2|\xi|h)]}\exp(2|\xi|h)A_1(\xi,p)-\frac{\exp(|\xi|h)\overline{D}}{\kappa_{11}|\xi|[1+\exp(2|\xi|h)]}-\frac{e_{15}\exp[(\gamma+2|\xi|)h]}{\kappa_{11}\mu\gamma[1+\exp(2|\xi|h)]}\times\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right) \tag{26}$$

$$A_5(\xi,p)=\frac{\mu\gamma[1-\exp(2\gamma h)][1+\exp(2|\xi|h)]+e_{15}^2/\kappa_{11}|\xi|[1+\exp(2\gamma h)][\exp(2|\xi|h)-1]}{c_{441}\gamma_1[1+\exp(2|\xi|h)]}A_1(\xi,p)+\frac{2e_{15}|\xi|\exp(|\xi|h)\overline{D}}{\kappa_{11}c_{441}|\xi|\gamma_1[1+\exp(2|\xi|h)]}+\frac{\exp(\gamma h)}{c_{441}\gamma_1}\left\{\frac{e_{15}^2|\xi|[\exp(2|\xi|h)-1]}{\kappa_{11}\mu\gamma[1+\exp(2|\xi|h)]}-1\right\}\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right) \tag{27}$$

where

$$\overline{\tau}=\int_{-\infty}^{\infty}\tau^*(x,p)\exp(i\xi x)dx \tag{28}$$

$$\overline{D}=\int_{-\infty}^{\infty}D^*(x,p)\exp(i\xi x)dx. \tag{29}$$

The remaining unknown $A_1(\xi,p)$ may then be determined from the mixed boundary conditions (23*d*) and (23*e*). To reduce the mixed boundary conditions into an integral equation, we first define the following new dislocation function:

$$\varphi(x,p)=\frac{\partial}{\partial x}[w_1^*(x,0,p)-w^*(x,0,p)]. \tag{30}$$

Then, from (23*e*) we have

$$\int_{-a}^{a}\varphi(\alpha,p)d\alpha=0. \tag{31}$$

Substitution of (12) and (15) into (30) yields

$$\varphi(x,p)=-\frac{1}{2\pi}\int_{-\infty}^{\infty}i\xi[A_5(\xi,p)-A_1(\xi,p)-A_2(\xi,p)]\times\exp(-i\xi x)d\xi. \tag{32}$$

From the above equation and the definition of Fourier transform, we obtain

$$A_5(\xi,p)-A_1(\xi,p)-A_2(\xi,p)=-\frac{1}{i\xi}\int_{-a}^{a}\varphi(\alpha,p)\exp(i\xi\alpha)d\alpha. \tag{33}$$

Further, from (24), (26), and (33), we have

$$A_1(\xi,p)=-\frac{c_{441}\gamma_1[1+\exp(2|\xi|h)]}{i\xi F_1(\xi,p)}\int_{-a}^{a}\varphi(\alpha,p)\exp(i\xi\alpha)d\alpha+\frac{c_{441}\gamma_1[1+\exp(2|\xi|h)]F_2(\xi,p)}{F_1(\xi,p)} \tag{34}$$

where

$$F_1(\xi,p)=[1+\exp(2|\xi|h)]\{\mu\gamma[1-\exp(2\gamma h)]-c_{441}\gamma_1[1+\exp(2\gamma h)]\}+\frac{e_{15}^2}{\kappa_{11}}|\xi|[1+\exp(2\gamma h)]\times[\exp(2|\xi|h)-1] \tag{35}$$

$$F_2(\xi,p)=\left\{\frac{1}{\mu\gamma}+\frac{1}{c_{441}\gamma_1}+\frac{e_{15}^2|\xi|[1-\exp(2|\xi|h)]}{\kappa_{11}c_{441}\mu\gamma\gamma_1[1+\exp(2|\xi|h)]}\right\}\times\exp(\gamma h)\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right)-\frac{2e_{15}\exp(|\xi|h)\overline{D}}{\kappa_{11}c_{441}\gamma_1[1+\exp(2|\xi|h)]}. \tag{36}$$

By using Eq. (23*d*), it is shown that

$$\int_{-\infty}^{\infty}\frac{\gamma_1F(\xi,p)\exp(-i\xi x)}{i\xi}\int_{-a}^{a}\varphi(\alpha,p)\exp(i\xi\alpha)d\alpha d\xi=\int_{-\infty}^{\infty}F_3(\xi,p)\gamma_1\exp(-i\xi x)d\xi,\quad |x|<a \tag{37}$$

where

$$F(\xi,p)=\frac{\mu\gamma[1+\exp(2|\xi|h)][1-\exp(2\gamma h)]+e_{15}^2/\kappa_{11}|\xi|[\exp(2|\xi|h)-1][1+\exp(2\gamma h)]}{F_1(\xi,p)} \tag{38}$$

$$F_3(\xi,p)=\frac{2\exp(\gamma h)[1+\exp(2|\xi|h)]}{F_1(\xi,p)}\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right)-\frac{2e_{15}\exp(|\xi|h)[1+\exp(2\gamma h)]\overline{D}}{\kappa_{11}F_1(\xi,p)}. \tag{39}$$

The right-hand side of (37) represents the load acting on the crack faces. From (39), one can see that the load depends on the electric displacement $\overline{D}$. We further notice that when $p\to 0$, $F_3(\xi,p)$ is independent of $\overline{D}$. From the initial-value theorem concerning Laplace transform, this indicates that only in the static case will the load become independent of the electric field. Therefore, the assumption that the mechanical loads acting on the crack faces are independent of the applied electric fields is inappropriate. It is also shown that the dynamic response of cracked piezoelectric materials and composites will present different behaviors from the static solution. In the static case, the electric field has no influence on the crack-tip fields when the permeable condition is considered.

The kernel of the infinite integration in Eq. (37) tends to a constant when $|\xi|\to\infty$, which corresponds to the singular part of the integration. After performing the appropriate asymptotic analysis, we find

$$\lim_{|\xi|\to\infty}\frac{\gamma_1F(\xi,p)}{\xi}=\frac{c_{44}}{c_{44}+c_{441}}\mathrm{sign}(\xi). \tag{40}$$

In order to extract the singular part from the integration, we employ the term $c_{44}\,\mathrm{sign}(\xi)/(c_{44}+c_{441})$ in the kernel. By interchanging the integration order in Eq. (37) and after the appropriate treatment, the governing equation for determining the unknown function $\varphi(\alpha,p)$ is obtained in terms of the following singular integral equation:

$$\int_{-a}^{a}\frac{\varphi(\alpha,p)}{\alpha-x}d\alpha+\int_{-a}^{a}k(\alpha,x,p)\varphi(\alpha,p)d\alpha=\frac{c_{44}+c_{441}}{2c_{44}}\int_{-\infty}^{\infty}F_3(\xi,p)\gamma_1\exp(-i\xi x)d\xi,\quad |x|<a \tag{41}$$

where

$$k(\alpha,x,p)=\int_0^{\infty}\left[\frac{(c_{44}+c_{441})\gamma_1F(\xi,p)}{c_{44}\xi}-1\right]\sin[\xi(\alpha-x)]d\xi. \tag{42}$$

It is clear that Eq. (31) is satisfied.

Equation (41) is a singular integral equation of the first kind, its solution includes the well-known square-root singularity and can be expressed as

$$\varphi(\alpha,p)=\sum_{j=0}^{\infty}\frac{B_j(p)}{\sqrt{1-\alpha^2/a^2}}T_j(\alpha/a) \tag{43}$$

where $T_j(\alpha/a)$ are Chebyshev polynomials of the first kind and $B_j(p)$ are unknown functions. From the orthogonality conditions of Chebyshev polynomials, Eq. (31) leads to $B_0(p)=0$. Substituting Eq. (43) into (41), the following algebraic equation for $B_j(p)$ is obtained:

$$\sum_{j=1}^{\infty}B_j(p)U_{j-1}(x/a)+\sum_{j=1}^{\infty}B_j(p)L_j(x,p)=\tau_1(x,p),\quad |x|<a \tag{44}$$

where $U_j(x/a)$ represent Chebyshev polynomials of the second kind, with

$$L_j(x,p)=\int_{-a}^{a}\frac{1}{\pi\sqrt{1-\alpha^2/a^2}}k(\alpha,x,p)T_j(\alpha/a)d\alpha \quad (45)$$

and

$$\tau_1(x,p)=\frac{c_{44}+c_{441}}{2\pi c_{44}}\int_{-\infty}^{\infty}F_3(\xi,p)\gamma_1\exp(-i\xi x)d\xi. \quad (46)$$

Truncating the Chebyshev polynomials in Eq. (43) to the Nth term and assuming that Eq. (44) is satisfied at N collocation points along the crack faces,

$$x_m=a\cos\left(\frac{m\pi}{N+1}\right),\quad m=1,2,\ldots,N. \quad (47)$$

Equation (44) can be reduced to a linear algebraic system of equations of the following form:

$$\sum_{j=1}^{N}B_j(p)\sin\left(\frac{mj\pi}{N+1}\right)\Big/\sin\left(\frac{m\pi}{N+1}\right)+\sum_{j=1}^{N}B_j(p)L_j(x_m,p)$$
$$=\tau_1(x_m,p),\quad m=1,2,\ldots,N. \quad (48)$$

Once $B_j(p)$ are determined from (48), the stress components can be obtained. Then, the dynamic stress intensity factor can be evaluated using the following expressions:

$$K_3^*(p)=\lim_{x\to a^+}\sqrt{2\pi(x-a)}\tau_{yz}^*(x,0,p)=-\frac{c_{44}c_{441}\sqrt{\pi a}}{c_{44}+c_{441}}\sum_{j=1}^{\infty}B_j(p). \quad (49)$$

4 Solution of the Impermeable Crack Problem

Consider now the impermeable crack problem. Performing Laplace transform to the boundary conditions (10a)–(10g) leads to

$$\tau_{yz}^*(x,-h,p)=\tau^*(x,p),\quad -\infty<x<\infty \quad (50a)$$

$$D_y^*(x,-h,p)=D^*(x,p),\quad -\infty<x<\infty \quad (50b)$$

$$\tau_{yz}^*(x,0,p)=\tau_{yz1}^*(x,0,p),\quad -\infty<x<\infty \quad (50c)$$

$$\tau_{yz}^*(x,0,p)=0,\quad |x|<a \quad (50d)$$

$$w^*(x,0,p)=w_1^*(x,0,p),\quad |x|>a \quad (50e)$$

$$D_y^*(x,0,p)=0,\quad |x|<a \quad (50f)$$

$$\phi^*(x,0,p)=0,\quad |x|>a \quad (50g)$$

From (50a)–(50c), it can be seen that

$$A_2(\xi,p)=A_1(\xi,p)\exp(2\gamma h)+\frac{\exp(\gamma h)}{\mu\gamma}\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right) \quad (51)$$

$$A_4(\xi,p)=A_3(\xi,p)\exp(2|\xi|h)-\frac{\exp(|\xi|h)\overline{D}}{\kappa_{11}|\xi|} \quad (52)$$

$$A_5(\xi,p)=\frac{\mu\gamma}{c_{441}\gamma_1}[1-\exp(2\gamma h)]A_1(\xi,p)+\frac{e_{15}|\xi|}{c_{441}\gamma_1}[1-\exp(2|\xi|h)]A_3(\xi,p)-\frac{\exp(\gamma h)}{c_{441}\gamma_1}\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right)+\frac{e_{15}\exp(|\xi|h)\overline{D}}{\kappa_{11}c_{441}\gamma_1} \quad (53)$$

Let us again introduce the following dislocation function:

$$\varphi_1(x,p)=\frac{\partial}{\partial x}[w_1^*(x,0,p)-w^*(x,0,p)] \quad (54)$$

and the following definition:

$$\varphi_2(x,p)=-\frac{\partial\phi^*(x,0,p)}{\partial x}. \quad (55)$$

According to Eqs. (50e) and (50g), those two functions must satisfy

$$\int_{-a}^{a}\varphi_1(\alpha,p)d\alpha=0 \quad (56)$$

$$\int_{-a}^{a}\varphi_2(\alpha,p)d\alpha=0. \quad (57)$$

By using Eqs. (54), (55), (12)–(15), and the definition of Fourier transform, we obtain

$$A_5(\xi,p)-A_1(\xi,p)-A_2(\xi,p)=-\frac{1}{i\xi}\int_{-a}^{a}\varphi_1(\alpha,p)\exp(i\xi\alpha)d\alpha \quad (58)$$

$$\frac{e_{15}}{\kappa_{11}}[A_1(\xi,p)+A_2(\xi,p)]+A_3(\xi,p)+A_4(\xi,p)=\frac{1}{i\xi}\int_{-a}^{a}\varphi_2(\alpha,p)\exp(i\xi\alpha)d\alpha. \quad (59)$$

Substitution of (51)–(53) into (58) and (59) yields

$$\begin{aligned}A_1(\xi,p)=&-\frac{c_{441}\gamma_1[1+\exp(2|\xi|h)]}{i\xi F_1(\xi,p)}\int_{-a}^{a}\varphi_1(\alpha,p)\exp(i\xi\alpha)d\alpha\\&-\frac{e_{15}|\xi|[1-\exp(2|\xi|h)]}{i\xi F_1(\xi,p)}\\&\times\int_{-a}^{a}\varphi_2(\alpha,p)\exp(i\xi\alpha)d\alpha+\left\{\left(1+\frac{c_{441}\gamma_1}{\mu\gamma}\right)\right.\\&\times[1+\exp(2|\xi|h)]\\&\left.+\frac{e_{15}^2|\xi|[1-\exp(2|\xi|h)]}{\kappa_{11}\mu\gamma}\right\}\frac{\exp(\gamma h)}{F_1(\xi,p)}\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right)\\&-\frac{2e_{15}\exp(|\xi|h)\overline{D}}{\kappa_{11}F_1(\xi,p)}\end{aligned} \quad (60)$$

$$\begin{aligned}A_3(\xi,p)=&\frac{c_{441}e_{15}\gamma_1[1+\exp(2\gamma h)]}{i\kappa_{11}\xi F_1(\xi,p)}\int_{-a}^{a}\varphi_1(\alpha,p)\exp(i\xi\alpha)d\alpha\\&+\frac{\mu\gamma[1-\exp(2\gamma h)]-c_{441}\gamma_1[1+\exp(2\gamma h)]}{i\xi F_1(\xi,p)}\\&\times\int_{-a}^{a}\varphi_2(\alpha,p)\exp(i\xi\alpha)d\alpha\\&-\frac{2e_{15}\exp(\gamma h)}{\kappa_{11}F_1(\xi,p)}\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right)+\left\{1\right.\\&\left.+\frac{2e_{15}^2|\xi|[1+\exp(2\gamma h)]}{\kappa_{11}F_1(\xi,p)}\right\}\frac{\exp(|\xi|h)\overline{D}}{\kappa_{11}|\xi|[1+\exp(2|\xi|h)]}.\end{aligned} \quad (61)$$

From (50d) and (50f), we have

$$\int_{-\infty}^{\infty}\left\{\gamma[\exp(2\gamma h)-1]A_1(\xi,p)+\frac{\exp(\gamma h)}{\mu}\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right)\right\}\times\exp(-i\xi x)d\xi=0,\ |x|<a \quad (62)$$

$$\int_{-\infty}^{\infty}\left\{|\xi|[\exp(2|\xi|h)-1]A_3(\xi,p)-\frac{\exp(|\xi|h)\overline{D}}{\kappa_{11}}\right\}\times\exp(-i\xi x)d\xi=0,\ |x|<a. \quad (63)$$

By substituting Eqs. (60), (61) into (62) and (63), we obtain

$$\frac{1}{i}\int_{-\infty}^{\infty} a_{11}(\xi,p)\exp(-i\xi x)\int_{-a}^{a}\varphi_1(\alpha,p)\exp(i\xi\alpha)d\alpha d\xi + \frac{1}{i}\int_{-\infty}^{\infty} a_{12}(\xi,p)\exp(-i\xi x)\int_{-a}^{a}\varphi_2(\alpha,p)\exp(i\xi\alpha)d\alpha d\xi = \tau_2(x,p), \quad |x|<a \tag{64}$$

$$\frac{1}{i}\int_{-\infty}^{\infty} a_{21}(\xi,p)\exp(-i\xi x)\int_{-a}^{a}\varphi_1(\alpha,p)\exp(i\xi\alpha)d\alpha d\xi + \frac{1}{i}\int_{-\infty}^{\infty} a_{22}(\xi,p)\exp(-i\xi x)\int_{-a}^{a}\varphi_2(\alpha,p)\exp(i\xi\alpha)d\alpha d\xi = \tau_3(x,p), \quad |x|<a \tag{65}$$

where

$$a_{11}(\xi,p)=\frac{c_{441}\gamma\gamma_1[1+\exp(2|\xi|h)][\exp(2\gamma h)-1]}{\xi F_1(\xi,p)} \tag{66}$$

$$a_{12}(\xi,p)=\frac{e_{15}|\xi|\gamma[1-\exp(2|\xi|h)][\exp(2\gamma h)-1]}{\xi F_1(\xi,p)} \tag{67}$$

$$a_{21}(\xi,p)=\frac{e_{15}c_{441}|\xi|\gamma_1[\exp(2|\xi|h)-1][\exp(2\gamma h)+1]}{\kappa_{11}\xi F_1(\xi,p)} \tag{68}$$

$$a_{22}(\xi,p)=\frac{|\xi|[\exp(2|\xi|h)-1]\{\mu\gamma[1-\exp(2\gamma h)]-c_{441}\gamma_1[1+\exp(2\gamma h)]\}}{\xi F_1(\xi,p)} \tag{69}$$

$$\tau_2(x,p)=-\int_{-\infty}^{\infty}\frac{2c_{441}\gamma_1[1+\exp(2|\xi|h)]+2e_{15}^2/\kappa_{11}|\xi|[1-\exp(2|\xi|h)]}{\mu F_1(\xi,p)}\exp(\gamma h-i\xi x)\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right)d\xi - \int_{-\infty}^{\infty}\frac{2e_{15}\gamma[\exp(2\gamma h)-1]\exp(|\xi|h-i\xi x)\overline{D}}{\kappa_{11}F_1(\xi,p)}d\xi \tag{70}$$

$$\tau_3(x,p)=\int_{-\infty}^{\infty}\frac{2e_{15}|\xi|[\exp(2|\xi|h)-1]}{\kappa_{11}F_1(\xi,p)}\exp(\gamma h-i\xi x)\left(\overline{\tau}+\frac{e_{15}}{\kappa_{11}}\overline{D}\right)d\xi + \int_{-\infty}^{\infty}\frac{2\mu\gamma[1-\exp(2\gamma h)]-2c_{441}\gamma_1[1+\exp(2\gamma h)]}{\kappa_{11}F_1(\xi,p)}\exp(|\xi|h-i\xi x)\overline{D}d\xi. \tag{71}$$

One can see that the right-hand side of (70) and (71) depends on both $\overline{\tau}$ and $\overline{D}$. This again indicates that the assumption that the loads acting on the crack faces are independent of the applied electric fields is inappropriate. However, by the use of the initial-value theorem as applied to Laplace transform, we can deduce that in the static case the electric field does have influence on the crack-tip fields when the impermeable condition is considered.

Performing appropriate asymptotic analysis leads to

$$\lim_{|\xi|\to\infty} a_{11}(\xi,p)=-\frac{c_{441}}{c_{44}+c_{441}}\operatorname{sign}(\xi) \tag{72}$$

$$\lim_{|\xi|\to\infty} a_{12}(\xi,p)=\frac{e_{15}}{c_{44}+c_{441}}\operatorname{sign}(\xi) \tag{73}$$

$$\lim_{|\xi|\to\infty} a_{21}(\xi,p)=-\frac{e_{15}c_{441}}{\kappa_{11}(c_{44}+c_{441})}\operatorname{sign}(\xi) \tag{74}$$

$$\lim_{|\xi|\to\infty} a_{22}(\xi,p)=\left[1+\frac{e_{15}^2}{\kappa_{11}(c_{44}+c_{441})}\right]\operatorname{sign}(\xi). \tag{75}$$

In a similar fashion to Section 3 and from Eqs. (64), (65), and (72)–(75), we have the following singular integral equations:

$$-\frac{c_{441}}{c_{44}+c_{441}}\int_{-a}^{a}\frac{\varphi_1(\alpha,p)}{\alpha-x}d\alpha+\frac{e_{15}}{c_{44}+c_{441}}\int_{-a}^{a}\frac{\varphi_2(\alpha,p)}{\alpha-x}d\alpha + \int_{-a}^{a}k_{11}(\alpha,x,p)\varphi_1(\alpha,p)d\alpha+\int_{-a}^{a}k_{12}(\alpha,x,p)\varphi_2(\alpha,p)d\alpha = \frac{1}{2}\tau_2(\xi,p), \quad |x|<a \tag{76}$$

$$-\frac{e_{15}c_{441}}{\kappa_{11}(c_{44}+c_{441})}\int_{-a}^{a}\frac{\varphi_1(\alpha,p)}{\alpha-x}d\alpha+\left[1+\frac{e_{15}^2}{\kappa_{11}(c_{44}+c_{441})}\right]\times\int_{-a}^{a}\frac{\varphi_2(\alpha,p)}{\alpha-x}d\alpha+\int_{-a}^{a}k_{21}(\alpha,x,p)\varphi_1(\alpha,p)d\alpha + \int_{-a}^{a}k_{22}(\alpha,x,p)\varphi_2(\alpha,p)d\alpha=\frac{1}{2}\tau_3(\xi,p), \quad |x|<a \tag{77}$$

where

$$k_{11}(\alpha,x,p)=\int_0^{\infty}\left[a_{11}(\xi,p)+\frac{c_{441}}{c_{44}+c_{441}}\right]\sin[\xi(\alpha-x)]d\xi \tag{78}$$

$$k_{12}(\alpha,x,p)=\int_0^{\infty}\left[a_{12}(\xi,p)-\frac{e_{15}}{c_{44}+c_{441}}\right]\sin[\xi(\alpha-x)]d\xi \tag{79}$$

$$k_{21}(\alpha,x,p)=\int_0^\infty\left[a_{21}(\xi,p)+\frac{e_{15}c_{441}}{\kappa_{11}(c_{44}+c_{441})}\right]\sin[\xi(\alpha-x)]d\xi \tag{80}$$

$$k_{22}(\alpha,x,p)=\int_0^\infty\left\{a_{22}(\xi,p)-\left[1+\frac{e_{15}^2}{\kappa_{11}(c_{44}+c_{441})}\right]\right\}\times\sin[\xi(\alpha-x)]d\xi. \tag{81}$$

The functions $\varphi_1(\alpha,p)$ and $\varphi_2(\alpha,p)$ are defined in terms of the Chebyshev polynomials:

$$\varphi_1(\alpha,p)=\sum_{j=0}^{\infty}\frac{B_j(p)}{\sqrt{1-\alpha^2/a^2}}T_j(\alpha/a),$$

$$\varphi_2(\alpha,p)=\sum_{j=0}^{\infty}\frac{E_j(p)}{\sqrt{1-\alpha^2/a^2}}T_j(\alpha/a). \tag{82}$$

From (56) and (57), it follows that $B_0(p)=E_0(p)=0$. By truncating the series to a reasonable number of terms and by using a simple collocation technique, we can determine the remaining unknowns using the following algebraic equations:

$$\sum_{j=1}^{N}\left[\frac{\sin\left(\frac{mj\pi}{N+1}\right)}{\sin\left(\frac{m\pi}{N+1}\right)}-L_{11j}(x_m,p)\right]B_j(p)-\sum_{j=1}^{N}\left[\frac{e_{15}\sin\left(\frac{mj\pi}{N+1}\right)}{c_{441}\sin\left(\frac{m\pi}{N+1}\right)}+L_{12j}(x_m,p)\right]E_j(p)=-\frac{c_{44}+c_{441}}{2\pi c_{441}}\tau_2(x_m,p) \tag{83}$$

$$-\sum_{j=1}^{N}\left[\frac{\sin\left(\frac{mj\pi}{N+1}\right)}{\sin\left(\frac{m\pi}{N+1}\right)}-L_{21j}(x_m,p)\right]B_j(p)+\sum_{j=1}^{N}\left\{\frac{[\kappa_{11}(c_{44}+c_{441})+e_{15}^2]\sin\left(\frac{mj\pi}{N+1}\right)}{e_{15}c_{441}\sin\left(\frac{m\pi}{N+1}\right)}+L_{22j}(x_m,p)\right\}E_j(p)=\frac{\kappa_{11}(c_{44}+c_{441})}{2\pi e_{15}c_{441}}\tau_3(x_m,p) \tag{84}$$

$$m=1,2,\ldots,N$$

where

$$L_{11j}(x_m,p)=\frac{c_{44}+c_{441}}{c_{441}}\int_{-a}^{a}\frac{1}{\pi\sqrt{1-\alpha^2/a^2}}k_{11}(\alpha,x_m,p)T_j(\alpha/a)d\alpha \tag{85}$$

$$L_{12j}(x_m,p)=\frac{c_{44}+c_{441}}{c_{441}}\int_{-a}^{a}\frac{1}{\pi\sqrt{1-\alpha^2/a^2}}k_{12}(\alpha,x_m,p)T_j(\alpha/a)d\alpha \tag{86}$$

$$L_{21j}(x_m,p)=\frac{\kappa_{11}(c_{44}+c_{441})}{e_{15}c_{441}}\int_{-a}^{a}\frac{1}{\pi\sqrt{1-\alpha^2/a^2}}k_{21}(\alpha,x_m,p)T_j(\alpha/a)d\alpha \tag{87}$$

$$L_{22j}(x_m,p)=\frac{\kappa_{11}(c_{44}+c_{441})}{e_{15}c_{441}}\int_{-a}^{a}\frac{1}{\pi\sqrt{1-\alpha^2/a^2}}k_{22}(\alpha,x_m,p)T_j(\alpha/a)d\alpha. \tag{88}$$

Based on the solutions of (83) and (84), the dynamic stress intensity factor and electric displacement intensity factor can be obtained, as follows:

$$K_3^*(p)=\lim_{x\to a^+}\sqrt{2\pi(x-a)}\,\tau_{yz}^*(x,0,p)=-\frac{\mu c_{441}\sqrt{\pi a}}{c_{44}+c_{441}}\sum_{j=1}^{\infty}B_j(p)+\frac{\mu e_{15}\sqrt{\pi a}}{c_{44}+c_{441}}\sum_{j=1}^{\infty}E_j(p)-\frac{e_{15}}{\kappa_{11}}K_D^*(p) \tag{89}$$

$$K_D^*(p)=\lim_{x\to a^+}\sqrt{2\pi(x-a)}D_y^*(x,0,p)=-\frac{e_{15}c_{441}\sqrt{\pi a}}{c_{44}+c_{441}}\sum_{j=1}^{\infty}B_j(p)+\sqrt{\pi a}\left(\kappa_{11}+\frac{e_{15}^2}{c_{44}+c_{441}}\right)\sum_{j=1}^{\infty}E_j(p). \tag{90}$$

5 Numerical Results and Discussions

Numerical calculations have been carried out to show the influence of the pertinent parameters. In the following calculations, the piezoelectric material is assumed to be the commercially available piezoceramic PZT-4, and the elastic material is assumed to be aluminum. The elastic, piezoelectric, and dielectric properties of the materials are as follows ([17]):

$$c_{44}=2.56*10^{10}\ \text{N/m}^2,\quad e_{15}=12.7\ \text{C/m}^2,$$
$$\kappa_{11}=64.6*10^{-10}\ \text{C/vm},\quad \rho=7500\ \text{kg/m}^3;$$
$$c_{441}=2.65*10^{10}\ \text{N/m}^2,\quad \rho_1=2706\ \text{kg/m}^3.$$

For the sake of simplicity, the electromechanical loads are assumed to act uniformly over the range $-b\leqslant x\leqslant b$ and in the form of a Heaviside step function; namely, $\tau(x,t)=\tau_0H(t)$ and $D(x,t)=D_0H(t)$, where τ_0 and D_0 are constants. In this case, we have $\overline{\tau}=2\tau_0\sin(b\xi)/(p\xi)$, $D=2D_0\sin(b\xi)/(p\xi)$.

To check the convergence of the expansions in (43) and (82), a number of runs with varying number of terms were used. We found that good convergence (2 percent difference between two successive runs) can be reached when the number exceeds 15 terms. In all our calculations, we used 20 terms.

Both the dynamic intensity factor and the electric displacement intensity factor in the physical plane are obtained by the numerical inversion of Laplace transform. Numerous numerical techniques have been developed for the inversion. Naraynan and Beskos [18] made a comparison study of those techniques and found that the Durbin's method ([19]) can give reliable results even for complicated functions. In the present analysis, we tried two techniques: the Durbin's method and the method developed by Miller and Guy [20]. Due to its simplicity, Miller and Guy's method has been used in most of the existing studies concerning the transient response of cracked piezoelectric materials (see, e.g., Chen and Yu [11],

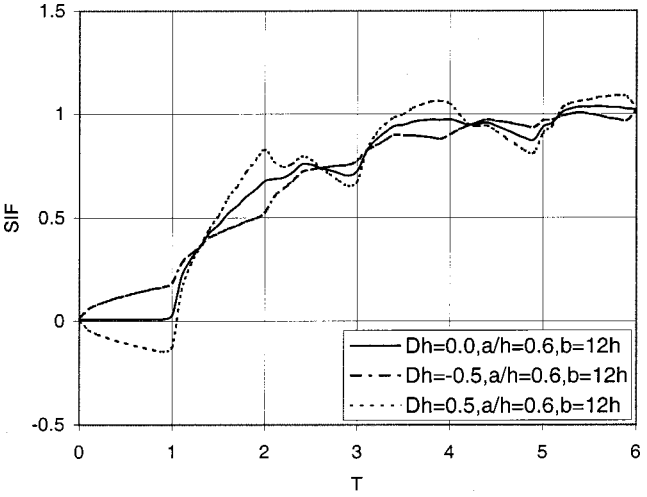


Fig. 2 Normalized SIF versus normalized time for various electromechanical loads and the permeable boundary condition

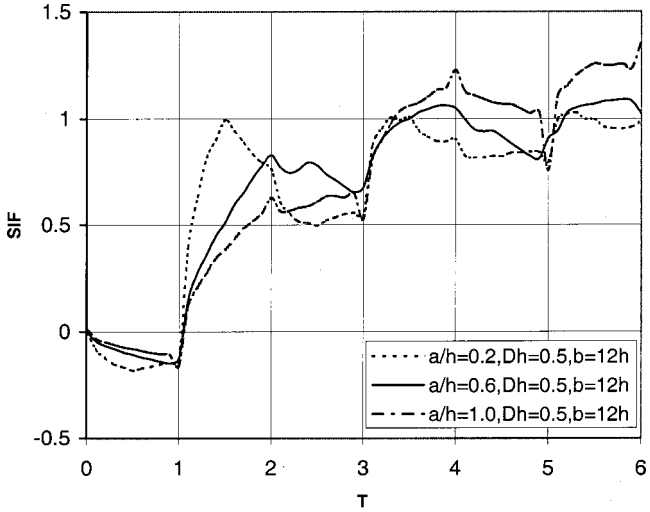


Fig. 3 Normalized SIF versus normalized time for various *a*/*h* and the permeable boundary condition

Chen and Karihaloo [12], Chen and Meguid [13], Wang and Yu [14], Shin et al. [15]). However, we found that this method gives diverging and inaccurate results. This is because the method uses very little information of the transformed domain, as can also be seen from the work of Naraynan and Beskos [18]. Therefore, only the results of the Durbin's method are presented here (Figs. 2–11). In these figures, normalized parameters are used with $\mathrm{SIF}=K_3(t)/(\tau_0\sqrt{\pi a})$, $\mathrm{EDIF}=K_D(t)/(D_0\sqrt{\pi a})$, $T=c_2t/h$ and $Dh=e_{15}D_0/(\kappa_{11}\tau_0)$.

Figures 2–4 are concerned with the results of the permeable crack problem. Specifically, Fig. 2 shows the influence of the applied electromechanical loads on the dynamic stress intensity factor. It is seen that if $T<1$, the incident stress wave does not arrive and the medium is completely at rest if no electric field is applied ($Dh=0$). This is reflected by the fact that the SIF is zero in that period of time. In the presence of the electric field, the SIF increases or decreases gradually with increasing time, depending on the direction of the electric field. The positive electric field ($Dh=0.5$) induces negative SIF, and the negative electric field ($Dh=-0.5$) causes positive SIF. When $T=1$, the incident stress wave arrives. Thereafter, the SIF rises rapidly with increasing time, and reaches a peak, then decreases until the arrival of the first reflecting wave from the free surface ($T=3$). It is shown that this process is intensified by the presence of the positive electric field, and alleviated by the presence of the negative electric field. The same phenomenon can also be observed, if we consider the first reflecting wave ($3\leqslant T<5$) and the second reflecting wave ($T\geqslant 5$). In the calculation of this figure, we also checked the accuracy of the Durbin's method. From the physical viewpoint, it is expected the SIF to be zero before $T=1$, since there is no electric field applied. The calculation shows that the error is less than 1 percent. The physical model indicates that the times at which the first and the second reflecting waves arrive should be $T=3$ and $T=5$, respectively. However, our calculations show that they are $T=2.9$ and $T=4.9$. This discrepancy is due to the approximation used in the model.

Figure 3 displays the variation of the SIF with various a/h at $Dh=0.5$. Due to the arrival of the incident stress wave, the decrease of a/h results in an increase in the SIF. However, when the reflecting waves arrive at $T=3$ and $T=5$, the decrease of a/h results in a decrease in the SIF. The effects of the loading range on the dynamic stress intensity factor are shown in Fig. 4.

Figures 5–8 are concerned with the results of the impermeable crack problem. Generally, similar observations can be deduced from Figs. 5–7. However, the effect of the electric field is more pronounced, as depicted in Fig. 5. It is seen that before the arrival of the incident stress wave, only the electric field is present. At $T=0$, the SIF has a jump. Then, a small overshoot takes place. From Figs. 6 and 7, one can deduce the influence of a/h and the loading range on this process.

The variation of the electric displacement intensity factor with various a/h at $Dh=0.5$ is depicted in Fig. 8. From this figure we can observe the dynamic overshoot phenomenon. Moreover, the phenomenon is intensified with the increase of a/h. This is quite different from the earlier results obtained in Chen and Yu [11], Chen and Karihaloo [12], Chen and Meguid [13], Wang and Yu [14], and Wang et al. [16], where the electric displacement intensity factor is in the form of a Heaviside step function.

Figures 9–11 compare the solutions of the dynamic stress intensity factor for the permeable and impermeable conditions. When $a/h=0.2$, the difference between the two solutions is negligible. However, when the ratio of a/h becomes large, the difference is appreciable. This indicates that for smaller cracks, the permeable and impermeable conditions provide comparable results for the local stress fields. For larger cracks, the two boundary conditions lead to different solutions for those local stress fields.

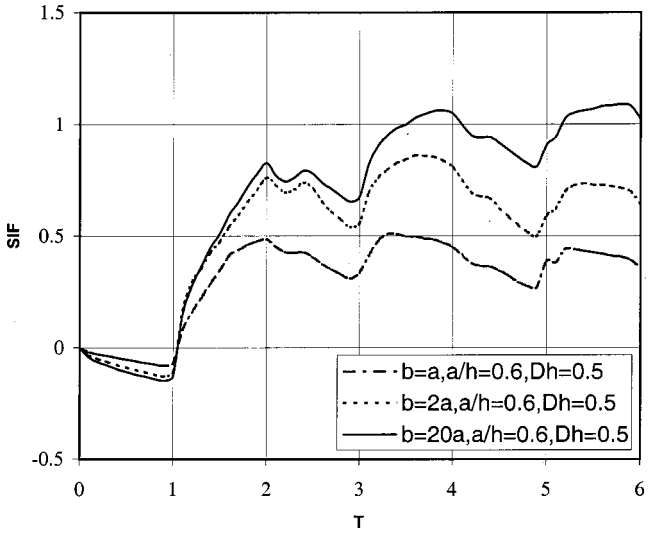


Fig. 4 Effects of loading range on the normalized SIF for the permeable boundary condition

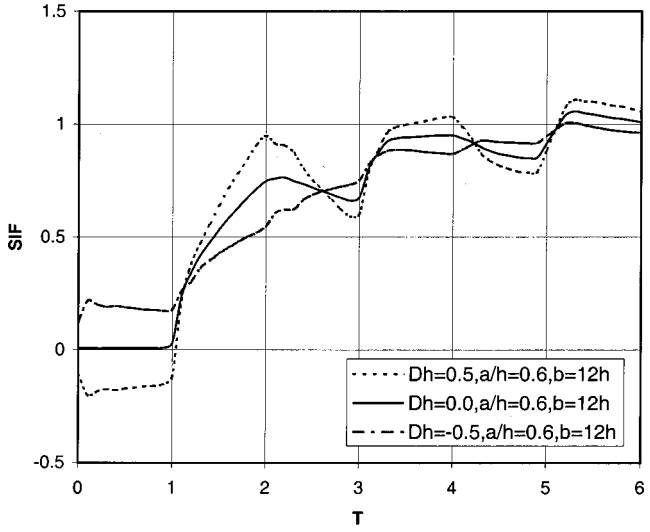


Fig. 5 Normalized SIF versus normalized time for various electromechanical loads and the impermeable boundary condition

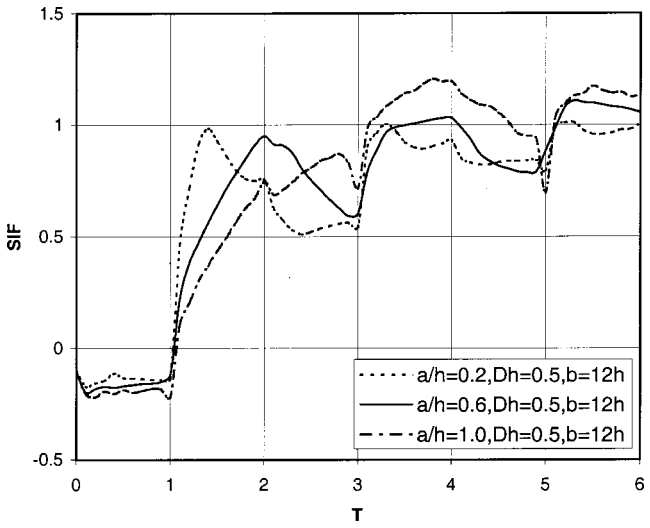


Fig. 6 Normalized SIF versus normalized time for various *a*/*h* and the impermeable boundary condition

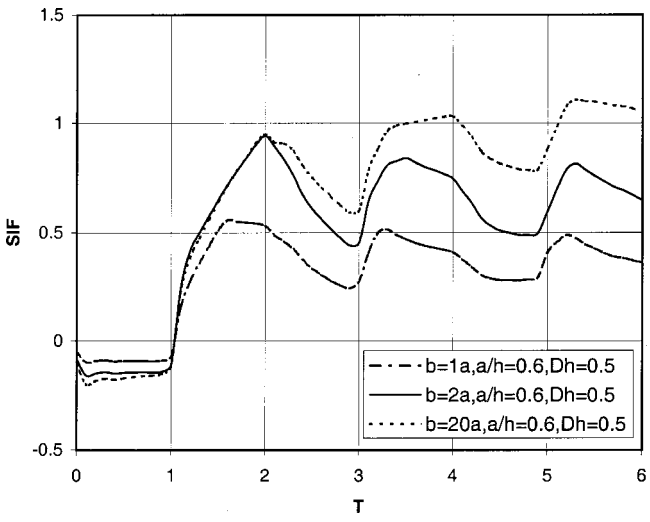


Fig. 7 Effect of loading range on the normalized SIF for the impermeable boundary condition

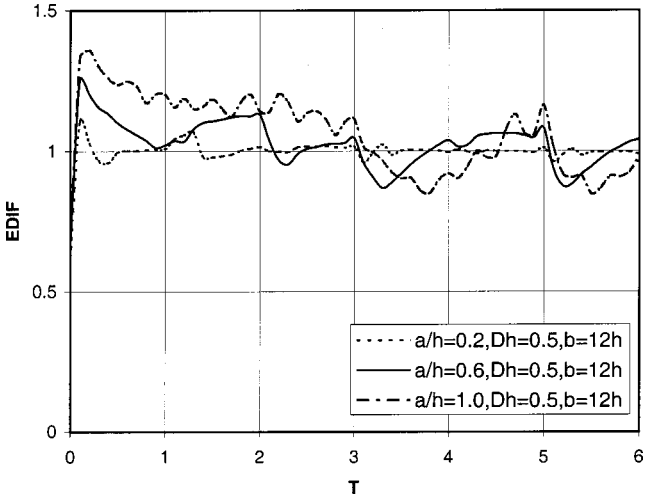


Fig. 8 Normalized EDIF versus normalized time for various *a*/*h* and the impermeable boundary condition

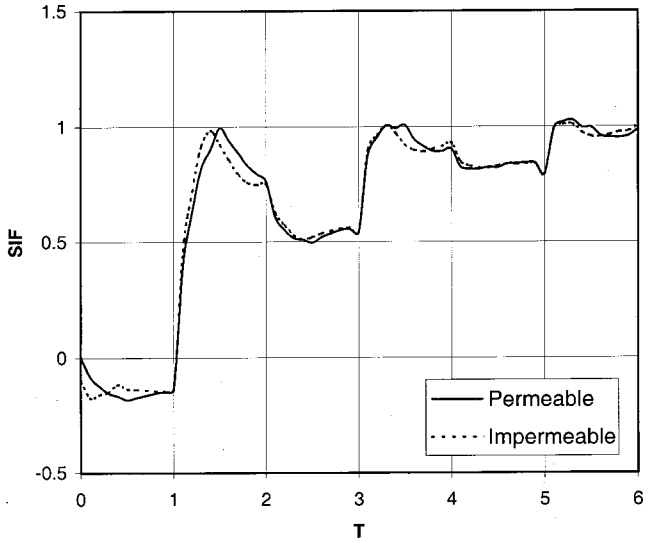


Fig. 9 Comparison between solutions of the permeable and impermeable boundary conditions for *a*/*h*=0.2, *Dh*=0.5, and *b*=12 h

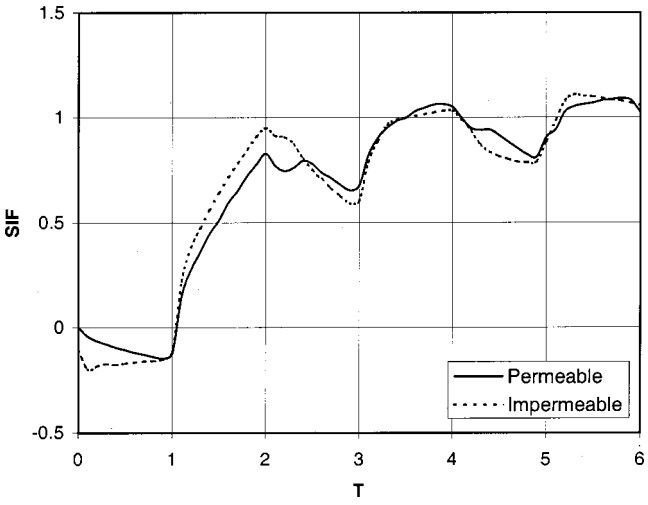


Fig. 10 Comparison between solutions of the permeable and impermeable boundary conditions for *a*/*h*=0.6, *Dh*=0.5, and *b*=12 h

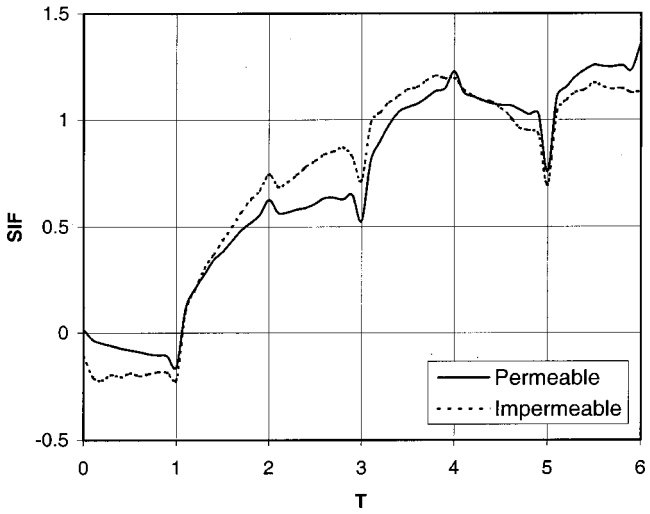


Fig. 11 Comparison between solutions of the permeable and impermeable boundary conditions for $a/h=1.0$, $Dh=0.5$, and $b=12\,h$

6 Conclusions

The dynamic problem of bonded piezoelectric and elastic half-space with an interface crack subjected to transient electromechanical load is investigated. The analysis is based on the use of integral transform techniques and integral equation methods. Numerical calculations are carried out to study the effect of the applied electric field, electric boundary conditions along the crack faces and a free surface on the resulting dynamic stress intensity factor and the electric displacement intensity factor. The study reveals that

1 The presence of the electric field results in an increase or a decrease of the dynamic stress intensity factor, depending on the direction of the electric field and the stage of responses. This indicates that both a positive electric field and a negative electric field can retard or promote the propagation of a crack. It is worth noting that an electric field alone can induce the stress intensity factor in cases involving transient loading. This result is different from the corresponding static problem. In the static case, the electric field has no influence on the crack tip fields when the permeable boundary condition is considered.

2 For both the permeable and impermeable boundary conditions, the interaction between a crack and the reflecting waves from a free surface is prevalent, which is manifested by the increase in the stress intensity factor due to the arrival of the reflecting waves.

3 An overshoot phenomenon exists for the electric displacement intensity factor when the impermeable boundary condition is considered. Moreover, the phenomenon is intensified with the increase in the crack length.

4 For a crack of small length (for the present case, $a/h<0.2$), both the permeable boundary condition and the impermeable boundary condition give comparable results for the local stress fields. When the crack length is relatively large, the difference between the two cases becomes evident.

References

[1] Shindo, Y., and Ozawa, E., 1990, ''Dynamic Analysis of a Cracked Piezoelectric Material,'' *Mechanical Modeling of New Electromagnetic Materials*, R. K. T. Hsieh, ed., Elsevier, Amsterdam, pp. 297–304.

[2] Norris, A. N., 1994, ''Dynamic Green's Functions in Anisotropic Piezoelectric, Thermoelastic and Poroelastic Solids,'' Proc. R. Soc. London, Ser. A, **447**, pp. 175–186.

[3] Khutoryyansky, N. M., and Sosa, H., 1995, ''Dynamic Representation Formulas and Fundamental Solutions for Piezoelectricity,'' Int. J. Solids Struct., **32**, pp. 3307–3325.

[4] Shindo, Y., Katsura, H., and Yan, W., 1996, ''Dynamic Stress Intensity Factor of a Cracked Dielectric Medium in a Uniform Electric Fields,'' Acta Mech., **117**, pp. 1–10.

[5] Narita, F., and Shindo, Y., 1998, ''Dynamic Anti-Plane Shear of a Cracked Piezoelectric Ceramic,'' Theor. Appl. Fract. Mech., **29**, pp. 169–180.

[6] Narita, F., and Shindo, Y., 1998, ''Scattering of Love Waves by a Surface-Breaking Crack in Piezoelectric Layered Media,'' JSME Int. J., Ser. A, **41**, pp. 40–48.

[7] Meguid, S. A., and Wang, X. D., 1998, ''Dynamic Anti-Plane Behavior of Interacting Cracks in a Piezoelectric Material,'' Int. J. Fract., **91**, pp. 391–403.

[8] Wang, X. D., 2001, ''On the Dynamic Behavior of Interacting Interfacial Cracks in Piezoelectric Media,'' Int. J. Solids Struct., **38**, pp. 815–831.

[9] Li, S., and Mataga, P. A., 1996, ''Dynamic Crack Propagation in Piezoelectric Materials. Part 1: Electrode Solution,'' J. Mech. Phys. Solids, **44**, pp. 1799–1830.

[10] Li, S., and Mataga, P. A., 1996, ''Dynamic Crack Propagation in Piezoelectric Materials. Part 2: Vacuum Solution,'' J. Mech. Phys. Solids, **44**, pp. 1831–1866.

[11] Chen, Z. T., and Yu, S. W., 1997, ''Crack Tip Field of Piezoelectric Materials Under Anti-Plane Impact,'' Chin. Sci. Bull., **42**, pp. 1613–1617.

[12] Chen, Z. T., and Karihaloo, B. L., 1999, ''Dynamic Response of a Cracked Piezoelectric Ceramic Under Arbitrary Electro-Mechanical Impact,'' Int. J. Solids Struct., **36**, pp. 5125–5133.

[13] Chen, Z. T., and Meguid, S. A., 2000, ''The Transient Response of a Piezoelectric Strip With a Vertical Crack Under Electro-Mechanical Impact Load,'' Int. J. Solids Struct., **37**, pp. 6051–6062.

[14] Wang, X., and Yu, S., 2000, ''Transient Response of a Crack in a Piezoelectric Strip Subjected to the Mechanical and Electrical Impacts,'' Int. J. Solids Struct., **37**, pp. 5795–5808.

[15] Shin, J. W., Kwon, S. M., and Lee, K. Y., 2001, ''An Eccentric Crack in a Piezoelectric Strip Under Anti-Plane Shear Impact Loading,'' Int. J. Solids Struct., **38**, pp. 1483–1494.

[16] Wang, B. L., Han, J. C., and Du, S. Y., 2000, ''Electroelastic Fracture Dynamics for Multi-layered Piezoelectric Materials Under Dynamic Anti-Plane Shearing,'' Int. J. Solids Struct., **37**, pp. 5219–5231.

[17] Narita, F., and Shindo, Y., 1999, ''Scattering of Antiplane Shear Waves by a Finite Crack in Piezoelectric Laminates,'' Acta Mech., **134**, pp. 27–43.

[18] Narayanan, G. V., and Beskos, D. E., 1982, ''Numerical Operational Methods for Time-Dependent Linear Problems,'' Int. J. Numer. Methods Eng., **18**, pp. 1829–1854.

[19] Durbin, F., 1974, ''Numerical Inversion of Laplace Transforms: An Efficient Improvement to Duber and Abate's Method,'' Comput. J., **17**, pp. 371–376.

[20] Miller, M. K., and Guy, W. T., 1966, ''Numerical Inversion of the Laplace Transform by Using Jocabi Polynomials,'' SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal., **3**, pp. 624–635.

Numerical Simulation of Pulsed Laser Bending

X. R. Zhang

G. Chen[1]

X. Xu[2]

e-mail: xxu@ecn.purdue.edu

School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907

The aim of this work is to develop an efficient method for computing pulsed laser bending. During pulsed laser bending, thousands of laser pulses are irradiated onto the target. Simulations of the thermomechanical effect and bending resulted from all the laser pulses would exceed the current computational capability. The method developed in this work requires only several laser pulses to be calculated. Therefore, the computation time is greatly reduced. Using the new method, it is also possible to increase the domain size of calculation and to choose dense meshes to obtain more accurate results. The new method is used to calculate pulsed laser bending of a thin stainless-steel plate. Results calculated for a domain with a reduced size are in good agreement with those obtained by computing all the laser pulses. In addition, experiments of pulsed laser bending are performed. It is found that experimental data and computational results are consistent.
[DOI: 10.1115/1.1459070]

1 Introduction

Laser bending or laser forming is a newly developed, flexible technique which modifies the curvature of sheet metal or hard material using energy of a laser. The schematic of a laser bending process is shown in Fig. 1. The target is irradiated by a focused laser beam passing across the target surface with a certain scanning speed. After laser heating, permanent bending is resulted, with the bending direction toward the laser beam (the positive *z*-direction shown in Fig. 1). Laser bending has been explained by the thermoelastoplastic theory ([1–4]). During the heating period, irradiation of the laser beam produces a sharp temperature gradient in the thickness direction, causing the upper layers of the heated material to expand more than the lower layers. This nonuniform thermal expansion causes the target to bend away from the laser beam. In the meantime, compressive stress and strain are produced by the bulk constraint of the surrounding materials. Because of the high temperature achieved, plastic deformations occur. During cooling, heat flows into the adjacent area and the stress changes from compressive to tensile due to thermal shrinkage. However, the compressive strain generated during heating is not completely cancelled. Therefore, the residual strain in the laser-irradiated area is compressive after the target cools, causing a permanent bending deformation toward the laser beam.

A large amount of experimental and numerical work has been conducted to study CW (continuous wave) laser bending of sheet metals ([5–10]). Applications of laser bending include forming complex shapes and straightening automobile body shells. Laser bending is also being used for high-precision curvature modification during hard disk manufacturing, in which low energy pulsed lasers are used ([4]). Chen et al. [11] studied bending by a line-shape pulsed laser beam using a two-dimensional finite element model. Since the laser beam intensity they used was uniform across the target surface (along the *y*-direction shown in Fig. 1), the effect of bending was calculated using a two-dimensional heat transfer model and a plane-strain model, and the calculation was greatly simplified. Relations between bending angles and pulsed laser parameters were determined by both computational and experimental methods.

Little work has been done on pulsed laser bending using a three-dimensional model. In a common pulsed laser bending operation such as the one used for curvature adjustment in hard disk manufacturing, thermal and thermomechanical phenomena involved are three-dimensional. Laser pulses with Gaussian intensity distributions and high repetition rates are irradiated along the scanning line, as shown in Fig. 2. The main difficulty for simulating pulsed laser bending is that thousands of laser pulses along the laser scanning direction need to be calculated. For example, at a scanning speed of 10 mm/s and a pulse repetition rate of 10 kHz, there will be a total of 2000 pulses irradiated on a 2-mm wide target. Also, the numbers of nodes and elements in a three-dimensional model are much more than that in a two-dimensional model. Direct simulations of any actual pulsed laser bending process are impractical in terms of both the computation time and the computer resource.

In this paper, an efficient calculation method is developed to simulate pulsed laser bending. Instead of calculating bending resulted from all the laser pulses, bending due to a fraction of the total laser pulses is computed. Then, the calculated strain distribution at a cross section perpendicular to the scanning direction is imposed onto the whole target as an initial condition to calculate bending. A computational algorithm is developed. The accuracy of this method is verified by both numerical calculations and experimental measurements.

2 Numerical Procedure

2.1 Calculation of Deformations From the Strain Field

In most pulsed laser bending processes, constant stress and strain fields along the laser scanning direction are obtained. Although a single laser pulse generates nonuniform stress and strain distributions, in practice, laser pulses with same pulse energy, separated by a very small distance compared with the laser beam radius are used. Thus, the laser-induced stress and strain vary little along the scanning direction. With this in mind, it is only necessary to calculate several laser pulses until the stress and strain fields in an *x*-*z* cross-sectional area are not changed by a new laser pulse. Then, the residual strain field in this cross section can be imposed onto the whole domain to calculate the deformation (bending). In other words, a strain field $\{\varepsilon_r\}$, which can be used to calculate displacements of the target after pulsed laser scanning, is generated by calculating only a fraction of the total pulses.

[1]Current address: CNH Global NV, Burr Ridge, IL.

[2]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, November 9, 2000; final revision, May 8, 2001. Associate Editor: K. T. Ramesh. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

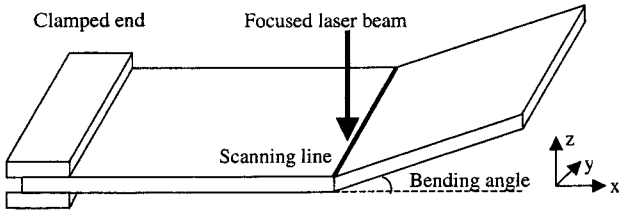


Fig. 1 Schematic of the laser bending process. The laser beam scans along a line in the *y*-direction, causing residual stress and strain in the laser irradiated area and permanent bending.

Before discussing the method of calculating displacements from a strain field, it is worth mentioning that the residual stress field couldn't be used to calculate displacements. The reason is that displacements are dependent not only on the stress but also on the load path when the plastic strain is involved. Different displacements will result from different load paths; even the residual stress fields are the same. On the other hand, there is a one-to-one correspondence between the strain and displacement fields. Therefore, the displacement field of the target can be completely determined by the strain field.

The finite element solver, ABAQUS (HKS, Inc., Pawtucket, RI) is used for the numerical calculation. In ABAQUS, only the stress field can be used as an initial condition for computation. Therefore, an initial stress field, which can produce the strain field equal to the laser produced strain field $\{\varepsilon_r\}$, needs to be obtained first. The method for calculating this stress field is described below.

Consider an undeformed domain without any external forces, but with an initial stress field $\{\sigma_i\}$. In order to satisfy force equilibrium, this initial stress should relax completely. For stress relaxation, the stress field in the domain can be written by

$$\{\sigma\}=\{\sigma_i\}+[E]\{\varepsilon\} \tag{1}$$

where $[E]$ is the matrix of elastic stiffness, $\{\varepsilon\}$ is the strain field due to stress relaxation, and $\{\sigma\}$ is the stress field. After stress relaxation, $\{\sigma\}\rightarrow\{0\}$. The strain field can be obtained by

$$\{\varepsilon\}=-\{\sigma_i\}/[E]. \tag{2}$$

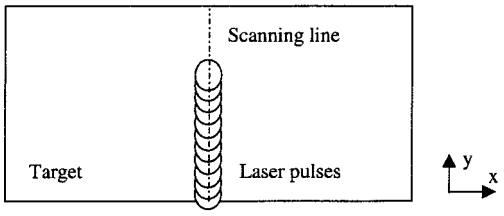


Fig. 2 Irradiation of laser pules on the target surface. The laser scans in the positive *y*-direction.

This equation determines the relationship between an initial stress field and the resulted strain field after stress relaxation. It can be seen that, if an initial stress field $\{\sigma_i\}=-[E]\{\varepsilon_r\}$ is used in the stress relaxation calculation, the resulted strain field will be identical to the strain field $\{\varepsilon_r\}$.

Therefore, in a brief summary, the computation starts with calculating a strain field $\{\varepsilon_r\}$ from several pulses and impose this strain field to the entire domain. Then a stress field $\{\sigma_i\}$ is obtained by computing $-[E]\{\varepsilon_r\}$. This stress field is applied to an undeformed domain followed by a stress relaxation calculation. This calculation yields both the strain $\{\varepsilon_r\}$ as well as the displacement (bending).

To verify this simulation method and use it to compute the pulsed laser bending process, a three-dimensional model is built and simulations of pulsed laser bending are conducted. In the first case, a full-hard 301 stainless steel sample that is 400 μm long, 120 μm wide, and 100 μm thick is irradiated by a pulsed laser. The scanning speed of the laser beam is set to be 195 mm/s, resulting in a total of fourteen pulses along the scanning line; and a 9 μm step size between two adjacent laser pulses. Although the domain size used here is smaller than many of those used in practice, the reduced domain size makes it possible to calculate the temperature, stress, and strain distributions produced by all the 14 laser pulses. On the other hand, to test the new calculation method, the strain distribution in the *x*-*z* cross section at $y=60$ μm after eight laser pulses is imposed onto the whole domain, and the procedures outlined above are used to compute the deformation caused by all the pulses. Results from the two approaches are then compared. In the second case, a full-hard 301 stainless steel sample that is 8 mm long, 1.2 mm wide, and 0.1 mm thick is irradiated by a pulsed laser. The laser scanning speed is also 195 mm/s, resulting in a total of 134 pulses. In this case, only the new method is used since it is impossible to complete the computation of all the 134 pulses within a reasonable amount of time. Experiments are conducted on samples with same dimensions and processing parameters, and the results of experiments and simulations are compared. The laser parameters used in the simulation and the experiment are summarized in Table 1.

The computational domain and mesh for the first case are shown in Fig. 3. Only half of the target is calculated because the central plane is approximated as a symmetry plane. A dense mesh is used around the laser path and then stretched away in length and thickness directions (*x* and *z*-directions). In the dense mesh region, eight elements are used in the *x*-direction, 33 elements in the *z*-direction, and 24 elements in the *y*-direction. A total of 9944 elements are used in the mesh. The same mesh is used for thermal analyses and stress-displacement calculations. The mesh tests are conducted by increasing the mesh density until the calculation result is independent of the mesh density.

Dissipation of energy by the plastic deformation is negligible compared with the high laser energy density during bending. Therefore, it is assumed that the thermal and mechanical problems are decoupled, so that the thermal analysis and the stress and strain calculation can be conducted separately.

Table 1 Pulsed laser parameters

Laser wavelength	1.064 μm
Laser pulse full width	120 ns
Laser pulse repetition	22 kHz
Laser pulse energy	4.4 – 6.4 μJ
Laser beam diameter	50 μm
Laser scanning speed	195 mm/s
Pulse step distance	9 μm

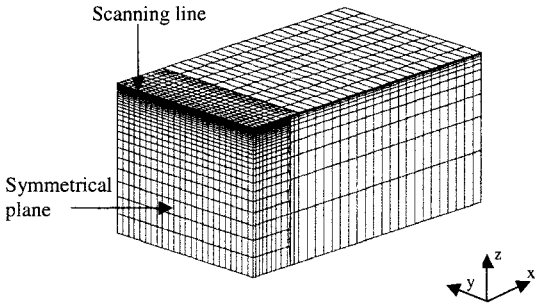


Fig. 3 Computational mesh (x:200 μm, y:120 μm, z:100 μm)

2.2 Thermal Analysis. The thermal analysis is based on solving the three-dimensional heat conduction equation. The initial condition is that the whole specimen is at the room temperature (300 K). Since the left and right boundaries as well as the bottom surface are far away from the laser beam, the boundary conditions at these boundaries are prescribed as the room temperatures. The laser flux is handled as a volumetric heat source absorbed by the target. The laser intensity at the target surface is considered as having a Gaussian distribution in both x and y-directions, which can be expressed as

$$I_s(x,y,t)=I_0(t)\ \exp\left(-2\frac{x^2+(y-y_0)^2}{w^2}\right) \tag{3}$$

where $I_0(t)$ is the time-dependent laser intensity at the center of the laser beam ($x=0;y=y_0$) and w is the beam radius. The temporal profile of the laser intensity is treated as increasing linearly from zero to the maximum at 60 ns, then decreasing to zero at the end of the pulse at 120 ns. The local radiation intensity $I(x,y,z,t)$ within the target is calculated considering exponential attenuation and surface reflection as

$$I(x,y,z,t)=(1-R_f)I_s(x,y,t)e^{-\alpha z} \tag{4}$$

where R_f is the optical reflectivity. α is the absorption coefficient given by

$$\alpha=\frac{4\pi\kappa}{\lambda}. \tag{5}$$

The imaginary part of the refractive index κ of stainless steel 301 at the laser wavelength 1.064 μm is unknown, and $\kappa=4.5$ of iron is applied. Properties used in the calculation are considered as temperature-dependent, and are shown in Fig. 4.

Sensitivity of calculated bending with respect to optical reflectivity has been studied ([11]). It was found that a 10% change of optical reflectivity value would cause a 23% difference in the bending angle. Therefore, the uncertainty in reflectivity does influence calculation results significantly. In this work, the reflectivity is measured to be 0.66, which has an uncertainty less than 5%.

The thermal analysis is carried out for laser pulse energy of 4.4 μJ, 5.4 μJ, and 6.4 μJ, respectively. The maximum temperatures obtained are all lower than the melting point of steel (1650 K).

2.3 Stress and Strain Calculation. For each laser pulse, the transient temperature field obtained from the thermal analysis is used as thermal loading, and residual stress and strain fields of the previous pulse are input as initial conditions to solve the quasi-static force equilibrium equations. The material is assumed to be linearly elastic-perfectly plastic. The Von Mises yield criterion is used to model the onset of plasticity. The boundary conditions are zero displacement in the x-direction and no rotations around y and z-axes in the symmetry plane, and all other surfaces are stress free. Details of the equations to be solved have been described elsewhere ([10]).

As shown in Fig. 4, material properties including density, yield stress, and Young's modulus are considered temperature-dependent. However, the strain rate enhancement effect is neglected because temperature-dependent data are unavailable. A constant value (0.3) of Poisson's ratio is used. Sensitivity of unknown material properties on the computational results has been studied ([11]). It was found that possible errors resulting from extrapolating material properties at high temperatures and using a constant Poisson's ratio were within a few percent.

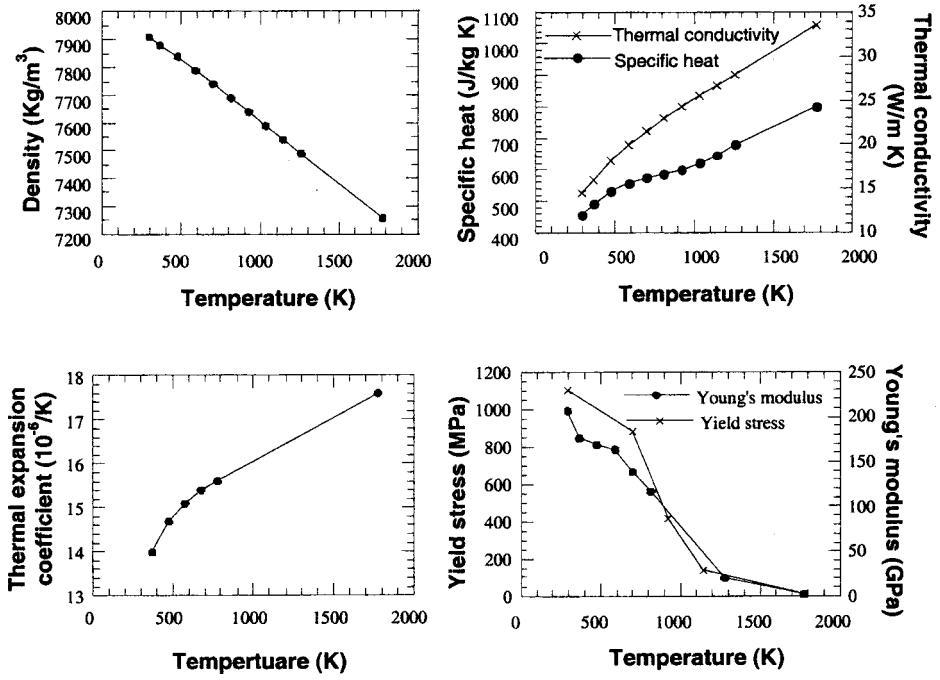


Fig. 4 Thermal and mechanical properties of full-hard 301 stainless steel

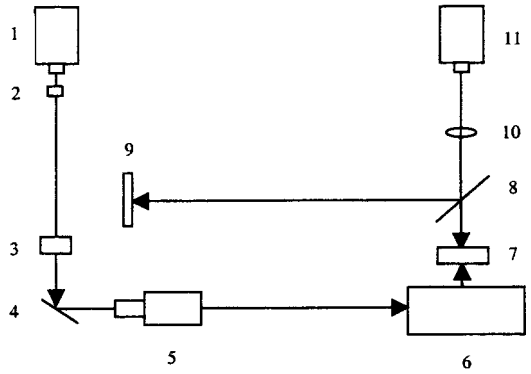


Fig. 5 Experimental setup for pulsed laser bending and for measuring the bending angle [1–ND:VA laser, 2–shutter, 3–polarizing beam splitter, 4–mirror, 5–beam expander, 6–X&Y scanner, 7–specimen, 8–beam splitter, 9–position-sensitive detector, 10–lens, 11–He-Ne laser]

3 Experimental Measurements

Experiments of bending of stainless steel are performed to verify the calculation results. The laser used in experiments is a pulsed Nd:VA laser with the same operation parameters shown in Table 1. Figure 5 illustrates the experimental setup for performing pulsed laser bending as well as for measuring the bending angle. The Nd:VA laser beam scans the specimen surface along the *y*-axis (Fig. 1) at a speed of 195 mm/s. The scanning speed is accurately controlled by a digital scanning system and the pulse step is 9 μm at this speed. An He-Ne laser beam is focused at the free end of the target to measure the bending angle in the *z*-direction. The reflected He-Ne laser beam is received by a position sensitive detector (PSD) with 1-μm sensitivity in position measurements. The accuracy of the bending angle measurement is about ± 1.5 μrad when the distance between the specimen and the PSD is set to 750 mm in the experiment. After laser scanning, the target bends toward the laser beam, causing the reflected He-Ne laser beam to move across the PSD. The position change of He-Ne laser beam can be converted to the bending angle of the specimen using geometrical calculations. The whole apparatus is set on a vibration-isolation table. Polished full hard 301 stainless steel sheets are used as targets.

4 Results and Discussion

Results calculated using a reduced domain size are presented to illustrate the temperature and residual strain and stress distributions induced by laser pulses. Bending deformations obtained by the new calculation method and by computing all laser pulses are then compared. Bending deformations resulted from different laser pulse energy are also presented. For the second case for which a larger sample is used, the calculated bending angles using the new method are compared with the experimental data.

4.1 Results Calculated Using a Reduced Domain Size Temperature distributions along *x* and *y*-directions and at different times are shown in Fig. 6. The laser pulse energy is 5.4 μJ and the pulse center is located at $y=54$ μm. Figure 6(*a*) shows the temperature distribution along the scanning line (the *y*-direction). It can be seen that the maximum temperature, $T_{\max}$, is reached at the pulse center. $T_{\max}$ increases once the laser pulse is irradiated on the surface and reaches its peak value 988.1 K at 87.7 ns, and then drops slowly to 365.5 K at 2.2 μs. It can be estimated that the laser-heated region is around 30 μm in radius. Figure 6(*b*) is the temperature distribution along the depth direction (the *z*-direction), beginning from the upper surface of the target. The maximum temperature is obtained at the upper surface and reaches 988.1 K at $t=87.7$ ns. The heat propagation depth is

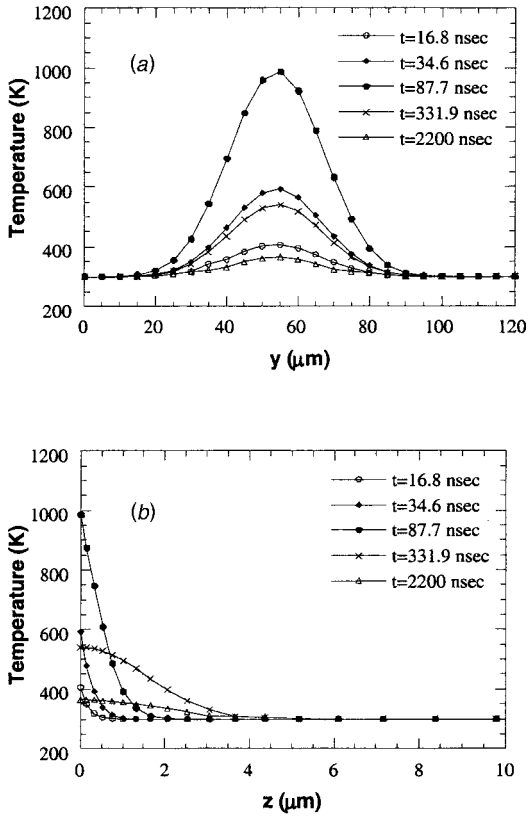
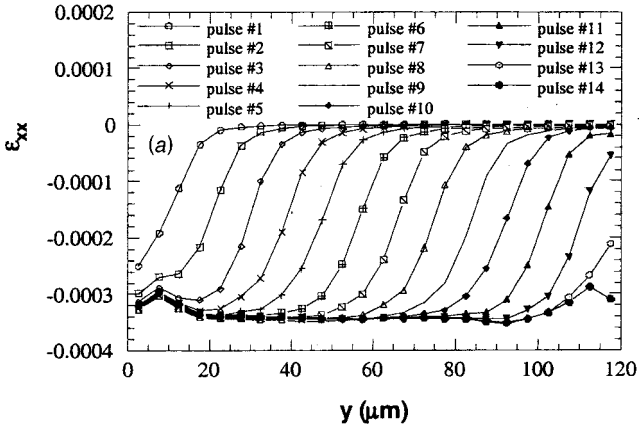


Fig. 6 Temperature distributions induced by the seventh pulse (pulse energy 5.4 μJ; pulse center at $y=54$ μm) (*a*) along the scanning line, (*b*) along the *z*-direction

around 4 μm at 2.2 μs and the temperature gradient during heating period is as high as 350 K/μm. This sharp temperature gradient causes nonuniform plastic strains in the target and the permanent bending deformation after laser heating.

Residual strain ε_{xx} and stress σ_{xx} distributions along the laser scanning path obtained from calculating all the fourteen pulses are plotted in Fig. 7(*a*) and Fig. 7(*b*), respectively. Only the components ε_{xx} and σ_{xx} are plotted since they are more important to the bending deformation than other components. It can be seen from Fig. 7(*a*) that after four pulses, the strain field in regions about 15 μm behind the new laser pulse is no longer changed. In other words, in the *y*-direction, each pulse only affects the stress and strain field within 15 μm from its center. It is also seen that after the laser pulses pass the whole target width, the residual stress and strain fields of the target are independent of the *y*-coordinate with the exception near the two edges, which is caused by the free stress boundary conditions. The uniform stress and strain along the *y*-direction are consistent with the assumption used in the calculation.

Residual strain ε_{xx} and stress σ_{xx} distributions along the *x*-direction at the upper surface are shown in Fig. 8(*a*) and Fig. 8(*b*), respectively. They are obtained after eight laser pulses in the cross section $y=60$ μm. It can be seen from Fig. 8(*a*) that the strain ε_{xx} is compressive within 15 μm from the center of the laser pulse. This agrees with the theoretical prediction that the compressive residual strain will be obtained near the center of laser-irradiated area where the temperature is the highest and the plastic deformation occurs ([4]). The residual strain ε_{xx} becomes positive (tensile strain) at locations more than 15 μm away from the center. The tensile strain in this region is due to the tensile force produced by thermal shrinkage during cooling. The total strained region is about 30 μm from the center of the laser beam and is slightly larger than the radius of the laser beam (25 μm). In



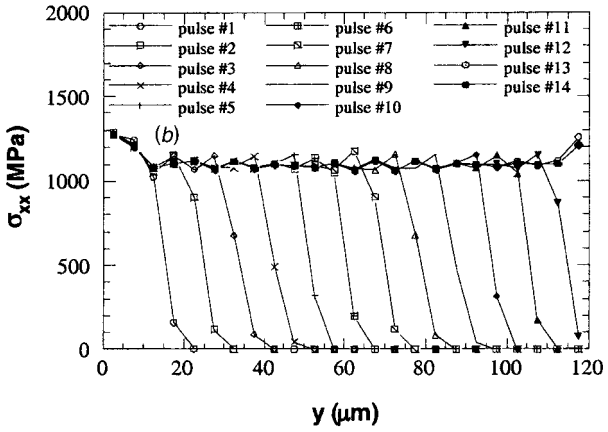
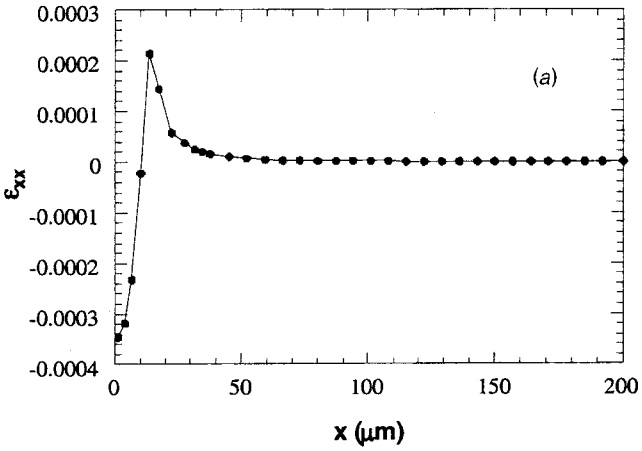


Fig. 7 (*a*) Residual strain (ε_{xx}), (*b*) residual stress (σ_{xx}) distributions along the scanning line induced by each laser pulse (pulse energy 5.4 μJ; scanning speed 195 mm/s)



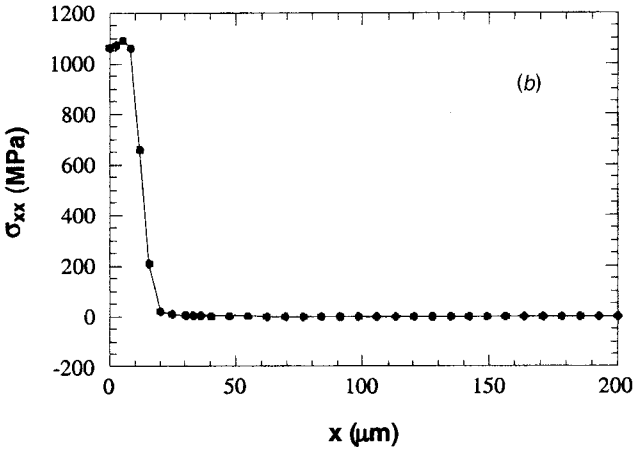


Fig. 8 (*a*) Residual strain (ε_{xx}), (*b*) residual stress (σ_{xx}) distributions along the *x*-direction ($y=60$ μm and $z=0$ μm) after eight pulses

Fig. 8(*b*), the stress σ_{xx} is tensile and its value is around 1.1 GPa in the region within 15 μm from the pulse center. This large tensile stress cancels more than 90% of the plastic strain produced during heating in this region. The tensile stress drops quickly to zero at about 25 μm from the center of the laser beam.

The strain distribution ε_{xx} calculated from the initial stress field $\{\sigma_i\}$ using the new simulation method is shown in Fig. 9. The average value of ε_{xx} obtained from the new method is -3.47×10^{-4}, comparing with the value of -3.42×10^{-4} calculated from all the 14 pulses. The two strain values are in very good agreement except at two edges. Again, the difference is caused by the free boundary conditions at the edges.

The off-plane displacement w is of prime interest since it reflects the amount of bending. The comparison between the deformation calculated from the initial stress $\{\sigma_i\}$ and that obtained by calculating all the pulses is shown in Fig. 10. Results at the cross section $y=60$ μm are plotted. It can be seen that displacements w of the two approaches are consistent and the bending angles are almost identical. The difference between the two curves is located around the transition mesh region. This is because that the element size and the shape in the transition region are not all the same, and errors are produced when the residual strain of one x-z cross section is imposed to the whole domain. It is seen from Fig. 10 that a "V" shape surface deformation is resulted after laser scanning, with the valley located at around 10 μm from the center of the scanning line. The positive off-plane displacement near the center of the scanning line is produced by thermal expansion along the positive z-direction because of the free-surface boundary condition.

Figure 11 shows the off-plane displacement w of the central point on the free edge of the surface ($x=200$ μm,$y=60$ μm,$z=100$ μm) produced after each laser pulse with pulse energy of 4.4 μJ, 5.4 μJ, and 6.4 μJ, respectively. As expected, laser pulses with high energy produce more bending. It is also seen that w increases almost linearly with the number of pulses for all the three cases.

4.2 Comparison Between Experimental and Numerical Results. Bending angles obtained experimentally are compared with calculated values as shown in Fig. 12. Laser energy of 4.4 μJ, 5.4 μJ, and 6.4 μJ is used in the experiment. On the other hand, calculations are carried out using the new method, in which the strain distribution obtained after eight laser pulses is imposed onto the entire computation domain. The size of the computation domain is 0.2 mm$\times$1.2 mm$\times$0.1 mm, which is identical to the sample size used in the experiment in the y and z-directions. Using a smaller size in the x-direction does not affect the computation results, since regions at x greater than 0.2 mm undergo a rigid rotation only. From the figure, it is seen that the experimental results agree with the calculated values within the experimental uncertainty. Both the experiment and simulation show the bending angle increases almost linearly with the pulse energy.

The agreements between the results of two numerical methods, and between the experimental and numerical results show that the

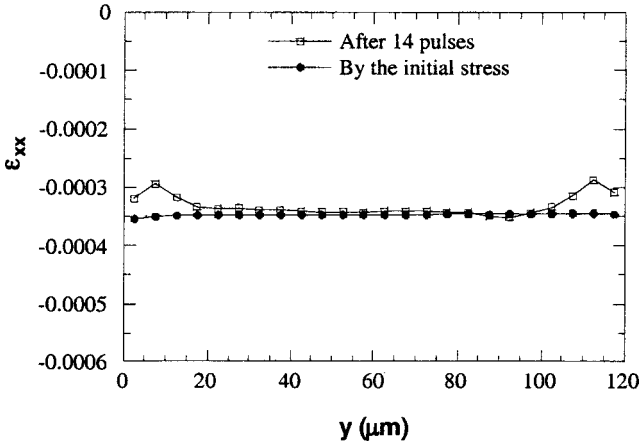


Fig. 9 Residual strain ε_{xx} along the scanning line on the top surface obtained by calculating all the 14 pulses and by using the new method (pulse energy 5.4 μJ; scanning speed 195 mm/s)

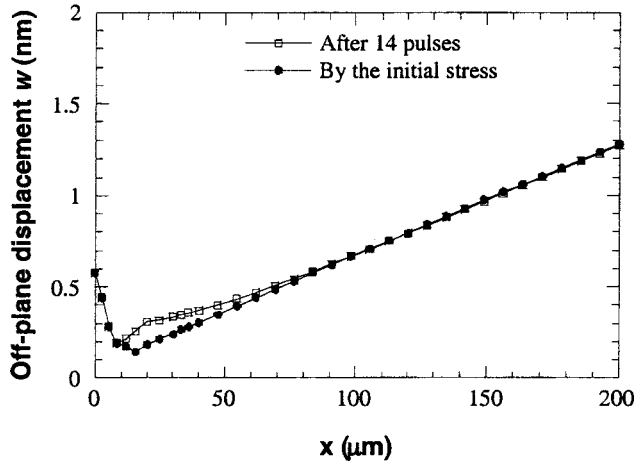


Fig. 10 Displacement w along the x-direction ($y=60$ μm and $z=0$ μm) obtained by calculating all the laser pulses and by using the new method (pulse energy 5.4 μJ; scanning speed 195 mm/s)

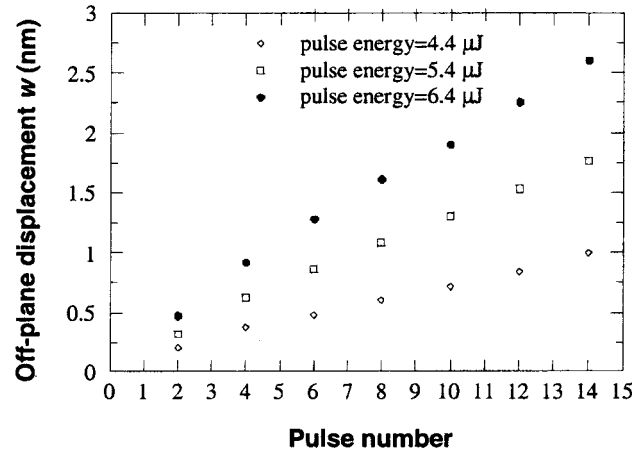


Fig. 11 Calculated displacement w at the free edge after each laser pulse as a function of laser energy

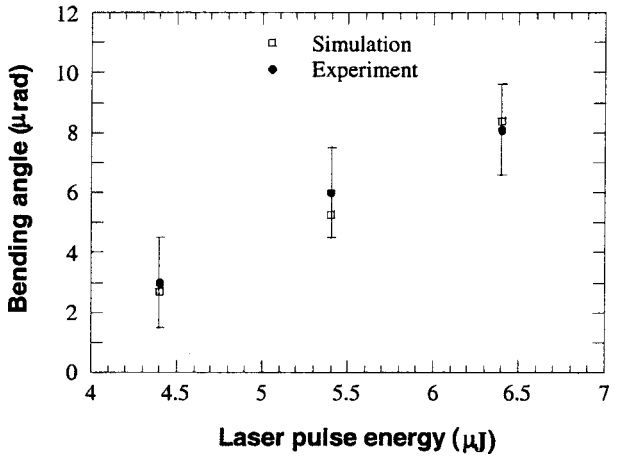


Fig. 12 Comparison between the measured bending angles and the simulation results obtained using the new method

newly developed method is indeed capable of computing pulsed laser bending. As indicated previously, the advantage of the new method is that the computation time is greatly reduced. For each laser pulse in the first case, about two hours are needed for the temperature calculation and four hours for the stress calculation using an 800 MHz Dell PC Workstation. It takes about 84 hours to obtain the bending deformation resulted from all the 14 pulses, and 50 hours when the new method is used. On the other hand, for the second case, it would have taken more than 10,000 hours to obtain the bending deformation if all the pulses were to be calculated. Using the new method, it only takes about 100 hours to complete the calculation. Thus, even for a sample as small as a few mm in size, bending can only be calculated with the use of the new method.

One concern of using the new method for calculating pulsed laser bending is when the laser beam scans the surface at a very high speed, thus the pulse step-size becomes large enough to cause nonuniform stress and strain along the scanning line. However, if the laser-induced stress and strain distribution is periodic, i.e., produced by high-speed scanning of the laser beam with constant energy per pulse, this method still works. The strain distribution within a period along the y-direction can be imposed to the whole domain, and the remaining steps follow those described previously in Section 2.1.

5 Conclusion

A new efficient method for computing pulsed laser bending is developed. The total computation time is greatly reduced and results are found to agree with those obtained using a conventional computation method. Experimental studies are also carried out to verify the simulation results. It is found that the calculated results agree with the experimental values. For most pulsed laser bending processes, the newly developed method is the only possible way to compute bending within a reasonable amount of time.

Acknowledgment

Support to this work by the National Science Foundation (DMI-9908176) is gratefully acknowledged. The authors also thank Dr. Andrew C. Tam of IBM Almaden Research Center for collaborations on this work.

References

[1] Namba, Y., 1986, "Laser Forming in Space," *International Conference on Lasers*, C. P. Wang et al., eds., STS Press, Las Vegas, NV, pp. 403–407.
[2] Geiger, M., and Vollertsen, F., 1993, "The Mechanisms of Laser Forming," Ann. CIRP, **42**, pp. 301–304.
[3] Vollertsen, F., 1994, "Mechanisms and Models for Laser Forming," *Laser Assisted Net Shape Engineering, Proc. of the LANE*, Vol. 1, M. Geiger et al., eds., Meisenbach, Bamberg, Germany, pp. 345–360.

[4] Chen, G., Xu, X., Poon, C. C., and Tam, A. C., 1998, ''Laser-Assisted Microscale Deformation of Stainless Steels and Ceramics,'' Opt. Eng., **37**, pp. 2837–2842.
[5] Scully, K., 1987, ''Laser Line Heating,'' J. Ship Res., **3**, pp. 237–246.
[6] Alberti, N., Fratini, L., and Micari, F., 1994, ''Numerical Simulation of the Laser Bending Processing by a Coupled Thermal Mechanical Analysis,'' *Laser Assisted Net Shape Engineering, Proc. of the LANE*, Vol. 1, M. Geiger et al., eds., Meisenbach, Bamberg, Germany, pp. 327–336.
[7] Pridham, M., and Thomson, G., 1995, ''An Investigation of Laser Forming Using Empirical Methods and Finite Element Analysis,'' J. Des. Manufact., **5**, pp. 203–211.
[8] Magee, J., Watkins, K. G., and Steen, W. M., 1997, ''Laser Forming of Aerospace Alloys,'' *Proc. of Laser Institute of America, ICALEO*, **83**, Laser Institute of America, Orlando, FL, pp. 156–165.
[9] Hsiao, Y. C., Shimizu, H., Firth, L., Maher, W., and Masubuchi, K., 1997, ''Finite Element Modeling of Laser Forming,'' *Proc. of Laser Institute of America, ICALEO*, **83**, Laser Institute of America, Orlando, FL, pp. 31–40.
[10] Chen, G., and Xu, X., 2000, ''3D CW Laser Forming of Thin Stainless Steel Sheets,'' ASME J. Manuf. Sci. Eng., in press.
[11] Chen, G., Xu, X., Poon, C. C., and Tam, A. C., 1999, ''Experimental and Numerical Studies on Microscale Bending of Stainless Steel With Pulsed Laser,'' ASME J. Appl. Mech., **66**, pp. 772–779.

Coupled Torsion-Lateral Stability of a Shaft-Disk System Driven Through a Universal Joint

H. A. DeSmidt
Graduate Research Assistant

K. W. Wang
Willian E. Diefenderfer Chaired Professor in Mechanical Engineering, Fellow ASME

E. C. Smith
Associate Professor of Aerospace Engineering

The Pennsylvania State University, 157 Hammond Building, University Park, PA 16802

Understanding the instability phenomena of rotor-shaft and driveline systems incorporating universal joints is becoming increasingly important because of the trend towards light-weight, high-speed supercritical designs. In this paper, a nondimensional, periodic, linear time-varying model with torsional and lateral degrees-of-freedom is developed for a rotor shaft-disk assembly supported on a flexible bearing and driven through a U-joint. The stability of this system is investigated utilizing Floquet theory. It is shown that the interaction between torsional and lateral dynamics results in new regions of parametric instability that have not been addressed in previous investigations. The presence of load inertia and misalignment causes dynamic coupling of the torsion and lateral modes, which can result in torsion-lateral instability for shaft speeds near the sum-type combinations of the torsion and lateral natural frequencies. The effect of angular misalignment, static load-torque, load-inertia, lateral frequency split, and auxiliary damping on the stability of the system is studied over a range of shaft operating speeds. Other than avoiding the unstable operating frequencies, the effectiveness of using auxiliary lateral viscous damping as a means of stabilizing the system is investigated. Finally, a closed-form technique based on perturbation expansions is derived to determine the auxiliary damping necessary to stabilize the system for the least stable case (worst case).
[DOI: 10.1115/1.1460907]

1 Introduction

Universal joints, commonly known as U-joints, are used in many power transmission applications when power must be transmitted across noncollinear shafts. U-Joints are widely used because they are relatively inexpensive and easy to maintain compared with many other types of shaft couplings. Furthermore, their design allows them to accommodate relatively large angular misalignments. Additionally, U-joints have high torque capability and can withstand relatively large axial loads.

Despite the advantages of the U-Joint, their nonconstant velocity nature can lead to undesirable vibration and even instability under certain conditions. Since the joint rotations are about two successive axes, there is a nonconstant velocity relationship between the driving and the driven end of the U-joint. This difference between the driving and driven shaft speed is commonly known as the Cardan Error. When the driving and driven shafts are not collinear, the driven shaft speed fluctuates at a frequency twice the driving shaft speed, even if the driving shaft speed is constant. This fluctuation can excite torsional vibration of the system and can lead to instability. Furthermore, torque transmitted through a U-joint in the presence of angular misalignment generates lateral moments that fluctuate at twice the driving shaft speed. These lateral moments can excite lateral motion of the shaft and can also destabilize the system.

Several researchers have investigated the dynamics and stability of rotating shafts driven through U-joints. Iwatsubo and Saigo [1] have studied the effect of load-torque on the transverse vibration of a nominally aligned rigid rotor disk driven through a U-joint. They derived the expressions for the parametric and self-exciting transverse moments created by torque transmitted through the U-joint. It was determined that static load-torque can induce parametric instabilities at the sum-type combinations of the transverse natural frequencies. Additionally, it was shown that load-torque also induces flutter instability ([1]).

Similar to [1], Mazzei et al. [2] considered a flexible shaft U-joint system without the rotor disk. The effect of static load-torque on the stability of a misaligned, slender, flexible shaft supported by compliant bearings was investigated. Torsion dynamics were neglected since the shaft was slender with no rotor disk. Similar to results presented in [1], Mazzei et al. [2] found that the static load-torque produced parametric instability near the sum-type combinations of the lateral bending natural frequencies of the shaft as well as flutter instability. Also, [2] compared the stability of a flexible shaft U-joint system to a rigid shaft U-joint system. They showed that the instability zones of the flexible shaft model associated with modes involving mainly rigid motion were very similar to the instability zones predicted by the rigid shaft model. Thus, the presence of shaft flexibility only adds additional instability zones associated with the flexible modes, but does not fundamentally alter the instability zones predicted by the rigid model.

Rosenberg [3] considered the effect of static misalignment on the stability of a flexible shaft driven by a U-joint. Here the misalignment angle is modeled as a static nominal angle plus a dynamic misalignment due to shaft vibration. Here again, torsional dynamics are neglected, and only lateral shaft flexibility is considered. Compared to the effect of static load-torque in [1,2] the results from [3] show that angular misalignment only weakly affects the stability near the sum-type lateral combination frequencies of the shaft.

Xu and Marangoni [4,5] examined how static angular misalignment between two shafts affects the lateral moments created by a U-joint. In this analysis, they considered only the static portion of the angular misalignment. Thus the effect of the misalignment was modeled by time periodic transverse moment forcing terms which contain integer multiple harmonics of twice the driving shaft speed. However, since dynamic misalignment due to shaft vibration was neglected, the potentially destabilizing parametric terms were not included in the analysis.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Feb. 15, 2001; final revision, Oct. 3, 2001. Editor: N. C. Perkins. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

Kato and Ota, [6], studied U-joint frictional effects for a statically misaligned shaft. Here, they derived expressions for lateral moments generated by viscous and coulomb friction between the yokes and cross piece of the U-joint. They concluded that internal friction in the U-joint generates harmonic lateral moments that occur at even multiples of the shaft operating speed, i.e., 2Ω, 4Ω, ..., etc. Additionally, they demonstrated that the viscous friction-induced lateral moments are suppressed if the friction coefficients at the driven and driving yokes are equal.

Asokanthan and Hwang [7] and Asokanthan and Wang [8] both studied the stability of two torsionally flexible, misaligned shafts coupled by a U-joint. In their analyses, the shafts were driving an inertia load and the orientations were fixed, thus lateral motion was neglected and only torsional dynamics were considered. Asokanthan and Hwang [7] concluded that shaft speed variation kinematics due to static angular misalignment caused fundamental and sum type parametric instabilities. They also showed that the width of the parametric instability zones increased with increasing misalignment angle.

2 Problem Statement and Research Objective

For a shaft/U-joint system carrying an inertia load, it has been shown that angular misalignment causes periodic speed variation of the driven shaft and induces torsional dynamics. Furthermore, other studies have shown that load-torque transmitted across a U-joint generates lateral moments that excite lateral dynamics. However, since these two phenomena have been investigated separately, the important interaction between the torsion and lateral dynamics has not been addressed in previous investigations. Because the effective misalignment is the sum of some nominal static misalignment plus the dynamic slope of the shaft at the U-joint due to lateral vibration, the misalignment-induced speed variation and resulting torsional dynamics are a function of the lateral dynamics. Thus, lateral shaft dynamics induce torsional dynamics. On the other hand, torsional dynamics generate a dynamic load-torque proportional to the inertia load. This torsion dynamics-induced dynamic load-torque is transmitted through the U-joint and generates lateral moments, which means torsional dynamics excite lateral dynamics. Therefore, the torsion and lateral modes of vibration are dynamically coupled, and due to the nature of the U-joint, this dynamic coupling is periodic.

As will be seen in the next few sections, this torsion-lateral coupling phenomena is important for supercritical rotor-shaft applications since it creates shaft speed regions of parametric instability not previously identified in the literature. The objective of this research is to address this critical but unexamined issue, that is, to understand the interaction between lateral and torsional dynamics and explore the bounds of the torsion-lateral instability regions. Specifically, the dynamic stability of a torsionally flexible, misaligned, U-joint driven shaft-disk assembly mounted on a compliant bearing/damper is examined.

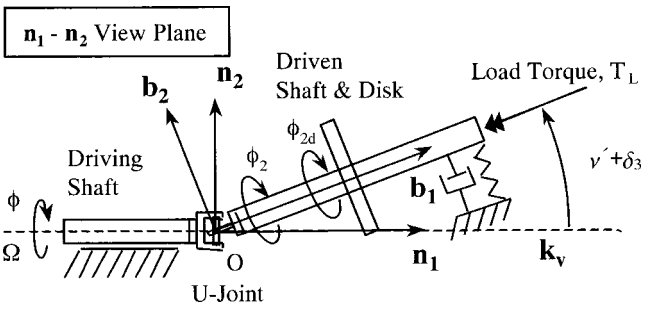


Fig. 1 Misaligned shaft and disk driven through a U-joint

3 System Description and Model

Equations of motion are derived for the system illustrated in Fig. 1. The torsionally flexible shaft, carrying a rotor disk, is subjected to a follower torque load and mounted on a spring bearing/damper while being driven through a U-joint coupling. The driving shaft speed, Ω, and the magnitude of the follower load-torque, T_L, are assumed constant. Since the driven shaft is considered to be rigid in bending and pinned at the U-joint, the orientation can be completely described in the fixed frame, $\{\mathbf{n}\}$, by two orthogonal projected slopes in the $\mathbf{n}_1$-$\mathbf{n}_2$ and $\mathbf{n}_1$-$\mathbf{n}_3$ planes. The statically misaligned equilibrium operating condition of the driven shaft is defined by static misalignment angles δ_2 and δ_3 in the $\mathbf{n}_1$-$\mathbf{n}_3$ and $\mathbf{n}_1$-$\mathbf{n}_2$ planes, respectively. The dynamic portion of the misalignment is then measured from the static misaligned operating condition by projected slopes, v' and ω', which correspond to the $\mathbf{n}_1$-$\mathbf{n}_3$ and $\mathbf{n}_1$-$\mathbf{n}_2$ planes, respectively. Figure 2 shows the rotation sequence from the fixed frame $\{\mathbf{n}\}$, to the body fixed frame of the driven shaft, $\{\mathbf{b}\}$.

From the fixed $\{\mathbf{n}\}$ frame, the intermediate frame, $\{\mathbf{n}'\}$ which follows the driven shaft, is defined by the combined static and dynamic projected slopes $v'+\delta_3$ and $\omega'+\delta_2$. Next, the body fixed frame, $\{\mathbf{b}\}$, is obtained by the driven shaft spin angle, ϕ_2, about the intermediate $\mathbf{n}_1'$-axis. The coordinate transformation matrix from the $\{\mathbf{n}\}$ frame to the $\{\mathbf{b}\}$ frame is shown in Eq. (1).

$$\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{bmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\phi_2 & \sin\phi_2\\ 0 & -\sin\phi_2 & \cos\phi_2\end{pmatrix}\begin{pmatrix}1-\dfrac{(\delta_3+v')^2}{2}-\dfrac{(\delta_2+w')^2}{2} & \delta_3+v' & \delta_2+w'\\ -(\delta_3+v') & 1-\dfrac{(\delta_3+v')^2}{2} & \dfrac{(\delta_3+v')(\delta_2+w')}{2}\\ -(\delta_2+w') & -\dfrac{(\delta_3+v')(\delta_2+w')}{2} & 1-\dfrac{(\delta_2+w')^2}{2}\end{pmatrix}\begin{bmatrix}\mathbf{n}_1\\ \mathbf{n}_2\\ \mathbf{n}_3\end{bmatrix} \quad (1)$$

Since the shaft is torsionally flexible, the total spin angle of the disk, ϕ_{2d}, is the sum of the driven yoke spin, ϕ_2, plus the elastic twist angle, $\hat{\phi}_{2d}$. Due to the kinematics of the U-joint, the spin angle of the driven yoke, ϕ_2, is a function of the driving yoke spin angle ϕ, and the misalignments, $v'+\delta_3$ and $\omega'+\delta_2$, of the driven shaft relative to the driving shaft. Since the driving shaft speed, Ω, is a given constant, the spin angle $\phi=\Omega t$, is also known. Thus ϕ_2 can be determined as a function of ϕ, v', ω', δ_2, and δ_3.

In order to determine the above relationship for ϕ_2, the coordinate transformation from $\{\mathbf{n}\}$ to $\{\mathbf{b}\}$ is represented in terms of the driving yoke spin angle, ϕ, about the $\mathbf{n}_1$-axis, followed by two

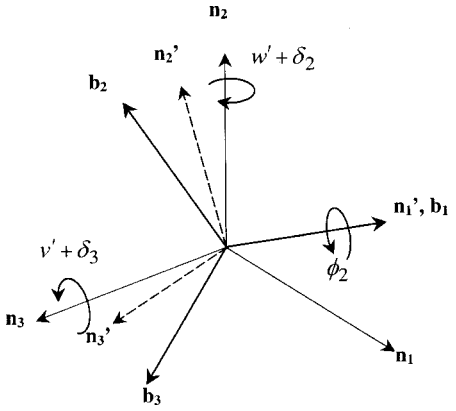


Fig. 2 Projected slopes $v'+\delta_3$ and $\omega'+\delta_2$ and driven shaft spin angle, ϕ_2 from $\{n\}$ to $\{b\}$

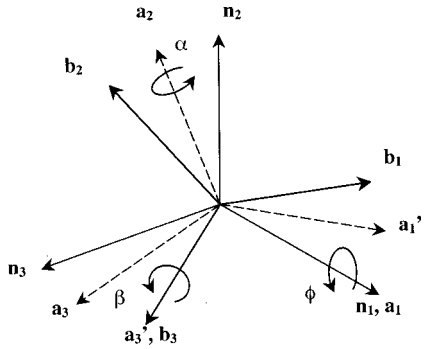


Fig. 3 Driving shaft spin angle, ϕ, and Euler angles α and β from $\{n\}$ to $\{b\}$

Euler angles, α and β, about successive axes, $\mathbf{a}_2$ and $\mathbf{a}_3'$, defined by the U-joint center cross piece. See Fig. 3.

The resulting coordinate transformation matrix from $\{\mathbf{n}\}$ to $\{\mathbf{b}\}$ is shown in Eq. (2).

$$\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{bmatrix}=\begin{pmatrix}\cos\alpha\cos\beta & \sin\beta & -\sin\alpha\cos\beta\\ -\sin\beta\cos\alpha & \cos\beta & \sin\beta\sin\alpha\\ \sin\alpha & 0 & \cos\alpha\end{pmatrix}\times\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\phi & \sin\phi\\ 0 & -\sin\phi & \cos\phi\end{pmatrix}\begin{bmatrix}\mathbf{n}_1\\ \mathbf{n}_2\\ \mathbf{n}_3\end{bmatrix} \tag{2}$$

By equating the two coordinate transformation expressions given in Eqs. (1) and (2) and dropping higher order terms, the two Euler angles α and β and the driven yoke spin angle, ϕ_2, are determined as functions of ϕ, v', ω', δ_2, and δ_3 for small misalignments. See Eqs. (3) and (4).

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=\begin{pmatrix}\sin\phi & -\cos\phi\\ \cos\phi & \sin\phi\end{pmatrix}\begin{bmatrix}v'+\delta_3\\ w'+\delta_2\end{bmatrix}+O(v'^2\delta) \tag{3}$$

$$\phi_2=\phi+\frac{\sin 2\phi}{4}[(\delta_3+v')^2-(\delta_2+w')^2]-\frac{\cos 2\phi}{2}(\delta_3+v')\times(\delta_2+w')+\frac{1}{2}[\delta_2 v'-\delta_3 w']+O(v'^3\delta) \tag{4}$$

Differentiating Eq. (4) with respect to time yields the expression for the driven yoke speed, Ω_2 in Eq. (5).

$$\Omega_2=\Omega+\frac{\Omega}{2}\cos 2\phi[(v'+\delta_3)^2-(w'+\delta_2)^2]+\Omega\sin 2\phi(v'+\delta_3)(w'+\delta_2)+\frac{1}{2}[\delta_2\dot{v}'-\delta_3\dot{w}']+\frac{\sin 2\phi}{2}[(v'+\delta_3)\dot{v}'-(w'+\delta_2)\dot{w}']-\frac{\cos 2\phi}{2}[(w'+\delta_2)\dot{v}'+(v'+\delta_3)\dot{w}'] \tag{5}$$

Note that the above expression for Ω_2 is more comprehensive than expressions used by other researchers since terms involving time derivatives of v' and ω' are included in the expression. Since the elastic twist of the shaft at the U-joint end is zero, and since the polar inertia of the disk is much greater than the shaft polar inertia, it is assumed that all of the torsion dynamics are lumped at the disk. Thus, only one elastic twist degree-of-freedom is needed to describe the torsion dynamics. The functional relationships of the spin, ϕ_{2d}, and spin rate, Ω_{2d}, of the disk about the $\mathbf{b}_1$-axis due to the driven yoke spin, ϕ_2, and spin rate, Ω_2, and elastic twist $\hat{\phi}_{2d}$ are shown in Eq. (6).

$$\begin{aligned}\phi_{2d}&=\phi_2(\phi,v',w',\delta_2,\delta_3)+\hat{\phi}_{2d}(t)\\ \Omega_{2d}&=\Omega_2(\Omega,\dot{v}',\dot{w}',v',w',\delta_2,\delta_3,t)+\dot{\hat{\phi}}_{2d}(t)\end{aligned} \tag{6}$$

The body fixed angular velocity of the disk can be expressed as

$$\mathbf{N}_\omega\mathbf{B}=\left[\Omega_{2d}+\frac{1}{2}(\dot{v}'w'-\dot{w}'v')\right]\mathbf{b}_1+[\sin(\phi_{2_d})\dot{v}'-\cos(\phi_{2_d})\dot{w}']\mathbf{b}_2+[\sin(\phi_{2_d})\dot{w}'+\cos(\phi_{2_d})\dot{v}']\mathbf{b}_3. \tag{7}$$

The total kinetic energy of the system is

$$T=\frac{I_m}{2}(\dot{v}'^2+\dot{w}'^2)+\frac{J_m}{2}(\Omega_{2d}+\Omega_{2d}[w'\dot{v}'-\dot{w}'v']) \tag{8}$$

where the total moments of inertia about point O are

$$I_m\equiv I_{ms}+I_d+m_dL_d^2 \quad \text{and} \quad J_m\equiv J_{ms}+J_d. \tag{9}$$

Here, I_{ms} and J_{ms} are the transverse and polar moments of inertia of the shaft about point O. I_d and J_d are the transverse and polar moments of inertia of the disk. m_d is the mass of the disk, and L_d is the axial location of the disk measured from point O. The total potential energy of the system is

$$V=\frac{k_\phi}{2}\hat{\phi}_{2d}^2+\frac{k_vL^2}{2}v'^2+\frac{k_wL^2}{2}w'^2 \tag{10}$$

With the lumped torsion spring stiffness k_ϕ is defined as

$$k_\phi=\frac{GJ_{cs}}{L_d}. \tag{11}$$

Also, k_v and k_w are the transverse bearing spring stiffness values corresponding to the v' and ω'-directions, respectively. L is the shaft length, which is also the axial location of the bearing measured from point O. GJ_{cs} is the torsional stiffness of the shaft. The virtual work due to the follower torque load, T_L is expressed as

$$\delta W=Q_{T_L}\,\delta q=-T_L\delta\theta_L, \text{ with } \delta\theta_L=\frac{\partial[{}^{\mathbf{N}}\boldsymbol{\omega}^{\mathbf{B}}\ \mathbf{b}_1]^T}{\partial\dot{q}}\delta q \tag{12}$$

where Q_{TL} is the generalized force vector and the degree-of-freedom vector, $\mathbf{q}$, is defined as

$$\mathbf{q}=[v'\ \ w'\ \ \hat{\phi}_{2d}]^T. \tag{13}$$

Also to account for viscous damping, a Rayleigh dissipation function, D, is defined as

$$D=\frac{1}{2}(c_v L^2\dot{v}'^2+c_w L^2\dot{w}'^2+c_\phi\dot{\hat{\phi}}_{2d}^2) \tag{14}$$

Where c_v, c_ω, and c_ϕ, are viscous damping terms. Using the above expressions derived for the kinetic and potential energies along with the virtual work and dissipation terms, the equations of motion are obtained from Lagrange's equations. Furthermore, recognizing the fact that

$$\frac{d}{dt}\left[\frac{\partial\Omega_{2d}}{\partial\dot{q}}\right]-\frac{\partial\Omega_{2d}}{\partial q}=0. \tag{15}$$

The resulting nonlinear equations of motion are shown below in terms of Ω_{2d}.

$$\begin{aligned}
&\frac{d}{dt}\left[\frac{\partial T}{\partial\dot{q}}\right]-\frac{\partial T}{\partial q}+\frac{\partial V}{\partial q}+\frac{\partial D}{\partial\dot{q}}-Q_{T_L}\\
&=\begin{bmatrix} I_m & 0 & 0\\ 0 & I_m & 0\\ 0 & 0 & J_m\end{bmatrix}\ddot{q}+\Omega_{2d}\begin{bmatrix} c_v L^2 & J_m & 0\\ -J_m & c_w L^2 & 0\\ 0 & 0 & c_\phi\end{bmatrix}\dot{q}\\
&+\begin{bmatrix} k_v L^2 & 0 & 0\\ 0 & k_w L^2 & 0\\ 0 & 0 & k_\phi\end{bmatrix}q+(J_m\dot{\Omega}_{2d}+T_L)\begin{bmatrix}\dfrac{\partial\Omega_{2d}}{\partial\dot{v}'}+\dfrac{w'}{2}\\ \dfrac{\partial\Omega_{2d}}{\partial\dot{w}'}-\dfrac{v'}{2}\\ 1\end{bmatrix}\\
&+\frac{J_m}{2}(w'\ddot{v}'-\ddot{w}'v')\begin{bmatrix}\dfrac{\partial\Omega_{2d}}{\partial\dot{v}'}\\ \dfrac{\partial\Omega_{2d}}{\partial\dot{w}'}\\ 0\end{bmatrix}=0
\end{aligned} \tag{16}$$

By using the expressions given in Eqs. (5) and (6) for Ω_{2d}, and taking the necessary partial and time derivatives of Ω_{2d}, the full nonlinear equations are derived.

Since the torque load, T_L, is constant, the equations of motion can be linearized about a static twist angle. Thus, the total elastic twist angle can be written as

$$\hat{\phi}_{2d}\equiv\psi_s+\psi_d \quad\text{with}\quad \psi_s=\frac{T_L}{k_\phi} \tag{17}$$

where ψ_s is the static part of the elastic twist due to T_L, and ψ_d is the relatively small dynamic portion of the elastic twist. After linearizing about the static twist operating condition, the equations of motion still contain many higher order terms. An ordering scheme, based on the order-of-magnitude assumptions shown in Eq. (18), is used to determine the dominant terms.

$$\begin{aligned}
O(\delta_3)=O(\delta_2)=O(\psi_s)&=1e^{-2}\approx 1 \ \text{deg}\\
O(v')=O(w')=O(\psi_d)&=1e^{-4}\approx 0.01 \ \text{deg}
\end{aligned} \tag{18}$$

After dropping terms with order of magnitude smaller than $1e^{-6}$, and nondimensionalizing, the result is the nondimensional linear periodic system shown in Eqs. (19) through (21). Here the "*" operator indicates differentiation with respect to nondimensional time.

$$\begin{aligned}
&[\mathbf{I}+\mathbf{M}_0+\mathbf{M}_{s2}\sin 2\phi+\mathbf{M}_{c2}\cos 2\phi]\overset{**}{\mathbf{x}}\\
&+[\mathbf{C}_{sd}+\mathbf{C}_{\text{aux}}+\mathbf{G}+\mathbf{C}_{s2}\sin 2\phi+\mathbf{C}_{c2}\cos 2\phi]\overset{*}{\mathbf{x}}\\
&+[\Lambda+\mathbf{K}_0+\mathbf{K}_{s2}\sin 2\phi+\mathbf{K}_{c2}\cos 2\phi]\mathbf{x}\\
&=\mathbf{F}_0+\mathbf{F}_{s2}\sin 2\phi+\mathbf{F}_{c2}\cos 2\phi+O(v'\delta^2,v'^2)
\end{aligned} \tag{19}$$

The nominal system consists of $\mathbf{I}$, $\mathbf{C}_{sd}$, $\mathbf{G}$, and Λ, which are the nondimensional mass, structural damping, gyroscopic, and stiffness matrices, respectively. The matrix $\mathbf{C}_{\text{aux}}$, accounts for the damping provided by an auxiliary viscous damper. See Eq. (20).

$$\begin{aligned}
\mathbf{I}&=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\quad \mathbf{C}_{sd}=\begin{bmatrix}2\zeta_v f_v&0&0\\0&2\zeta_w f_w&0\\0&0&2\zeta_\phi f_\phi\end{bmatrix},\\
\mathbf{C}_{\text{aux}}&=\begin{bmatrix}c_d&0&0\\0&c_d&0\\0&0&0\end{bmatrix},\quad \mathbf{G}=\begin{bmatrix}0&f\eta&0\\-f\eta&0&0\\0&0&0\end{bmatrix},\\
\Lambda&=\begin{bmatrix}f_v^2&0&0\\0&f_w^2&0\\0&0&f_\phi^2\end{bmatrix}\\
&\text{with}\quad \mathbf{x}=[v' \ \ w' \ \ \psi_d]^T
\end{aligned} \tag{20}$$

The matrices in Eq. (19) that account for the static angular misalignments, δ_2 and δ_3, and static load-torque, T_L, are defined in Eq. (21).

$$\begin{aligned}
\mathbf{M}_0&=\frac{1}{2}\begin{bmatrix}0&0&\delta_2\eta\\0&0&-\delta_3\eta\\\delta_2&-\delta_3&0\end{bmatrix},\\
\mathbf{M}_{s2}&=\frac{1}{2}\begin{bmatrix}0&0&\delta_3\eta\\0&0&-\delta_2\eta\\\delta_3&-\delta_2&0\end{bmatrix},\quad \mathbf{M}_{c2}=\frac{1}{2}\begin{bmatrix}0&0&-\delta_2\eta\\0&0&-\delta_3\eta\\-\delta_2&-\delta_3&0\end{bmatrix}\\
\mathbf{C}_{s2}&=\begin{bmatrix}0&0&0\\0&0&0\\2f\delta_2&2f\delta_3&0\end{bmatrix},\quad \mathbf{C}_{c2}=\begin{bmatrix}0&0&0\\0&0&0\\2f\delta_3&-2f\delta_2&0\end{bmatrix},
\end{aligned} \tag{21}$$

$$\begin{aligned}
\mathbf{K}_0&=\begin{bmatrix}0&\tau&0\\-\tau&0&0\\0&0&0\end{bmatrix},\quad \mathbf{K}_{s2}=\begin{bmatrix}\tau&0&0\\0&-\tau&0\\-2f^2\delta_3&2f^2\delta_2&0\end{bmatrix},\\
\mathbf{K}_{c2}&=\begin{bmatrix}0&-\tau&0\\-\tau&0&0\\2f^2\delta_2&2f^2\delta_3&0\end{bmatrix}
\end{aligned}$$

The forcing terms due to misalignment and load torque on the right-hand side of Eq. (19) are shown in Eq. (22).

$$\begin{aligned}
\mathbf{F}_0&=\tau\begin{bmatrix}-\delta_2\\\delta_3\\0\end{bmatrix}+\frac{f^2\eta}{4}\begin{bmatrix}\delta_2^2\delta_3+\delta_3^3\\\delta_3^2\delta_2+\delta_2^3\\0\end{bmatrix},\\
\mathbf{F}_{s2}&=\tau\begin{bmatrix}-\delta_3\\\delta_2\\0\end{bmatrix}+\frac{f^2\eta}{2}\begin{bmatrix}\delta_3^2\delta_2-\delta_2^3\\\delta_2^2\delta_3-\delta_3^3\\2\delta_3^2-2\delta_2^2\end{bmatrix},\\
\mathbf{F}_{c2}&=\tau\begin{bmatrix}\delta_2\\\delta_3\\0\end{bmatrix}+f^2\eta\begin{bmatrix}-\delta_2^2\delta_3\\\delta_3^2\delta_2\\-2\delta_2\delta_3\end{bmatrix}
\end{aligned} \tag{22}$$

The equations have been nondimensionalized with respect to a reference frequency, Ω_0, such that the nondimensional shaft speed is $f=\Omega/\Omega_0$ and f_v, f_ω, and f_ϕ are the nondimensional natural frequencies. The modal damping ratios due to structural damping are denoted by ζ_v, ζ_ω, and ζ_ϕ, and the nondimensional

damping coefficient of the auxiliary damper is c_d. η is the torsional to lateral inertia ratio and τ is the nondimensional load-torque parameter, see Eq. (23).

$$\Omega_0=\sqrt{\frac{(k_v+k_w)L^2}{2I_m}},\quad c_d=\frac{C_dL^2}{I_m\Omega_0},\quad \eta=\frac{J_m}{I_m},\quad \tau=\frac{T_L}{2I_m\Omega_0^2}$$
$$f_v=\sqrt{\frac{k_vL^2}{I_m\Omega_0^2}},\quad f_w=\sqrt{\frac{k_wL^2}{I_m\Omega_0^2}},\quad f_\phi=\sqrt{\frac{k_\phi}{J_m\Omega_0^2}} \tag{23}$$

The static torque load, T_L, creates the stiffness coupling matrices, $\mathbf{K}_0$, $\mathbf{K}_{s2}$, and $\mathbf{K}_{c2}$, and the dynamic torque, $J_m\dot{\Omega}_{2d}$, creates the inertia coupling matrices, $\mathbf{M}_0$, $\mathbf{M}_{s2}$, and $\mathbf{M}_{c2}$.

Several more physically significant nondimensional parameters are defined for a solid circular crosssection shaft. First, the inertia ratio, η, is rewritten in terms of the following nondimensional parameters.

$$l_d=\frac{L_d}{L},\quad \varepsilon_s=\frac{r_s}{L},\quad \gamma=\frac{J_d}{J_{ms}},\quad \mu=\frac{m_d}{m_s},$$
$$\eta=\frac{J_m}{I_m}=\frac{2(1+\gamma)\varepsilon^2}{(1+\gamma)\varepsilon^2+\frac{4}{3}+4l_d\mu} \tag{24}$$

l_d is the nondimensional axial distance of the disk from point O. ε_s is the shaft slenderness ratio with shaft radius r_s. γ is the polar inertia ratio between the disk and the shaft, and μ is the mass ratio between the disk and the shaft. Also, the torsion-lateral stiffness ratio, Δ, and nondimensional torsion natural frequency, f_ϕ, are defined as

$$\Delta=\frac{k_\phi}{(k_v+k_w)\frac{L^2}{2}},\quad f_\phi^2=\frac{\Delta}{\eta}. \tag{25}$$

Also, since the reference frequency, Ω_0, is the root mean of the two lateral frequencies, the two nondimensional lateral frequencies, f_v and f_ω, are always evenly split about 1. Thus, the nondimensional lateral frequencies can be written as

$$f_v^2=1+\lambda \quad \text{and} \quad f_w^2=1-\lambda \tag{26}$$

where λ is the lateral mode frequency-split parameter. Since only rigid lateral modes are considered, the shaft bending stiffness must be higher than the bearing support stiffness. Also the load torque, τ, should be less than the load torque that exceeds the shear yield strain of the shaft material, $\tau_{\max}$. These two requirements are summarized below.

$$1)\ \frac{1}{2}[k_v+k_w]<\frac{3EI_{cs}}{L^3}\Rightarrow 3(1+\nu)l_df_\phi^2\eta>1$$
$$2)\ \tau<\tau_{\max}\ \text{ where }\ \tau_{\max}=\gamma_{\text{yield}}\frac{l_df_\phi^2\eta}{2\varepsilon} \tag{27}$$

Here ν is the Poisson's ratio, which relates the elastic modulus, E, to the shear modulus, G, and γ_{yield} is the shear yield strain of the shaft material.

4 Stability Analysis

Since the equations of motion shown in Eq. (19) are linear and periodic, Floquet theory is used to assess the stability of the system [9]. The stability is determined numerically by examining the eigenvalues of the Floquet transition matrix (FTM). This technique is numerically intensive, but deemed necessary to capture all the instability behavior of the equations of motion.

With equations in Eq. (19) cast in first order form and the forcing terms set to zero, the system is written as

$$\dot{\mathbf{X}}=\mathbf{A}(t)\mathbf{X},\ \text{with}\ \mathbf{X}=[\mathbf{x}\ \dot{\mathbf{x}}]^T\quad \text{and}\quad \mathbf{A}(T+t)=\mathbf{A}(t) \tag{28}$$

where $\mathbf{A}(t)$ is $2n\times 2n$ system matrix, with $n=3$. T is the period, which in this case is $T=\pi/\Omega$, and in the nondimensional case, $T^*=\pi/f$. Next, the FTM matrix, denoted by $\Phi(T^*)$, is generated, where

$$\Phi(T^*)=[\{\mathbf{x}_1(T^*)\},\{\mathbf{x}_2(T^*)\},\ldots,\{\mathbf{x}_{2n}(T^*)\}] \tag{29}$$

and $[\{\mathbf{x}_1(t)\},\{\mathbf{x}_2(t)\},\ldots,\{\mathbf{x}_{2n}(t)\}]$ are the $2n$ linearly independent solutions obtained by numerically integrating Eq. (28) from 0 to T^* with the following initial conditions:

$$\Phi(0)=\begin{bmatrix}1&0&\cdots&0\\0&1&\cdots&0\\&&\ddots&0\\0&0&0&1\end{bmatrix}_{2n\times 2n}. \tag{30}$$

The FTM matrix, $\Phi(T^*)$, maps the state of the system from some initial state, $\mathbf{X}_0$, to the state at time $t=kT^*$, such that $\mathbf{X}(kT^*)=\Phi(T^*)^k\mathbf{X}_0$. Thus the eigenvalues, λ_i, of $\Phi(T^*)$, which govern the stability of the mapping, also determine the stability of the system.

$$\frac{\ln\lambda_i}{T^*}=\alpha_i+j\omega_i\quad\begin{matrix}\text{stable}&\text{if}\ \alpha_i\leqslant 0\\ \text{unstable}&\text{if}\ \alpha_i>0\end{matrix}\quad \text{for}\ [i=1,2\ldots 2n] \tag{31}$$

where α_i is the effective damping of the time varying system. Furthermore, the equivalent modal damping ratio of the ith mode, z_i, is defined as

$$z_i=-\frac{\alpha_i}{\omega_i}. \tag{32}$$

As discussed in [10], it is expected that parametric instabilities may occur when the parametric excitation frequency, which is twice the driving shaft speed in this case, is in the neighborhood of the principle, sum combination and difference combination frequencies. Thus, the potential parametric instability zones are written as

$$2f=|f_{n_i}\pm f_{n_j}|+p,\quad [i,j=1,2,3,\ldots]. \tag{33}$$

Here the f_{ni}'s are the natural frequencies of the linear, time invariant portion of the system and p is a small frequency detuning parameter.

In this portion of the investigation, the effects of static load-torque, lateral frequency-split, and static operating misalignment angle on the stability of the system are examined. In terms of the nondimensional parameters, the individual and combined effects of τ, λ, δ_2, and δ_3 are studied. The damping is assumed to be purely structural with $\zeta_v=\zeta_w=\zeta_\phi=\zeta=0.01$ and no auxiliary damping, $c_d=0.0$.

The τ-f stability boundary calculated for the system with no static misalignment, i.e., $\delta_2=\delta_3=0$ deg is shown in Fig. 4. Other researchers [1,2] have shown that the constant skew-symmetric matrix, $\mathbf{K}_0$, due to the follower torque load, τ, causes flutter instability for sufficient torque and shaft speeds. The flutter stability boundary is a continuous curve that bisects the τ-f parameter space into a stable and an unstable region. It was shown in [1] that the lateral frequency-split, λ, increases the critical torque required to induce flutter. Additionally, the parametric stiffness matrices due to torque transmission through the U-joint, $\mathbf{K}_{s2}$ and $\mathbf{K}_{c2}$, induce parametric instability in frequency range I. Specifically the torque-induced parametric instabilities occur near the frequencies f_w, f_v, and $(f_v+f_w)/2$.

Figure 5 depicts how static misalignment between the driving and driven shafts affects stability. As shown in Eqs. (19) through (21), the static misalignment angles, δ_2 and δ_3, give rise to inertia

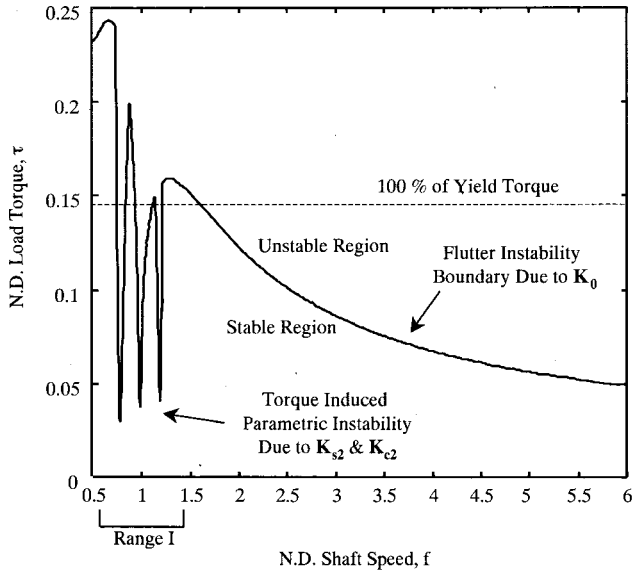


Fig. 4 τ-f stability boundary with no misalignment. $\delta_2=\delta_3=0$ deg, $\lambda=0.4$, $\Delta=3$, $\varepsilon_s=0.05$, $I_d=0.5$, $\mu=0.1$, $\gamma=10$, $c_d=0.0$.

coupling matrices, $\mathbf{M}_0$, $\mathbf{M}_{s2}$, and $\mathbf{M}_{c2}$, which couple the torsion and lateral degrees-of-freedom. The two parametric inertia coupling matrices, $\mathbf{M}_{s2}$ and $\mathbf{M}_{c2}$, alter the stability in the neighborhood of the torsion-lateral combination frequencies in range II, however, the torque-induced parametric instability zones in range I are not affected. The four torsion-lateral combination frequencies in range II are defined as

$$f_1=\frac{f_\phi-f_v}{2},\quad f_2=\frac{f_\phi-f_w}{2},\quad f_3=\frac{f_\phi+f_w}{2},\quad f_4=\frac{f_\phi+f_v}{2}. \tag{34}$$

Here $f_1\leqslant f_2\leqslant f_3\leqslant f_4$ where f_1 and f_2 are the difference-type combinations and f_3 and f_4 are the sum-type combinations. Since $f_v^2=1+\lambda$ and $f_w^2=1-\lambda$, combinations involving f_v are called stiff-mode combinations and those involving f_w are called soft-mode combinations. Since misalignment only affects stability near the torsion-lateral combination frequencies, a more detailed plot of the τ-f stability region over frequency range II is shown in Fig. 6.

The stability for several static misalignment angles is examined for both positive and negative values of the follower torque, τ. A positive τ is a load-torque that resists the rotation, and a negative τ is a driving torque that is applied in the same direction as the rotation. In most practical situations, $\tau>0$, since this corresponds to power being transmitted from the driving shaft to the driven shaft.

Misalignment is stabilizing for shaft speeds near the torsion-lateral difference combination frequencies, f_1 and f_2, and is destabilizing for speeds near the torsion-lateral sum combination frequencies, f_3 and f_4. Specifically, misalignment can be stabilizing since it increases the magnitude of the critical load-torque required to causes instability near $f=f_1$ and near $f=f_2$ on the upper and lower τ-f stability boundaries, respectively. On the other hand, misalignment can be destabilizing since it lowers the magnitude of the critical load-torque near $f=f_3$ and near $f=f_4$ on the lower and upper τ-f stability boundaries, respectively. Furthermore, even when $\tau=0$, misalignment alone can be sufficient to cause instability. This is shown in Figs. 5 and 6, where instability zones of finite width are present for $\tau=0$ near the torsion-lateral sum combination frequencies, f_3 and f_4.

Figure 7, shows the stabilizing effect of the lateral frequency-split, λ, on the upper τ-f stability boundary, $\tau\geqslant 0$, with $\delta_2=2$ deg over the shaft speed range of the torsion-lateral combination frequencies. λ has a stabilizing effect since increasing λ increases the critical destabilizing load torque, τ, across the entire shaft speed range. Misalignment causes the minimum destabilizing load-torque to occur near the sum-type, torsion-lateral combination frequency, f_4, regardless of λ. Thus the least stable configuration is when the lateral frequencies are equal, $\lambda=0.0$, and the least-stable operating speed is in the vicinity of the sum-type torsion-lateral combination frequency f_4 when $\tau\geqslant 0$.

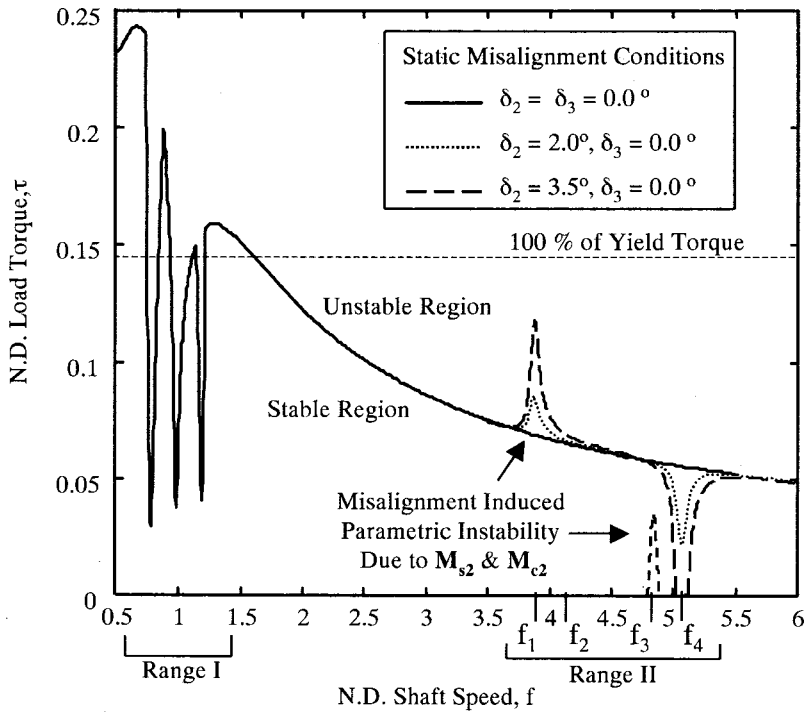


Fig. 5 τ-f stability boundary for several misalignments. $\lambda=0.4$, $\Delta=3$, $\varepsilon_s=0.05$, $I_d=0.5$, $\mu=0.1$, $\gamma=10$, $c_d=0.0$.

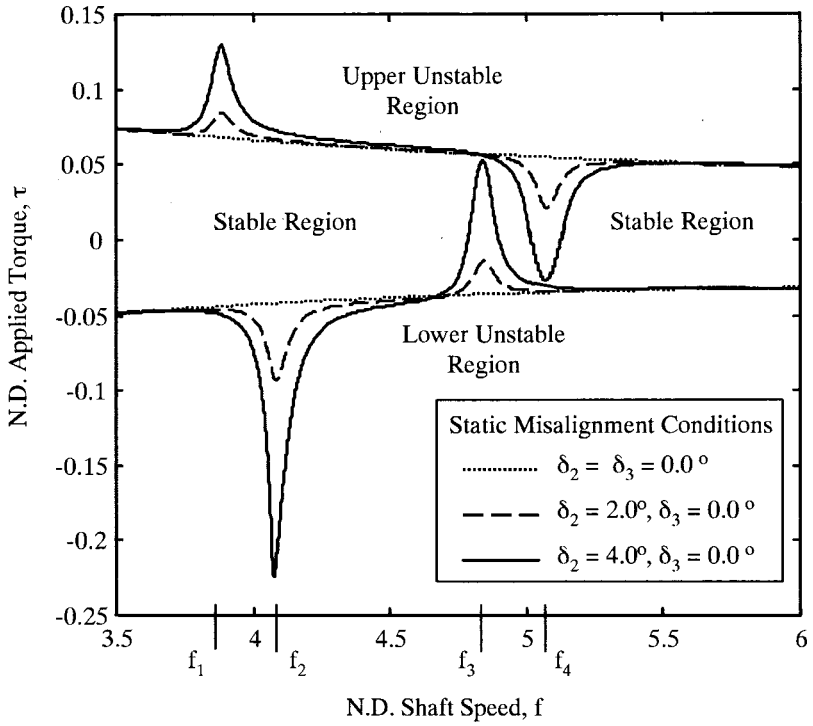


Fig. 6 τ-f **stability boundary for several misalignments.** $\lambda=0.4$, $\Delta=3$, $\varepsilon_s=0.05$, $I_d=0.5$, $\mu=0.1$, $\gamma=10$, $c_d=0.0$.

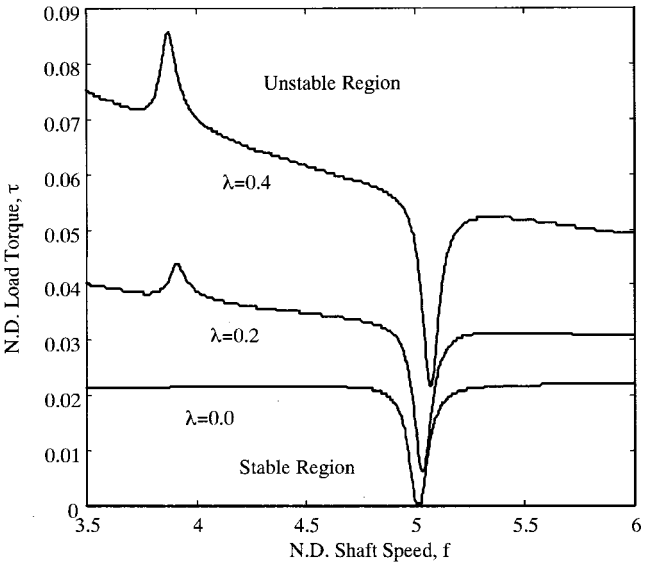


Fig. 7 τ-f **stability boundary for several** λ. $\delta_2=2.0$ **deg,** $\delta_3=0.0$ **deg,** $\Delta=3$, $\varepsilon_s=0.05$, $I_d=0.5$, $\mu=0.1$, $\gamma=10$, $c_d=0.0$.

5 Stabilization via Auxiliary Lateral Damping

In the previous section it is shown that misalignment causes instability when f is near the torsion-lateral sum combination frequencies, f_4 and f_3, for positive and negative values of load-torque, respectively. Therefore, if shaft alignment cannot be guaranteed due to external factors such as foundation deflection, etc., then it is important to avoid these frequency zones during shaft operation. However, avoiding these zones maybe too restrictive for certain supercritical shaft applications, especially if the shaft operating speed must be variable. Therefore, in this section, the use of auxiliary damping as a means of stabilizing the system is investigated. The system is augmented with an auxiliary viscous damper that dissipates energy from lateral motion of the shaft and has a nondimensional damping coefficient c_d. See Eqs. (20) and (23). The total system damping is due to both the inherent structural damping and the damping provided by the auxiliary damper.

Figure 8 shows the required auxiliary nondimensional damping coefficient, c_{req}, necessary to guarantee stability for several combinations of load-torque, τ, and static misalignment, δ_2, over the shaft speed range covering the torsion-lateral combination frequencies.

When misalignment is present with no load-torque, only parametric instability occurs, hence auxiliary damping is only required for shaft operating speeds near sum-type torsion-lateral combination frequencies f_3 and f_4. When both misalignment and load-torque are present, parametric and flutter instability occur, hence auxiliary damping is required across the entire shaft speed range shown. The stabilizing effect of the misalignment near $f=f_1$ is shown by a reduction in required auxiliary damping near the shaft speed f_1. The maximum required auxiliary damping coefficient, $c_{\text{max-req}}$, occurs at the torsion-lateral combination frequency f_4 for all $\tau\geqslant 0$. Thus, the damping coefficient of the auxiliary damper, c_d, must be at least as high as the required auxiliary damping at $f=f_4$ in order to guarantee stability for all shaft operating speeds. This condition is summarized in Eq. (35).

$$c_{\text{max-req}}\equiv\max_{f\geqslant 0}[c_{\text{req}}(f)]=c_{\text{req}}(f_4) \tag{35}$$

$$\text{for stablilty } c_d\geqslant c_{\text{max-req}}$$

Figure 9 shows how $c_{\text{max-req}}$ varies with the degree of angular misalignment, δ_2, for several values of load-torque, τ.

The auxiliary damping coefficient required to guarantee stability increases rapidly with the degree of angular misalignment. This is because the dynamic load-torque terms, accounted for by the inertia coupling matrices $\mathbf{M}_0$, $\mathbf{M}_{s2}$, and $\mathbf{M}_{c2}$, are proportional to the angular misalignment. Additionally, $c_{\text{max-req}}$, also increases with load torque, τ.

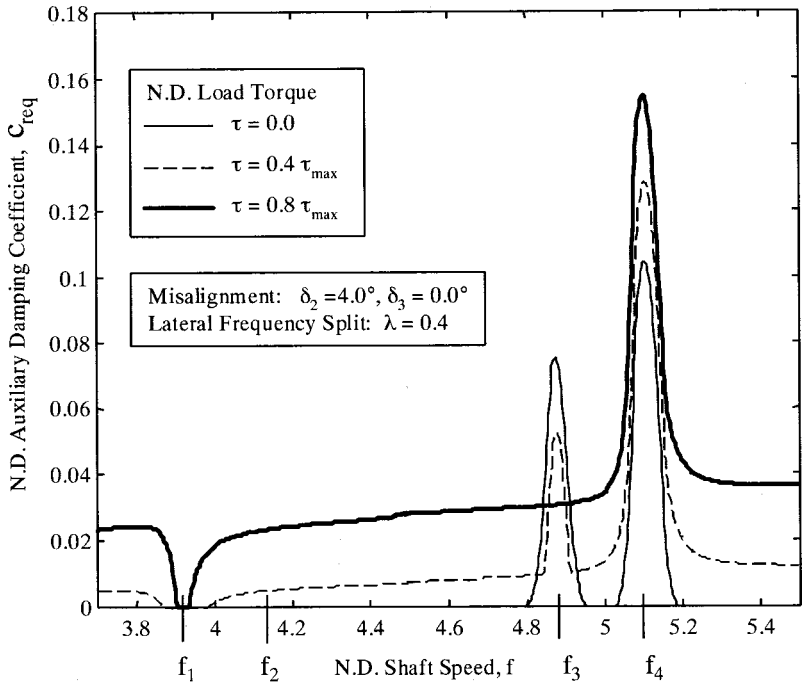


Fig. 8 Required auxiliary damping for stability versus shaft speed for several τ. ζ=0.005, λ=0.4, Δ=3, ε_s=0.05, I_d=0.5, μ=0.1, γ=10.

Finally, Fig. 10 shows how $c_{\text{max-req}}$ varies with misalignment for several disk-shaft polar inertia ratios, γ, with τ=0. As the disk-shaft polar inertia ratio, γ, is increased, more damping is required to stabilize the system for a given angular misalignment. Physically, this is because the dynamic torque load generated by misalignment-induced speed variation is proportional to the polar inertia load carried by the shaft.

6 Perturbation Analysis of Least Stable Case

Since determining stability by evaluating the Floquet transition matrix is computationally intensive, a stability criteria based on a perturbation approach is derived for the least stable case (worst case), i.e., $f=f_4$ and λ=0. Since the governing equation-of-motion, Eq. (19), is linear and periodically time-varying, the deri-

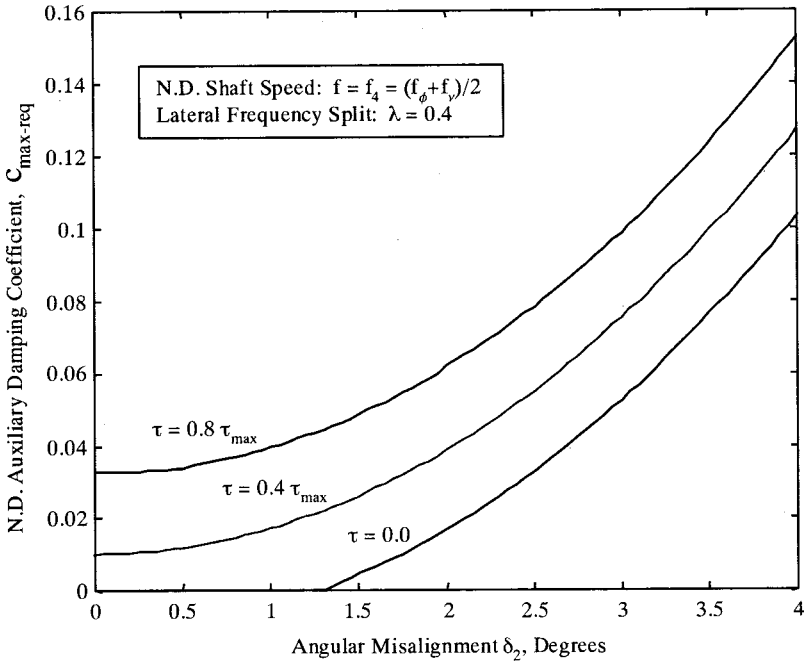


Fig. 9 Maximum required auxiliary damping for stability versus misalignment for several τ. $f=f_4$, $\lambda\lambda$=0.4, ζ=0.005, Δ=3, ε_s=0.05, I_d=0.5, μ=0.1 γ=10.

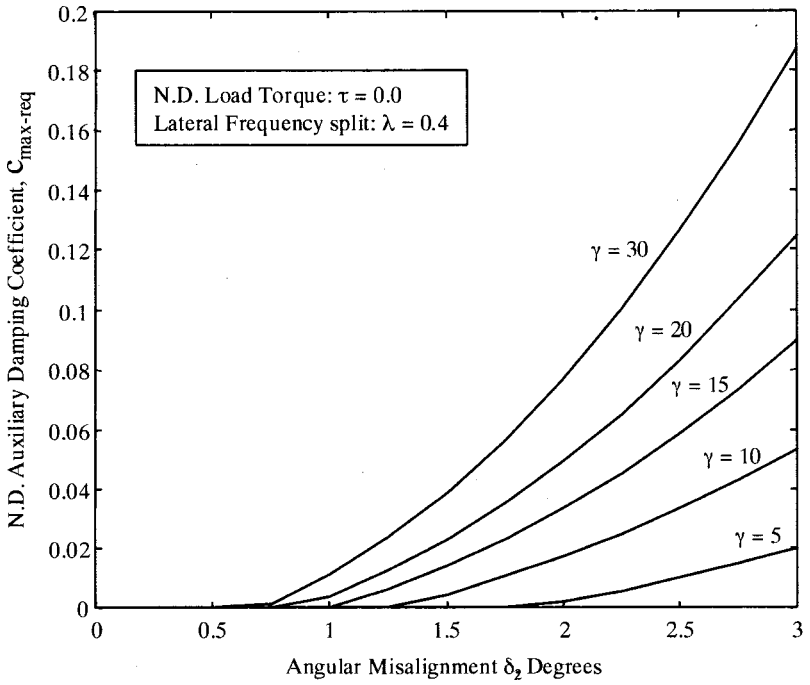


Fig. 10 Maximum required auxiliary damping for stability versus misalignment for several γ. $f=f_4$, $\lambda=0.4$, $\zeta=0.005$, $\Delta=3$, $\varepsilon_s=0.05$, $I_d=0.5$, $\mu=0.1$.

vation is based on a technique developed by Hsu in references [10], [11]. In order to apply Hsu's method, the equation must be in the following form:

$$\mathbf{I}\frac{d^2\mathbf{x}}{dt^2}+\varepsilon\mathbf{C}(t)\frac{d\mathbf{x}}{dt}+\mathbf{K}(t)\mathbf{x}=\mathbf{0},$$
$$\mathbf{C}(t)=\mathbf{C}(t+T) \quad \text{and} \quad \mathbf{K}(t)=\mathbf{K}(t+T). \tag{36}$$

However, the system in Eq. (19) does not have the same form as Eq. (36) because (19) contains inertia coupling terms, $\mathbf{M}_0$, $\mathbf{M}_{s2}$, and $\mathbf{M}_{c2}$, that result in an additional periodically time-varying inertia matrix $\mathbf{M}(t)$. In order to eliminate $\mathbf{M}(t)$ and get the equations into the same form as Eq. (36), Eq. (19) is multiplied through by an expression for $\mathbf{M}(t)^{-1}$. The expression for $\mathbf{M}(t)^{-1}$, which is valid to the first order, is given in Eq. (37).

$$\mathbf{M}(\mathbf{t})=[\mathbf{I}+\varepsilon\mathbf{M}_\Delta(\mathbf{t})]$$
$$\varepsilon\mathbf{M}_\Delta(\mathbf{t})=\mathbf{M}_0+\mathbf{M}_{s2}\sin 2\phi+\mathbf{M}_{c2}\cos 2\phi \tag{37}$$
$$\mathbf{M}(\mathbf{t})^{-1}=[\mathbf{I}+\varepsilon\mathbf{M}_\Delta(\mathbf{t})]^{-1}=\mathbf{I}-\varepsilon\mathbf{M}_\Delta(\mathbf{t})+O(\varepsilon^2)$$

Multiplying Eq. (19) by the approximate expression for $\mathbf{M}(t)^{-1}$ yields Eq. (38).

$$\mathbf{I}\overset{**}{\mathbf{x}}+\mathbf{M}(\mathbf{t})^{-1}\mathbf{C}(\mathbf{t})\overset{*}{\mathbf{x}}+\mathbf{M}(\mathbf{t})^{-1}\mathbf{K}(\mathbf{t})\mathbf{x}=\mathbf{0} \tag{38}$$

Which can be rewritten as

$$\mathbf{I}\overset{**}{\mathbf{x}}+\Lambda\mathbf{x}=-\varepsilon[\mathbf{c}_0+\mathbf{c}_{c2}\cos(2\phi)+\mathbf{c}_{s2}\sin(2\phi)]\overset{*}{\mathbf{x}}$$
$$-\varepsilon[\mathbf{k}_0+\mathbf{k}_{c2}\cos(2\phi)+\mathbf{k}_{s2}\sin(2\phi)]\mathbf{x} \tag{39}$$

where the perturbation terms $\mathbf{c}_0$, $\mathbf{c}_{c2}$, $\mathbf{c}_{s2}$, $\mathbf{k}_0$, $\mathbf{k}_{c2}$ and $\mathbf{k}_{s2}$ are defined in the Appendix. Next each row of the matrix equation in Eq. (39) is written in first-order form

$$\overset{*}{x}_i=y_i$$
$$y_i+\omega_i^2x_i=-\varepsilon\sum_{j=1}^{3}[c_{0_{ij}}+c_{c2_{ij}}\cos(2\phi)+c_{s2_{ij}}\sin(2\phi)]y_j$$
$$-\varepsilon\sum_{j=1}^{3}[k_{0_{ij}}+k_{c2_{ij}}\cos(2\phi)+k_{s2_{ij}}\sin(2\phi)]x_j \tag{40}$$

for $i=1$, 2, and 3, where ω_i are the nominal system natural frequencies

$$\omega_1=f_v, \quad \omega_2=f_w \quad \text{and} \quad \omega_3=f_\phi$$

Thus

$$f_3=\frac{\omega_3+\omega_2}{2} \quad \text{and} \quad f_4=\frac{\omega_3+\omega_1}{2}. \tag{41}$$

Following Hsu's approach and taking the form of the solution to be

$$x_i=A_i(t)\cos(\omega_i t)+B_i(t)\sin(\omega_i t)+\varepsilon\hat{x}_i(t)$$
$$y_i=\omega_i[-A_i(t)\sin(\omega_i t)+B_i(t)\cos(\omega_i t)]+\varepsilon\overset{*}{\hat{x}}_i(t) \tag{42}$$

the first two terms involving A_i and B_i on the right-hand side of Eq. (42) are the variational terms. The remaining terms involving ε are the perturbation terms. Substituting the assumed solution of Eq. (42) into Eq. (40) gives

$$\overset{*}{A}_i \cos(\omega_i t) + \overset{*}{B}_i \sin(\omega_i t) = 0$$

$$\overset{*}{B}_i \cos(\omega_i t) - \overset{*}{A}_i \sin(\omega_i t) + \varepsilon(\overset{**}{\hat{x}}_i + \omega_i^2 \hat{x}_i)$$

$$= -\frac{\varepsilon}{\omega_i} \sum_{j=1}^{3} [X_c^{ij} \cos(\omega_j t) + X_s^{ij} \sin(\omega_j t)]$$

$$-\frac{\varepsilon}{2\omega_i} \sum_{j=1}^{3} [H_1^{ij} \cos(\omega_j t + 2ft) + H_2^{ij} \cos(\omega_j t - 2ft)] \quad (43)$$

$$-\frac{\varepsilon}{2\omega_i} \sum_{j=1}^{3} [H_3^{ij} \sin(\omega_j t + 2ft) + H_4^{ij} \sin(\omega_j t - 2ft)$$

For $i = 1$, 2, and 3, where

$$X_c^{ij} = k_{0_{ij}} A_j + \omega_j c_{0_{ij}} B_j \quad \text{and} \quad X_s^{ij} = k_{0_{ij}} B_j - \omega_j c_{0_{ij}} A_j$$

$$H_1^{ij} = k_{c2_{ij}} A_j - k_{s2_{ij}} B_j + \omega_j c_{c2_{ij}} B_j + \omega_j c_{s2_{ij}} A_j$$

$$H_2^{ij} = k_{c2_{ij}} A_j + k_{s2_{ij}} B_j + \omega_j c_{c2_{ij}} B_j - \omega_j c_{s2_{ij}} A_j \quad (44)$$

$$H_3^{ij} = k_{c2_{ij}} B_j + k_{s2_{ij}} A_j - \omega_j c_{c2_{ij}} A_j + \omega_j c_{s2_{ij}} B_j$$

$$H_4^{ij} = k_{c2_{ij}} B_j - k_{s2_{ij}} A_j - \omega_j c_{c2_{ij}} A_j - \omega_j c_{s2_{ij}} B_j .$$

The key step in determining the stability of Eq. (43), as done in [10,11], is to examine the stability of the variational and perturbation equations separately. To do this, all terms on the right-hand side of Eq. (43), except those that would lead to a resonance or near resonance solution, are associated with the perturbation part. By treating the A_is and B_is as constant, the perturbation part becomes a linear constant-coefficient differential equation, whose stability is completely determined by the sign of the ω_is, the system natural frequencies. Since $\omega_i > 0$, the perturbation equation is stable, hence the overall stability of Eq. (43) is governed by the stability of the variational part.

In order to proceed with the stability analysis of the variational part, all the resonance producing terms on the right-hand side of Eq. (43) must be identified. Since we are interested in determining the stability in the vicinity of the least stable shaft speed, $f = f_4$ with least-stable value of lateral frequency-split, $\lambda = 0$, we have

$$\text{for } \lambda = 0, \ \omega_1 = \omega_2 = 1, \quad \text{and} \quad f_3 = f_4 = \frac{\omega_3 + \omega_1}{2} = \frac{\omega_3 + \omega_2}{2}. \quad (45)$$

It is apparent that if f is nearly equal to $(\omega_j \pm \omega_k)/2$, then certain terms on the right side of Eq. (43) become resonance producing terms for several sets of j and k, see Eq. (46).

$$f = \frac{\omega_j + \omega_k}{2} + \rho \quad \text{with} \quad \begin{matrix} k=1, \ j=3 \\ k=2, \ j=3 \\ k=3, \ j=1 \end{matrix} \quad \text{and} \quad 2 \quad (46)$$

Where ρ is a small real number that acts as a frequency detuning parameter. Next, substituting Eq. (46) into Eq. (43) for each of the three separate sets of j and k and taking only the resonance producing terms on the right-hand side, we obtain three pairs of coupled variational equations which govern the stability about the shaft speed $f = f_3 = f_4$. See Eq. (47).

$$\overset{*}{A}_1 \cos(\omega_1 t) + \overset{*}{B}_1 \sin(\omega_1 t) = 0, \quad k = 1 \quad (47a)$$

$$-\overset{*}{A}_1 \sin(\omega_1 t) + \overset{*}{B}_1 \cos(\omega_1 t) = -\frac{\varepsilon}{\omega_1} [X_c^{11} \cos(\omega_1 t) + X_s^{11} \sin(\omega_1 t) + X_c^{12} \cos(\omega_2 t) + X_s^{12} \sin(\omega_2 t)] - \frac{\varepsilon}{2\omega_1} [H_2^{13} \cos(\omega_1 t + 2\rho t) - H_4^{13} \sin(\omega_1 t + 2\rho t)]$$

$$\overset{*}{A}_2 \cos(\omega_2 t) + \overset{*}{B}_2 \sin(\omega_2 t) = 0, \quad k = 2 \quad (47b)$$

$$-\overset{*}{A}_2 \sin(\omega_2 t) + \overset{*}{B}_2 \cos(\omega_2 t) = -\frac{\varepsilon}{\omega_2} [X_c^{21} \cos(\omega_1 t) + X_s^{21} \sin(\omega_1 t) + X_c^{22} \cos(\omega_2 t) + X_s^{22} \sin(\omega_2 t)] - \frac{\varepsilon}{2\omega_2} [H_2^{23} \cos(\omega_2 t + 2\rho t) - H_4^{23} \sin(\omega_2 t + 2\rho t)]$$

$$\overset{*}{A}_3 \cos(\omega_3 t) + \overset{*}{B}_3 \sin(\omega_3 t) = 0, \quad k = 3 \quad (47c)$$

$$-\overset{*}{A}_3 \sin(\omega_3 t) + \overset{*}{B}_3 \cos(\omega_3 t)$$

$$= -\frac{\varepsilon}{\omega_3} [X_c^{33} \cos(\omega_3 t) + X_s^{33} \sin(\omega_3 t)] - \frac{\varepsilon}{2\omega_3} [(H_2^{31} + H_2^{32}) \times \cos(\omega_3 t + 2\rho t) - (H_4^{31} + H_4^{32}) \sin(\omega_3 t + 2\rho t)]$$

This is a more complicated situation than shown in [10] since Hsu studies the case where all the eigenvalues are distinct, thus only two pairs variational equations are needed to determine the stability around any particular combination frequency. In this case, since $\lambda = 0$ leads to the multiplicity $\omega_1 = \omega_2$, three pairs of variational equations are needed.

The variational equations in Eq. (47) are solved by averaging and then transformed to a set of complex coordinates. Specifically, (47) is rearranged so that dA_i/dt and dB_i/dt are on the left-hand sides and then both sides of the equations are integrated from 0 to 2π. The resulting averaged variational equations are shown in Eq. (48).

$$\overset{*}{A}_1 = \frac{\varepsilon}{\omega_1} \left[\frac{1}{2} (X_s^{11} + X_s^{12}) - \frac{1}{4} H_2^{13} \sin(2\rho t) - \frac{1}{4} H_4^{13} \cos(2\rho t) \right] \quad (48a)$$

$$\overset{*}{B}_1 = -\frac{\varepsilon}{\omega_1} \left[\frac{1}{2} (X_c^{11} + X_c^{12}) + \frac{1}{4} H_2^{13} \cos(2\rho t) - \frac{1}{4} H_4^{13} \sin(2\rho t) \right]$$

$$\overset{*}{A}_2 = \frac{\varepsilon}{\omega_2} \left[\frac{1}{2} (X_s^{21} + X_s^{22}) - \frac{1}{4} H_2^{23} \sin(2\rho t) - \frac{1}{4} H_4^{23} \cos(2\rho t) \right] \quad (48b)$$

$$\overset{*}{B}_2 = -\frac{\varepsilon}{\omega_2} \left[\frac{1}{2} (X_c^{21} + X_c^{22}) + \frac{1}{4} H_2^{23} \cos(2\rho t) - \frac{1}{4} H_4^{23} \sin(2\rho t) \right]$$

$$\overset{*}{A}_3=\frac{\varepsilon}{\omega_3}\left[\frac{1}{2}X_s^{33}-\frac{1}{4}(H_2^{31}+H_2^{32})\sin(2\rho t)-\frac{1}{4}(H_4^{31}+H_4^{32})\cos(2\rho t)\right] \tag{48c}$$

$$\overset{*}{B}_3=-\frac{\varepsilon}{\omega_3}\left[\frac{1}{2}X_c^{33}+\frac{1}{4}(H_2^{31}+H_2^{32})\cos(2\rho t)-\frac{1}{4}(H_4^{31}+H_4^{32})\sin(2\rho t)\right]$$

Next the averaged equations in Eq. (48) are transformed to the following complex coordinates:

$$X_1=A_1+jB_1 \quad Y_1=A_2+jB_2 \quad Z_1=A_3+jB_3$$
$$X_2=A_1-jB_1 \quad Y_2=A_2-jB_2 \quad Z_2=A_3-jB_3. \tag{49}$$

The transformed, averaged, variational equations become

$$\overset{*}{x}_1=-\frac{\varepsilon}{2}\left[c_{0_{11}}+j\frac{k_{0_{11}}}{\omega_1}\right]X_1-\frac{\varepsilon}{2\omega_1}[\omega_2 c_{0_{12}}+jk_{0_{12}}]Y_1+\frac{\varepsilon}{4\omega_1}e^{-j2\rho t}[\omega_3 c_{c2_{13}}+k_{s2_{13}}+j(\omega_3 c_{s2_{13}}-k_{c2_{13}})]Z_2 \tag{50a}$$

$$\overset{*}{Y}_1=-\frac{\varepsilon}{2\omega_2}[\omega_1 c_{0_{21}}+jk_{0_{21}}]X_1-\frac{\varepsilon}{2}\left[c_{0_{22}}+j\frac{k_{0_{22}}}{\omega_2}\right]Y_1+\frac{\varepsilon}{4\omega_2}e^{-j2\rho t}[\omega_3 c_{c2_{23}}+k_{s2_{23}}+j(\omega_3 c_{s2_{23}}-k_{c2_{23}})]Z_2 \tag{50b}$$

$$\overset{*}{Z}_2=\frac{\varepsilon}{4\omega_3}e^{j2\rho t}[\omega_1 c_{c2_{31}}+k_{s2_{31}}-j(\omega_1 c_{s2_{31}}-k_{c2_{31}})]X_1+\frac{\varepsilon}{4\omega_3}e^{j2\rho t}[\omega_2 c_{c2_{32}}+k_{s2_{32}}-j(\omega_2 c_{s2_{32}}-k_{c2_{32}})]Y_1+\frac{\varepsilon}{2}\left[-c_{0_{33}}+j\frac{k_{0_{33}}}{\omega_3}\right]Z_2. \tag{50c}$$

Note, a similar set of three equations involving X_2, Y_2, and Z_1 is also obtained, but the stability properties are identical so only one set is shown, see Eq. (50). Finally, due to the special structure of the above equations, the time dependence can be eliminated by assuming a solution with the following form:

$$X_1=X_{1_0}e^{(r-j\rho)t} \quad Y_1=Y_{1_0}e^{(r-j\rho)t} \quad Z_2=Z_{2_0}e^{(r+j\rho)t}. \tag{51}$$

Here, r is an unknown complex number called the characteristic exponent. Since $j\rho$ is purely imaginary, the sign of the real part of r completely determines the stability of the solution. The stability condition is summarized in the following equation:

$$\text{stable} \quad \text{if } \mathrm{Re}[r]<0$$
$$\text{unstable} \quad \text{if } \mathrm{Re}[r]\geqslant 0. \tag{52}$$

To solve for the characteristic exponents, r, substitute the assumed solution in Eq. (51) into Eq. (50) and then solve the resulting algebraic eigenvalue problem. The characteristic exponents, r, are just the eigenvalues of the 3×3 complex characteristic matrix $\mathbf{\Gamma}$, shown below in terms of physical parameters.

$$\mathbf{\Gamma}=\begin{bmatrix} -\frac{1}{2}(c_d+2\zeta) & -\frac{\eta}{4}(1+f_\phi+2\rho) & -\frac{\eta\delta_3}{8}f_\phi^2 \\ \frac{\eta}{4}(1+f_\phi+2\rho) & -\frac{1}{2}(c_d+2\zeta) & \frac{\eta\delta_2}{8}f_\phi^2 \\ -\frac{\delta_3}{8f_\phi}(f_\phi+2\rho)^2 & \frac{\delta_2}{8f_\phi}(f_\phi+2\rho)^2 & -f_\phi\zeta \end{bmatrix} + j\begin{bmatrix} \rho & -\frac{\tau}{2} & -\frac{\eta\delta_2}{8}f_\phi^2 \\ \frac{\tau}{2} & \rho & -\frac{\eta\delta_3}{8}f_\phi^2 \\ \frac{\delta_2}{8f_\phi}(f_\phi+2\rho)^2 & \frac{\delta_3}{8f_\phi}(f_\phi+2\rho)^2 & -\rho \end{bmatrix} \tag{53}$$

Therefore, the determination of the stability about the worst-case shaft operating speed, $f=f_4$, with $\lambda=0$ is reduced to solving for the eigenvalues of the 3×3 the matrix $\mathbf{\Gamma}$. This procedure is much more computationally efficient than generating the Floquet transition matrix, $\Phi(T)$, via numerical integration and then evaluating its eigenvalues. A comparison between the numerical and perturbation method is shown in Fig. 11.

The difference between the two methods increases with misalignment angle. Since the time-varying portion of the system matrices in Eq. (20) are proportional to the misalignment angles, δ_2 and δ_3, as the misalignment increases the $O(\varepsilon)$ assumptions made in Eqs. (37) and (42) become less appropriate. Despite this difference between the two methods, the perturbation method is still a useful design tool since it always over predicts the severity of the instability, and hence offers a factor of safety.

7 Summary and Conclusions

This research investigates the interaction between lateral and torsional dynamics and explores the bounds of the torsion-lateral instability regions of a torsionally flexible, misaligned shaft-disk system driven through a U-joint. Nondimensional equations-of-motion are derived and the effects of angular misalignment, static load-torque, load-inertia, lateral frequency-split, and auxiliary damping on the stability of the system over a range of shaft operating speeds is investigated. By including both torsional and lateral degrees-of-freedom, it is discovered that, when misalignment is present, the lateral and torsional degrees-of-freedom are dynamically coupled by periodic terms in the inertia matrix. Furthermore these periodic inertia-coupling terms are proportional to the misalignment angle and the torsional inertia load. Physically, misalignment causes speed variation of the driven shaft, which in turn excites elastic twist dynamics of the driven shaft. If the driven shaft is carrying a significant torsional inertia load, the speed variation and elastic twist dynamics induce a dynamic torque load at the U-joint. The dynamic torque load, in addition to any static load torque, generate lateral moments at the U-joint which excite lateral shaft motion resulting in dynamic misalignment. Both the dynamic and the original static misalignment in return cause shaft speed variation. Hence, a dynamic coupling mechanism exists between the torsional and lateral degrees-of-freedom. This torsional-lateral coupling results in periodic inertia coupling matrices, $\mathbf{M}_0$, $\mathbf{M}_{s2}$, and $\mathbf{M}_{c2}$, which vary with twice the driving shaft speed and are proportional to the angular misalignment and the inertia load.

The analysis shows that the torsion-lateral inertia coupling matrices, $\mathbf{M}_0$, $\mathbf{M}_{s2}$, and $\mathbf{M}_{c2}$, induced by misalignment, significantly affect stability when the shaft speed, f, is near the torsion-lateral combination frequencies f_1, f_2, f_3, and f_4. When the shaft speed is near the difference-type torsion-lateral combinations, f_1 and f_2, misalignment has a stabilizing effect. On the other hand, when f is near the sum-type torsion-lateral combination frequencies, f_3 and f_4, misalignment can destabilize the sys-

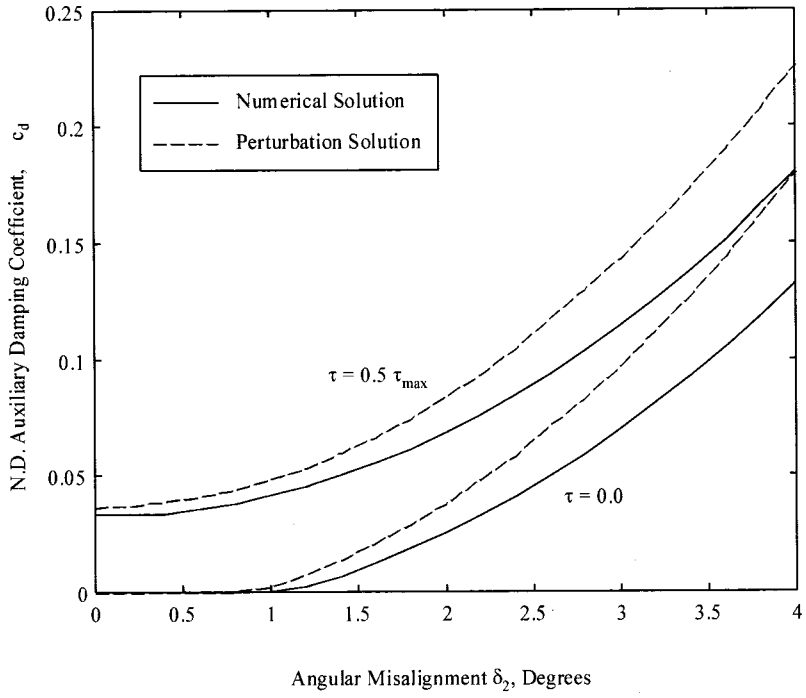


Fig. 11 Numerical and perturbation solution. $f=f_4$, $\zeta=0.005$, $\lambda=0.2$, $\Delta=3$, $\varepsilon_s=0.05$, $I_d=0.5$, $\mu=0.1$, $\gamma=10$.

tem. Furthermore, for sufficient misalignment, parametric instability zones of finite frequency width occur near both sum combination frequencies, f_3 and f_4, for all values of load-torque. Therefore, if shaft alignment cannot be guaranteed, it is important to avoid the sum-type torsion-lateral combination zones during shaft operation.

Since avoiding the sum-type torsion-lateral combination zones may be difficult for certain applications, it is shown, in this study, that such a restriction can be relaxed and stability can be guaranteed if the inherent structural damping is augmented with an auxiliary lateral damper with sufficient damping. To determine the amount of damping necessary for stability, the minimum, nondimensional damping coefficient required for stability, c_{req}, is computed over the shaft speed range for several values of misalignment and load-torque. For $\tau>0$, and nonzero misalignment, the required damping reaches a maximum, $c_{\text{max-req}}$, when $f=f_4$. Thus, the worst-case operating speed with respect to stability is $f=f_4$. It is demonstrated that $c_{\text{max-req}}$, increases with load-torque, τ and with misalignment, δ_2 and δ_3. Also, by increasing the lateral frequency-split parameter, λ, one can reduce the required damping c_{req} across the entire shaft speed range.

Finally, an efficient method for determining the stability at the least stable operating condition, $f=f_4$ and $\lambda=0$, is developed. Following Hsu's work [10,11], the method is based on an asymptotic solution of transformed averaged variational equations. Since the variational equations are obtained from a perturbation expansion about some nominal shaft speed, the stability analysis is only valid in the vicinity of this nominal shaft speed. After several transformations, the time-dependence of the variational equations is eliminated and the stability determination reduces to a complex eigenvalue problem in terms of the system parameters. The order of eigenvalue problem depends on the number of system modes involved in the instability.

With $\lambda=0$, i.e., no lateral frequency-split, the lateral modes have a multiplicity of two hence both lateral modes and the torsion mode are involved in the torsion-bending stability calculation. Since three modes are involved, the stability calculation is reduced to a 3×3 eigenvalue problem.

The perturbation approach stability calculation is much faster than the numerical Floquet-based calculation since no time integration is involved, however, there is some difference between the results especially at high misalignment angles (3 deg or 4 deg). This difference could be due to the inverse approximation or the assumption that the time-varying terms are all order ε in the perturbation expansion. Despite this, the method appears to always over predict the severity of the instability, thus it is still a useful design tool for estimating the necessary lateral damping required to suppress instability.

Acknowledgments

This research is sponsored by the U.S. Army Research Office MURI program and by the NASA Graduate Student Research Program with NASA Glenn Research Center.

Appendix

Using the inverse approximation, the perturbation matrices on the right-hand side of Eq. (39) are shown in Eqs. (*A*1) and (*A*2).

$$\mathbf{c}_0=\begin{bmatrix} 2\zeta f_v+c_d & f\eta & 0 \\ -f\eta & 2\zeta f_w+c_d & 0 \\ 0 & 0 & 2\zeta f_\phi \end{bmatrix} \tag{A1}$$

$$\mathbf{c}_{c2}=\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2f\delta_3 & -2f\delta_2 & 0 \end{bmatrix}, \quad \mathbf{c}_{s2}=\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2f\delta_2 & 2f\delta_3 & 0 \end{bmatrix}$$

$$\mathbf{k}_0=\begin{bmatrix} 0 & \tau & -\frac{1}{2}\delta_2\eta f_\phi^2 \\ -\tau & 0 & \frac{1}{2}\delta_3\eta f_\phi^2 \\ -\frac{1}{2}\delta_2 f_v^2 & \frac{1}{2}\delta_3 f_w^2 & 0 \end{bmatrix}$$

$$\mathbf{k}_{c2}=\begin{bmatrix} 0 & -\tau & \frac{1}{2}\delta_2\eta f_\phi^2 \\ -\tau & 0 & \frac{1}{2}\delta_3\eta f_\phi^2 \\ \delta_2\left(2f^2+\frac{1}{2}f_v^2\right) & \delta_3\left(2f^2+\frac{1}{2}f_w^2\right) & 0 \end{bmatrix} \tag{A2}$$

$$\mathbf{k}_{s2}=\begin{bmatrix} \tau & 0 & -\frac{1}{2}\delta_3\eta f_\phi^2 \\ 0 & -\tau & \frac{1}{2}\delta_2\eta f_\phi^2 \\ -\delta_3\left(2f^2+\frac{1}{2}f_v^2\right) & \delta_2\left(2f^2+\frac{1}{2}f_w^2\right) & 0 \end{bmatrix}$$

References

[1] Iwatsubo, T., and Saigo, M., 1984, ''Transverse Vibration of a Rotor System Driven by a Cardan Joint,'' J. Sound Vib., **95**, pp. 9–18.

[2] Mazzei, Jr., A. J., Argento, A., and Scott, R. A., 1999, ''Dynamic Stability of a Rotating Shaft Driven Through a Universal Joint,'' J. Sound Vib., **222**, pp. 19–47.

[3] Rosenberg, R. M., 1958, ''On the Dynamical Behavior of Rotating Shafts Driven by Universal (Hooke) Coupling,'' ASME J. Appl. Mech., **25**, pp. 47–51.

[4] Xu, M., and Marangoni, R. D., 1994, ''Vibration Analysis of a Motor-Flexible Coupling-Rotor System Subjected to Misalignment and Unbalance. Part I: Theoretical Model and Analysis,'' J. Sound Vib., **176**, pp. 663–679.

[5] Xu, M., and Marangoni, R. D., 1994, ''Vibration Analysis of a Motor-Flexible Coupling-Rotor System Subjected to Misalignment and Unbalance. Part II: Experimental Validation,'' J. Sound Vib., **176**, pp. 681–691.

[6] Kato, M., and Ota, H., 1990, ''Lateral Excitation of a Rotating Shaft Driven by a Universal Joint With Friction,'' ASME J. Vibr. Acoust., **112**, pp. 298–303.

[7] Asokanthan, S. F., and Hwang, M. C., 1996, ''Torsional Instabilities in a System Incorporating a Hooke's Joint,'' ASME J. Vibr. Acoust., **118**, pp. 368–374.

[8] Asokanthan, S. F., and Wang, X. H., 1996, ''Characterization of Torsional Instabilities in a Hooke's Joint Driven System via Maximal Lyapunov Exponents,'' J. Sound Vib., **194**, pp. 83–91.

[9] Bolotin, V. V., 1963, *Nonconservative Problems of the Theory of Elastic Stability*, Pergamon Press, New York.

[10] Hsu, C. S., 1963, ''On the Parametric Excitation of a Dynamic System Having Multiple Degrees of Freedom,'' ASME J. Appl. Mech., **30**, pp. 367–372.

[11] Hsu, C. S., 1965, ''Further Results on Parametric Excitation of a Dynamic System,'' ASME J. Appl. Mech., **32**, pp. 373–377.

W. Q. Zhu
Professor

M. L. Deng
Graduate Student

Z. L. Huang
Associate Professor

Department of Mechanics,
Zhejiang University,
Hangzhou 310027, P. R. China and
State Key Laboratory of Nonlinear Mechanics,
Institute of Mechanics,
Chinese Academy of Science,
Beijing 100008, P. R. China

First-Passage Failure of Quasi-Integrable Hamiltonian Systems

The first-passage failure of quasi-integrable Hamiltonian systems (multidegree-of-freedom integrable Hamiltonian systems subject to light dampings and weakly random excitations) is investigated. The motion equations of such a system are first reduced to a set of averaged Itô stochastic differential equations by using the stochastic averaging method for quasi-integrable Hamitonian systems. Then, a backward Kolmogorov equation governing the conditional reliability function and a set of generalized Pontryagin equations governing the conditional moments of first-passage time are established. Finally, the conditional reliability function, and the conditional probability density and moments of first-passage time are obtained by solving these equations with suitable initial and boundary conditions. Two examples are given to illustrate the proposed procedure and the results from digital simulation are obtained to verify the effectiveness of the procedure.
[DOI: 10.1115/1.1460912]

Introduction

In the theory of random vibration or stochastic structural dynamics, usually two failure models are studied: first-passage (first-excursion) failure and fatigue failure. In recent years, fatigue failure is treated as the propagation of a dominant crack to a critical size. Thus, fatigue failure becomes a special kind of first-passage failure. The first-passage failure is among the most difficult problems in the theory of random vibration or stochastic structural dynamics. At present, a mathematical exact solution is possible only if the random phenomenon in question can be treated as a diffusive Markov process. Still, known solutions are limited to the one-dimensional case ([1,2]).

The state space of a mechanical or structural system model is generally two-dimensional or higher. For such a system subject to Gaussian white noise excitation, the response is a vector diffusive Markov process, and a backward Kolmogorov equation governing the conditional reliability function and a set of generalized Pontryagin equations governing the conditional moments of first-passage time can be set up. However, these equations can usually be solved only numerically. For this purpose, a variety of numerical methods, such as finite element procedure and generalized cell mapping approach have been developed ([3–6]). Unfortunately, at present, the problems can be solved in this way are limited to two or three dimensional.

The response quantities of a quasi-Hamiltonian system (a linear or nonlinear conservative system subject to light dampings and weakly random excitations) can be divided into two categories: rapidly varying processes and slowly varying processes. Usually the slowly varying processes are much more significant for characterizing the long-term behavior of the system. Stochastic averaging is a method to derive the equations governing the slowly varying processes from the original equations of the system. The vector of slowly varying processes after averaging are (approximately) diffusive Markov process and the dimension of the averaged equations is usually much less than that of the original equations. Furthermore, the averaged equations are much more regular than the original equations since there is only one time scale in the former equations while there are two time scales in the later equations. Thus, the stochastic averaging method is a powerful approximate procedure to deal with quasi-Hamiltonian systems.

The first-passage failure of mechanical and structural system usually occurs rarely. It is a long-term behavior and the stochastic averaging method is suitable for studying it. The classical stochastic averaging method has been applied by many researchers to study the first-passage problem of single-degree-of-freedom oscillators with linear restoring force and with nonlinear restoring force ([7–17]). Recently, the stochastic averaging method for quasi-Hamiltonian systems has been developed ([18–20]). Except for response prediction, it has been applied to study the stochastic stability and bifurcation ([20–23]), the first-passage failure of quasi-non-integrable Hamiltonian systems ([24]) and the nonlinear stochastic optimal control ([25–29]).

In the present paper, the stochastic averaging method for quasi-integrable Hamiltonian systems is first reviewed briefly. Then the backward Kolmogorov equation governing the conditional reliability function and the generalized Pontryagin equations governing the conditional moments of first-passage time are derived from the averaged equations of quasi-integrable Hamiltonian systems, and the initial and boundary conditions are formulated. Finally, two examples are worked out and the results obtained by using the proposed procedure are compared with those from digital simulation and with those obtained by using the procedure for quasi-non-integrable Hamiltonian systems ([24]).

Stochastic Averaging of Quasi-Integrable Hamiltonian Systems

The stochastic averaging method for quasi-integrable Hamiltonian systems has been developed for nonresonant and resonant cases, and for white noise and wide-band excitations ([19,23]). Here, only the method for nonresonant case and for white noise excitation is briefly reviewed. Consider a quasi-Hamiltonian system of n-degree-of-freedom governed by the following equations of motion:

$$\dot{Q}_i = \frac{\partial \overline{H}}{\partial P_i}$$

$$\dot{P}_i = -\frac{\partial \overline{H}}{\partial Q_i} - \varepsilon c_{ij}\frac{\partial \overline{H}}{\partial P_j} + \varepsilon^{1/2} f_{ik} W_k(t) \tag{1}$$

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, Feb. 28, 2001; final revision, Sept. 27, 2001. Associate Editor: N. C. Perkins. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

$$i,j=1,2,\ldots,n; \quad k=1,2,\ldots,m$$

where Q_i and P_i are generalized displacements and momenta, respectively; $H=H(\mathbf{Q},\mathbf{P})$ is twice differentiable Hamiltonian; $c_{ij}=c_{ij}(\mathbf{Q},\mathbf{P})$ are functions representing quasi-linear damping coefficients; $f_{ik}=f_{ik}(\mathbf{Q},\mathbf{P})$ are functions representing excitation amplitudes; ε is a small positive parameter; $W_k(t)$ are Gaussian white noises in the sense of Stratonovich with correlation functions $E[W_k(t)W_l(t+\tau)]=2D_{kl}\delta(\tau)$.

Equation (1) can be modeled as the following set of Itô stochastic differential equations:

$$dQ_i=\frac{\partial \overline{H}}{\partial P_i}dt \tag{2a}$$

$$dP_i=-\left(\frac{\partial \overline{H}}{\partial Q_i}+\varepsilon c_{ij}\frac{\partial \overline{H}}{\partial P_j}-\varepsilon D_{kl}f_{jl}\frac{\partial f_{ik}}{\partial P_j}\right)dt+\varepsilon^{1/2}\sigma_{ik}dB_k(t)$$
$$i,j=1,2,\ldots,n; \quad k=1,2,\ldots,m \tag{2b}$$

where $B_k(t)$ are the independent unit Wiener processes and $\boldsymbol{\sigma}\boldsymbol{\sigma}^T=2\mathbf{D}\mathbf{f}\mathbf{f}^T$. The double summation terms on the right-hand side of Eq. (2*b*) are known as the Wong-Zakai correction terms. These terms usually can be split into two parts: one having the effect of modifying the conservative forces and another modifying the damping forces. The first part can be combined with $-\partial H/\partial Q_i$ to form an overall effective conservative forces $-\partial H/\partial Q_i$ with a modified Hamiltonian $H=H(\mathbf{Q},\mathbf{P})$ and with $\partial H/\partial P_i=\partial \overline{H}/\partial P_i$. The second part can be combined with $-\varepsilon c_{ij}\partial \overline{H}/\partial P_j$ to constitute an effective damping forces $-\varepsilon m_{ij}\partial H/\partial P_j$ with $m_{ij}=m_{ij}(\mathbf{Q},\mathbf{P})$. With these accomplished, Eqs. (2*a*) and (2*b*) can be rewritten as

$$dQ_i=\frac{\partial H}{\partial P_i}dt \tag{3a}$$

$$dP_i=-\left(\frac{\partial H}{\partial Q_i}+\varepsilon m_{ij}\frac{\partial H}{\partial P_j}\right)dt+\varepsilon^{1/2}\sigma_{ik}dB_k(t)$$
$$i,j=1,2,\ldots,n; \quad k=1,2,\ldots,m. \tag{3b}$$

Assume that the Hamiltonian system with Hamiltonian H is integrable and nonresonant. That is, in the Hamiltonian system there exist n independent first integrals (conserved quantities) $H_1,H_2,\ldots,H_n$, which are in involution. The words "in involution" implies that the Poisson bracket of any two of $H_1,H_2,\ldots,H_n$ vanishes. In principle, n pairs of action-angle variables I_i,θ_i can be introduced for an integrable Hamiltonian system of n-degrees-of-freedom. Non-resonance means that the n frequencies, $\omega_i=d\theta_i/dt$, do not satisfy the following resonant relation:

$$k_i^u\omega_i=0(\epsilon) \tag{4}$$

where k_i^u are integers with $\Sigma_{i=1}^n|k_i^u|<4$.

Introduce transformations

$$H_r=H_r(\mathbf{Q},\mathbf{P}), \quad r=1,2,\ldots,n. \tag{5}$$

The Itô stochastic differential equations for H_r are obtained from Eqs. (3*a*) and (3*b*) by using Itô differential rule as follows:

$$dH_r=\varepsilon\left(-m_{ij}\frac{\partial H}{\partial P_j}\frac{\partial H_r}{\partial P_i}+\frac{1}{2}\sigma_{ik}\sigma_{jk}\frac{\partial^2 H_r}{\partial P_i\partial P_j}\right)dt$$
$$+\varepsilon^{1/2}\frac{\partial H_r}{\partial P_i}\sigma_{ik}dB_k(t)$$
$$r,i,j=1,2,\ldots,n; \quad k=1,2,\ldots,m \tag{6}$$

where P_i are replaced by H_s in terms of Eq. (5). Now the system is governed by Eqs. (3*a*) and (6) and the state variables are Q_i and H_r. It is seen from these equations that Q_i are rapidly varying processes while H_r are slowly varying processes. According to the Khasminskii theorem ([30]), H_r converge weakly to an n-dimensional vector diffusion processes as $\varepsilon\to 0$ in a time interval $0\leqslant t\leqslant T$, where $T\sim 0(\varepsilon^{-1})$. For simplicity, the same symbols H_r are used to denote r components of this diffusion process.

The Itô stochastic differential equations for this n-dimensional vector diffusion process can be obtained by applying time averaging to Eq. (6). The result is

$$dH_r=a_r(\mathbf{H})dt+\overline{\sigma}_{rk}(\mathbf{H})d\overline{B}_k(t)$$
$$r=1,2,\ldots,n; \quad k=1,2,\ldots,m \tag{7}$$

where $\mathbf{H}=[H_1H_2\ldots H_n]^T$; $\overline{B}_k(t)$ are independent unit Wiener processes;

$$a_r(\mathbf{H})=\varepsilon\left\langle -m_{ij}\frac{\partial H}{\partial P_j}\frac{\partial H_r}{\partial P_i}+\frac{1}{2}\sigma_{ik}\sigma_{jk}\frac{\partial^2 H_r}{\partial P_i\partial P_j}\right\rangle_t$$

$$b_{rs}(\mathbf{H})=\overline{\sigma}_{rk}(\mathbf{H})\overline{\sigma}_{sk}(\mathbf{H})=\varepsilon\left\langle \sigma_{ik}\sigma_{jk}\frac{\partial H_r}{\partial P_i}\frac{\partial H_s}{\partial P_j}\right\rangle_t \tag{8}$$

$$\langle[\ ]\rangle_t=\lim_{T\to\infty}\frac{1}{T}\int_T^{t_0+T}[\ ]dt.$$

Note that H_r are kept constant in performing the time averaging.

The time averaging in Eq. (8) may be replaced by space averaging. For example, suppose that the Hamiltonian is separable and equal to the sum of n independent first integers, i.e.,

$$H(\mathbf{q},\mathbf{p})=\sum_{r=1}^n H_r(q_r,p_r) \tag{9}$$

and for each H_r there is a periodic orbit with period T_r. Then the averaged drift and diffusion coefficients of Eq. (7) can be obtained as follows:

$$a_r(\mathbf{H})=\frac{\varepsilon}{T}\oint\left(-m_{rj}\frac{\partial H_j}{\partial P_j}\frac{\partial H_r}{\partial P_r}+\frac{1}{2}\sigma_{rk}\sigma_{rk}\frac{\partial^2 H_r}{\partial P_r^2}\right)$$
$$\times\prod_{\mu=1}^n\left(1\Big/\frac{\partial H_\mu}{\partial P_\mu}\right)dq_\mu$$
$$b_{rs}(\mathbf{H})=\frac{\varepsilon}{T}\oint\left(\sigma_{rk}\sigma_{sk}\frac{\partial H_r}{\partial P_r}\frac{\partial H_s}{\partial P_s}\right)\times\prod_{\mu=1}^n\left(1\Big/\frac{\partial H_\mu}{\partial P_\mu}\right)dq_\mu \tag{10}$$

where $\oint[\ ]\Pi_{\mu=1}^n(\ )dq_\mu$ represents an n-fold loop integral and

$$T=T(\mathbf{H})=\prod_{\mu=1}^n T_\mu=\oint\prod_{\mu=1}^n\left(1\Big/\frac{\partial H_\mu}{\partial P_\mu}\right)dq_\mu. \tag{11}$$

In the case where action-angle variables I_i, θ_i are available, H_r can be replaced by I_r and averaged Itô Eq. (7) by

$$dI_r=\overline{a}_r(\mathbf{I})dt+\bar{\bar{\sigma}}_{rk}(\mathbf{I})d\bar{\bar{B}}_k(t)$$
$$r=1,2,\ldots,n; \quad k=1,\ldots,m \tag{12}$$

where $\mathbf{I}=[I_1I_2\ldots I_n]^T$;

$$\overline{a}_r(\mathbf{I})=\frac{\varepsilon}{(2\pi)^n}\int_0^{2\pi}\left(-m_{ij}\frac{\partial H}{\partial P_j}\frac{\partial I_r}{\partial P_i}+\frac{1}{2}\sigma_{ik}\sigma_{jk}\frac{\partial^2 I_r}{\partial P_i\partial P_j}\right)d\boldsymbol{\theta}$$
$$b_{rs}(\mathbf{I})=\bar{\bar{\sigma}}_{rk}(\mathbf{I})\bar{\bar{\sigma}}_{sk}(\mathbf{I})=\frac{\varepsilon}{(2\pi)^n}\int_0^{2\pi}\left(\sigma_{ik}\sigma_{jk}\frac{\partial I_r}{\partial P_i}\frac{\partial I_s}{\partial P_j}\right)d\boldsymbol{\theta} \tag{13}$$

in which $\boldsymbol{\theta}=[\theta_1\theta_2\ldots\theta_n]^T$; $\int_0^{2\pi}[\ ]d\boldsymbol{\theta}$ denotes an n-fold integral.

Note that averaged Eq. (7) or (12) is much simpler than original Eq. (1). The dimension of the former equation is only a half of that of the later equation. Equations (7) and (12) contain only

slowly varying process $\mathbf{H}(t)$ and $\mathbf{I}(t)$, respectively, and they are suitable for studying the long-term behavior of the system, such as the first-passage failure.

Backward Kolmogorov Equation and Generalized Pontryagin Equations

For most mechanical/structural systems Hamiltonian H represents the total energy of the system, and H_r the energy of the rth degree-of-freedom of the system. H_r may vary between H_{r0} and ∞, where H_{r0} is a constant, such as H for a Duffing oscillator with hardening spring, between $-\infty$ and H_{r0}, such as H for a Duffing oscillator with softening spring, or between H_{r0} and H_{rm}, where H_{rm} is a constant, such as H for a pendulum. The state of the averaged system of a quasi-integrable Hamiltonian system varies randomly in the n-dimensional domain defined by the direct product of the H_r intervals and the safety domain Ω is a bounded region with boundary Γ within the n-dimensional H_r domain. Suppose that the lower boundary of a safety domain for each H_r is at zero (it is always possible to make so by using coordinate transformation). Then the boundary Γ consists of Γ_0 (at least one of H_r vanishes) and critical boundary Γ_c. The first-passage failure occurs when $\mathbf{H}(t)$ crosses Γ_c for the first time, and it is characterized by the conditional reliability function, the conditional probability density or conditional moments of first-passage time, where the word ''conditional'' means under the given initial condition in the safety domain.

The conditional reliability function, denoted by $R(t|\mathbf{H}_0)$, is defined as the probability of $\mathbf{H}(t)$ being in safety domain Ω within time interval $(0,t]$ given initial state $\mathbf{H}_0=\mathbf{H}(0)$ being in Ω, i.e.,

$$R(t|\mathbf{H}_0)=P\{\mathbf{H}(\tau)\in\Omega,\tau\in(0,t]|\mathbf{H}_0\in\Omega\}. \tag{14}$$

It is the integral of the conditional transition probability density in Ω. The conditional transition probability density is the transition probability density of the sample functions which remain in Ω in time interval $[0,t]$. For an averaged system, the conditional transition probability density satisfies the backward Kolmogorov equation with drift and diffusion coefficients defined by Eqs. (8), (10), or (13). Thus, the following backward Kolmogorov equation can be derived for the conditional reliability function:

$$\frac{\partial R}{\partial t}=a_r(\mathbf{H}_0)\frac{\partial R}{\partial H_{r0}}+\frac{1}{2}b_{rs}(\mathbf{H}_0)\frac{\partial^2 R}{\partial H_{r0}\partial H_{s0}} \qquad r,s=1,2,\ldots,n \tag{15}$$

where $a_r(\mathbf{H}_0)$ and $b_{rs}(\mathbf{H}_0)$ are defined by Eqs. (8) or (10) with $\mathbf{H}$ replaced by $\mathbf{H}_0$. The initial condition is

$$R(0|\mathbf{H}_0)=1, \quad \mathbf{H}_0\in\Omega \tag{16}$$

which implies that the system is initially in the safety domain. The boundary conditions are

$$R(t|\Gamma_0)=finite \tag{17}$$

$$R(t|\Gamma_c)=0. \tag{18}$$

Equations (17) and (18) imply that Γ_0 is a reflecting boundary while Γ_c is the absorbing boundary.

The first-passage time T is defined as the time when the system reaches critical boundary Γ_c for the first time given $\mathbf{H}_0$ being in Ω. Noting that the conditional probability of the first-passage failure $F(t|\mathbf{H}_0)=1-R(t|\mathbf{H}_0)$, the conditional probability density of the first-passage time can be obtained from the conditional reliability function as follows:

$$p(T|\mathbf{H}_0)=\left.\frac{-\partial R(t|\mathbf{H}_0)}{\partial t}\right|_{t=T}. \tag{19}$$

The conditional moments of first-passage time are defined as

$$\mu_l(\mathbf{H}_0)=\int_0^\infty T^l p(T|\mathbf{H}_0)dT, \quad l=1,2,\ldots. \tag{20}$$

The equations governing the conditional moments of first-passage time can be obtained from Eq. (15) in terms of relationships (19) and (20) as follows:

$$\frac{1}{2}b_{rs}(\mathbf{H}_0)\frac{\partial^2\mu_{l+1}}{\partial H_{r0}\partial H_{s0}}+a_r(\mathbf{H}_0)\frac{\partial\mu_{l+1}}{\partial H_{r0}}=-(l+1)\mu_l \qquad r,s=1,2,\ldots,n; \quad l=0,1,2,\ldots. \tag{21}$$

It is easily seen from Eq. (20) that $\mu_0=1$. The boundary conditions associated with Eq. (21) are obtained from Eqs. (17) and (18) in terms of Eqs. (19) and (20). They are

$$\mu_l(\Gamma_0)=finite \tag{22}$$

$$\mu_l(\Gamma_c)=0. \tag{23}$$

Note that both boundary conditions (17) and (22) are qualitative rather than quantitative. They can be made to be quantitative by using Eqs. (15) and (21), respectively, based on the limiting behavior of the drift and diffusion coefficients in Eqs. (15) and (21) at boundary Γ_0 and it will be illustrated with the followingj examples.

The conditional reliability function is obtained from solving backward Kolmogorov Eq. (15) together with initial condition (16) and boundary conditions (17) and (18). The conditional probability density of first-passage time is obtained from the conditional reliability function by using Eq. (19). The conditional moments of first-passage time are obtained either from the conditional probability density of first-passage time by using definition (20) or directly from solving generalized Pontryagin Eq. (21) together with boundary conditions (22) and (23).

Examples

Example 1. Consider linearly and nonlinearly coupled two linear oscillators subject to external and parametric excitations of Gaussian white noises. The equations of motion of the system are of the form

$$\begin{aligned}\ddot{X}_1+\alpha_{11}\dot{X}_1+\alpha_{12}\dot{X}_2+\beta_1(X_1^2+X_2^2)\dot{X}_1+\omega_1^2X_1&=W_1(t)+X_1W_3(t)\\ \ddot{X}_2+\alpha_{21}\dot{X}_1+\alpha_{22}\dot{X}_2+\beta_2(X_1^2+X_2^2)\dot{X}_2+\omega_2^2X_2&=W_2(t)+X_2W_4(t)\end{aligned} \tag{24}$$

where α_{ij}, β_i, and $\omega_i(i,j=1,2)$ are constants; $W_k(t)(k=1,2,3,4)$ are independent Gaussian white noises with intensities $2D_k$; α_{ij}, β_i, and D_k are assumed of the same order of ε. The response of system (24) in both nonresonant and resonant cases with external excitations only has been studied by using the stochastic averaging method for quasi-integrable Hamiltonian systems ([19]). Here we study the first-passage failure of system (24) in a nonresonant case.

Let $X_1=Q_1$, $X_2=Q_2$, $\dot{X}_1=P_1$, $\dot{X}_2=P_2$. Equation (24) can be recast in the form of Eq. (1) as follows:

$$\begin{aligned}\dot{Q}_1&=P_1\\ \dot{Q}_2&=P_2\\ \dot{P}_1&=-\omega_1^2Q_1-[\alpha_{11}+\beta_1(Q_1^2+Q_2^2)]P_1-\alpha_{12}P_2+W_1(t)+Q_1W_3(t)\\ \dot{P}_2&=-\omega_2^2Q_2-[\alpha_{22}+\beta_2(Q_1^2+Q_2^2)]P_2-\alpha_{21}P_1+W_2(t)+Q_2W_4(t).\end{aligned} \tag{25}$$

Equation (25) can be modeled as Itô stochastic differential equations of the form of Eqs. (3*a*) and (3*b*). Since the Wong-Zakai

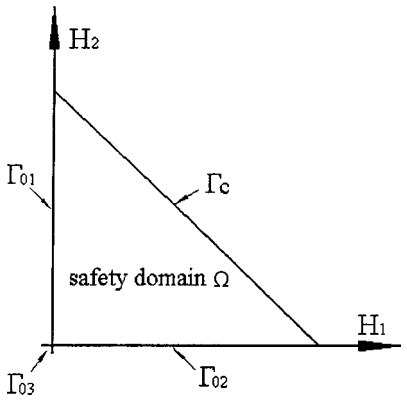


Fig. 1 Safety domain Ω and its boundary on plane H_1 and H_2 for system (24)

correction terms in this case vanish, the modified Hamiltonian associated with the Itô equations is the same as that associated with Eq. (25), i.e.,

$$H=H_1+H_2 \tag{26}$$

$$H_i=\frac{1}{2}(P_i^2+\omega_i^2Q_i^2), \quad i=1,2. \tag{27}$$

The Hamiltonian system with Hamiltonian H is integrable. Thus, system (25) is a quasi-integrable Hamiltonian system. By using the stochastic averaging method for quasi-integrable Hamiltonian systems, the following averaged Itô equations can be obtained in the nonresonant case:

$$\begin{gathered} dH_r=a_r(H_1,H_2)dt+\overline{\sigma}_{rk}(H_1,H_2)d\overline{B}_k(t) \\ r=1,2, \quad k=1,2,3,4 \end{gathered} \tag{28}$$

where

$$\begin{aligned} a_1&=-\alpha_{11}H_1-\frac{\beta_1}{2\omega_1^2}H_1^2-\frac{\beta_1}{\omega_2^2}H_1H_2+D_1+\frac{D_3}{\omega_1^2}H_1 \\ a_2&=-\alpha_{22}H_2-\frac{\beta_2}{2\omega_2^2}H_2^2-\frac{\beta_2}{\omega_1^2}H_1H_2+D_2+\frac{D_4}{\omega_2^2}H_2 \\ b_{11}&=\overline{\sigma}_{1k}\overline{\sigma}_{1k}=2D_1H_1+D_3\frac{H_1^2}{\omega_1^2} \\ b_{22}&=\overline{\sigma}_{2k}\overline{\sigma}_{2k}=2D_2H_2+D_4\frac{H_2^2}{\omega_2^2} \\ b_{12}&=b_{21}=\overline{\sigma}_{1k}\overline{\sigma}_{2k}=0. \end{aligned} \tag{29}$$

It is seen from Eq. (27) that H_i vary from 0 to ∞. So, the state of averaged system (28) varies randomly in the first quadrant of plane (H_1,H_2). Suppose that the limit state of the system is $H=H_1+H_2=H_c$, i.e.,

$$\Gamma_c: \ H_1+H_2=H_c, \quad H_1,H_2\geqslant 0. \tag{30}$$

The safety domain of the system is the inside of a right triangle with boundaries Γ_c in Eq. (30) and Γ_0 defined by

$$\begin{aligned} &\Gamma_0=\Gamma_{01}+\Gamma_{02}+\Gamma_{03}, \\ &\Gamma_{01}: \ H_1=0, \quad 0<H_2<H_c \\ &\Gamma_{02}: \ H_2=0, \quad 0<H_1<H_c \\ &\Gamma_{03}: \ H_1=H_2=0 \end{aligned} \tag{31}$$

(see Fig. 1).

Following Eq. (15), the conditional reliability function $R(t|H_{10},H_{20})$ of system (24) is governed by the following backward Kolmogorov equation:

$$\frac{\partial R}{\partial t}=a_1\frac{\partial R}{\partial H_{10}}+a_2\frac{\partial R}{\partial H_{20}}+\frac{1}{2}b_{11}\frac{\partial^2 R}{\partial H_{10}^2}+\frac{1}{2}b_{22}\frac{\partial^2 R}{\partial H_{20}^2} \tag{32}$$

where a_1, a_2, b_{11}, and b_{22} are defined by Eq. (29) with H_1, H_2 replaced by H_{10} and H_{20}, respectively. The initial condition is Eq. (16) with $\mathbf{H}_0=[H_{10}H_{20}]^T$. One boundary condition is Eq. (18) with Γ_c defined by Eq. (30). The other qualitative boundary condition, Eq. (17) with Γ_0 defined by Eq. (31), can be transformed into a quantitative one by using Eq. (32) and considering the limiting behavior of drift and diffusion coefficients in Eq. (29) at boundary Γ_0 defined by Eq. (31). It is

$$\begin{aligned} \frac{\partial R}{\partial t}&=D_1\frac{\partial R}{\partial H_{10}}+\left(D_2-\alpha_{22}H_{20}-\frac{\beta_2}{2\omega_2^2}H_{20}^2+\frac{D_4}{\omega_2^2}H_{20}\right)\frac{\partial R}{\partial H_{20}} \\ &\quad+\left(D_2H_{20}+D_4\frac{H_{20}^2}{2\omega_2^2}\right)\frac{\partial^2 R}{\partial H_{20}^2} \end{aligned} \tag{33}$$

for boundary Γ_{01};

$$\begin{aligned} \frac{\partial R}{\partial t}&=\left(D_1-\alpha_{11}H_{10}-\frac{\beta_2}{2\omega_1^2}H_{10}^2+\frac{D_3}{\omega_1^2}H_{10}\right)\frac{\partial R}{\partial H_{10}} \\ &\quad+D_2\frac{\partial R}{\partial H_{20}}+\left(D_1H_{10}+D_3\frac{H_{10}^2}{2\omega_1^2}\right)\frac{\partial^2 R}{\partial H_{10}^2} \end{aligned} \tag{34}$$

for boundary Γ_{02};

$$\frac{\partial R}{\partial t}=D_1\frac{\partial R}{\partial H_{10}}+D_2\frac{\partial R}{\partial H_{20}} \tag{35}$$

for boundary Γ_{03}.

Equation (32) is a two-dimensional parabolic partial differential equation and can be solved numerically together with the initial and boundary conditions by using the Peaceman-Rachford scheme of the finite difference method to yield the conditional reliability function of system (24). The conditional probability density of the first-passage time of system (24) is then obtained from the conditional reliability function by using Eq. (19).

Similarly, the generalized Pontryagin equations for the conditional moments of the first passage time of system (24) can be derived from the averaged Itô Eq. (28) as follows:

$$\begin{aligned} &\frac{1}{2}b_{11}\frac{\partial^2\mu_{l+1}}{\partial H_{10}^2}+\frac{1}{2}b_{22}\frac{\partial^2\mu_{l+1}}{\partial H_{20}^2}+a_1\frac{\partial\mu_{l+1}}{\partial H_{10}} \\ &\quad+a_2\frac{\partial\mu_{l+1}}{\partial H_{20}} \\ &\quad=-(l+1)\mu_l \end{aligned} \tag{36}$$

where a_1, a_2, b_1, and b_2 are defined by Eq. (29) with H_1 and H_2 replaced by H_{10} and H_{20}, respectively. One boundary condition is (23) with Γ_c defined by Eq. (30). The other qualitative boundary condition, Eq. (22) with Γ_0 defined by Eq. (31), can be transformed into quantitative one by using Eq. (36) and considering the limiting behavior of the drift and diffusion coefficients in Eq. (29) at boundary Γ_0. It is

$$\begin{aligned} &\left(D_2H_{20}+D_4\frac{H_{20}^2}{2\omega_2^2}\right)\frac{\partial^2\mu_{l+1}}{\partial H_{20}^2}+D_1\frac{\partial\mu_{l+1}}{\partial H_{10}}+\left(D_2-\alpha_{22}H_{20}\right. \\ &\quad\left.-\frac{\beta_2}{2\omega_2^2}H_{20}^2+\frac{D_4}{\omega_2^2}H_{20}\right)\frac{\partial\mu_{l+1}}{\partial H_{20}}=-(l+1)\mu_l \end{aligned} \tag{37}$$

for boundary Γ_{01};

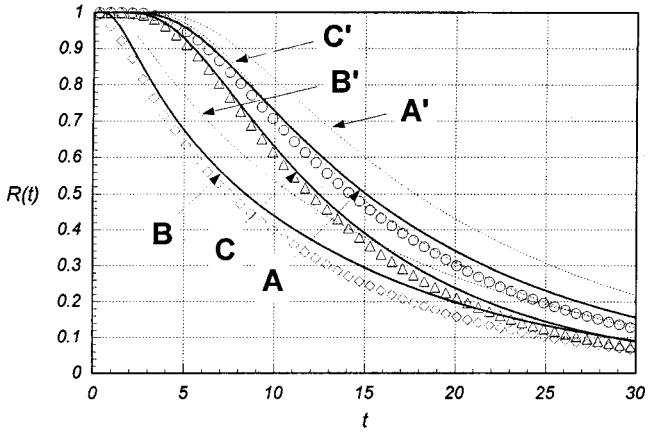


Fig. 2 Reliability function of system (24) for given initial condition. $\alpha_{11}=0.01$, $\alpha_{12}=0.03$, $\beta_1=0.1$, $\omega_1=1.0$, $\alpha_{21}=0.04$, $\alpha_{22}=0.04$, $\beta_2=0.4$, $\omega_2=0.707$, $2D_1=0.03$, $2D_2=0.01$, $H_c=0.3$. The other parameters are $2D_3=2D_4=0$, $H_{10}=H_{20}=0$ for A and A′; $2D_3=2D_4=0$, $H_{10}=0.09$, $H_{20}=0.03$ for B and B′; $2D_3=0.1$, $2D_4=0.01$, $H_{10}=H_{20}=0$ for C and C′. ——analytical result by using the present proposed procedure; – – – –analytical result by using the procedure proposed in [24]; ○ ◇ △ from digital simulation.

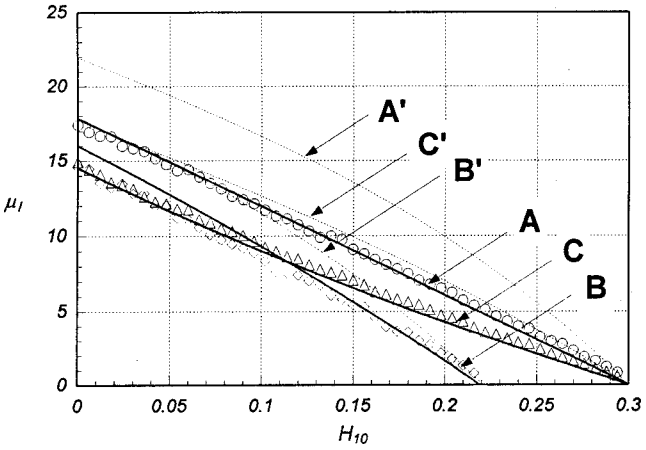


Fig. 4 Mean first-passage time of system (24) as function of H_{10} for given H_{20}. $2D_3=2D_4=0$, $H_{20}=0$ for A and A′; $2D_3=2D_4=0$, $H_{20}=0.08$ for B and B′; $2D_3=0.1$, $2D_4=0.01$, $H_{20}=0$ for C and C′. The other parameters and symbols are the same as those in Fig. 2.

$$\left(D_1H_{10}+D_3\frac{H_{10}^2}{2\omega_1^2}\right)\frac{\partial^2\mu_{l+1}}{\partial H_{10}^2}+\left(D_1-\alpha_{11}H_{10}-\frac{\beta_1}{2\omega_1^2}H_{10}^2+\frac{D_3}{\omega_1^2}H_{10}\right)\frac{\partial\mu_{l+1}}{\partial H_{10}}+D_2\frac{\partial\mu_{l+1}}{\partial H_{20}}=-(l+1)\mu_l \tag{38}$$

for boundary Γ_{02};

$$D_1\frac{\partial\mu_{l+1}}{\partial H_{10}}+D_2\frac{\partial\mu_{l+1}}{\partial H_{20}}=-(l+1)\mu_l. \tag{39}$$

Equation (36) is a two-dimensional elliptical partial differential equation and can be solved numerically together with boundary conditions by using the five-point scheme of the finite difference method to yield the conditional moments of first-passage time of system (24).

Some numerical results for the conditional reliability function, the conditional probability density and the conditional mean of the first passage time of system (24) obtained by using the above procedure are shown in Figs. 2–4. Similar results from digital simulation are also shown for comparison. It is seen that the two results are in excellent agreement. Note that the conditional reliability function is a monotonously decreasing function of time. Some results for the reliability function, the probability density, and the mean of first-passage time of system (24) as functions of the initial condition are shown in Figs. 5–7. It is seen that both the reliability and mean first-passage time are monotonously decreasing functions of H_{10} and/or H_{20}.

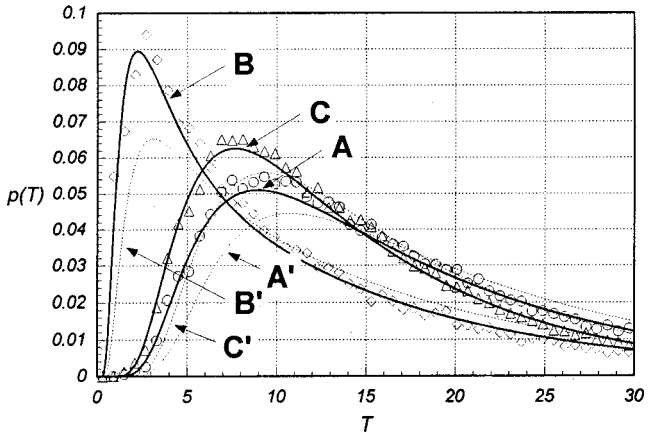


Fig. 3 Probability density of first-passage time of system (24) for given initial condition. The parameters and symbols are the same as those in Fig. 2.

As indicated above, system (24) is a quasi-integrable Hamiltonian system. However, the procedure for evaluating the conditional reliability function and the statistics of first-passage time for quasi-non-integrable Hamiltonian systems developed in [24] can also be applied to system (24). It is interesting to see if this method yields good results.

Treat system (24) as a quasi-non-integrable Hamiltonian system, the averaged Itô equation is of the form

$$dH=a(H)dt+\overline{\sigma}(H)d\overline{B}(t) \tag{40}$$

where H is defined by Eqs. (26) and (27),

$$\begin{aligned}a(H)&=D_1+D_2-\frac{1}{6}(\beta_1+\beta_2)\left(\frac{1}{\omega_1^2}+\frac{1}{\omega_2^2}\right)\\&\quad-\frac{1}{2}\left(\alpha_{11}+\alpha_{22}-\frac{D_3}{\omega_1^2}-\frac{D_4}{\omega_2^2}\right)H\\b(H)&=\overline{\sigma}^2(H)=\frac{1}{3}\left(\frac{D_3}{\omega_1^2}+\frac{D_4}{\omega_2^2}\right)H^2+(D_1+D_2)H.\end{aligned} \tag{41}$$

The conditional reliability function $R(t|H_0)$ of system (40) is governed by the following one-dimensional backward Kolmogorov equation:

$$\frac{\partial R}{\partial t}=a(H_0)\frac{\partial R}{\partial H_0}+\frac{1}{2}b(H_0)\frac{\partial^2R}{\partial H_0^2} \tag{42}$$

where $a(H_0)$ and $b(H_0)$ are defined by Eq. (41) with H replaced by H_0. The boundary conditions are

$$R(t|H_c)=0 \tag{43}$$

$$R(t|0)=finite. \tag{44}$$

The later condition is qualitative and can be made to be quantitative by using Eq. (42) and the limiting behavior of $a(H_0)$ and $b(H_0)$ near $H_0=0$. It is

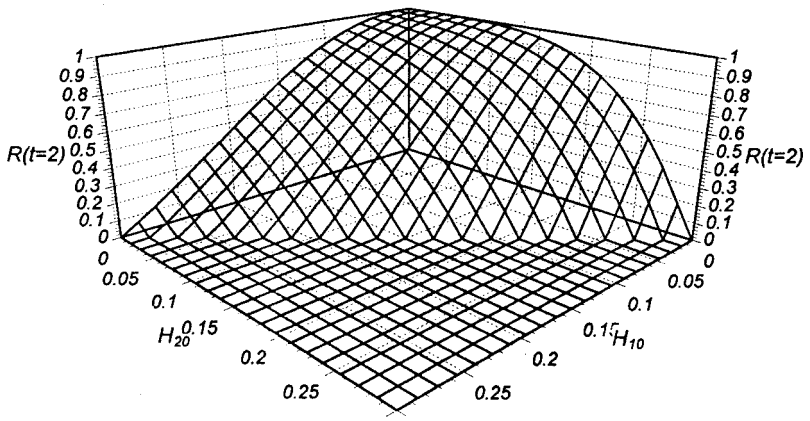


Fig. 5 Reliability of system (24) at $t=2$ (second) as function of H_{10} and H_{20}. $2D_3=0.1$, $2D_4=0.01$. The other parameters are the same as those in Fig. 2.

$$\frac{\partial R}{\partial t}=\left[D_1+D_2-\frac{1}{6}(\beta_1+\beta_2)\left(\frac{1}{\omega_1^2}+\frac{1}{\omega_2^2}\right)\right]\frac{\partial R}{\partial H_0}. \tag{45}$$

The initial condition is

$$R(0|H_0)=1. \tag{46}$$

The one-dimensional boundary-initial value problem, Eqs. (42), (43), (45), and (46), can be solved by using the finite difference method of Crank-Nicolson type. The conditional probability density of first-passage time can be obtained from $R(t|H_0)$ as follows:

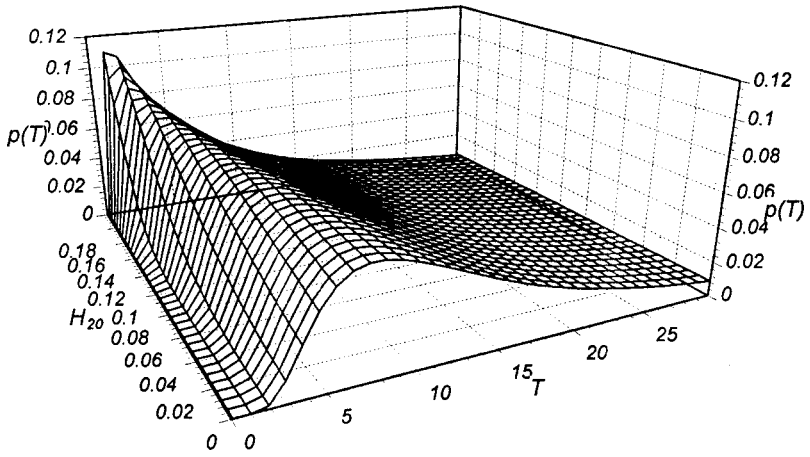


Fig. 6 Probability density of first-passage time of system (24) as function of H_{20} and t for given $H_{10}=0$. $2D_3=0.1$, $2D_4=0.01$. The other parameters are the same as those in Fig. 2.

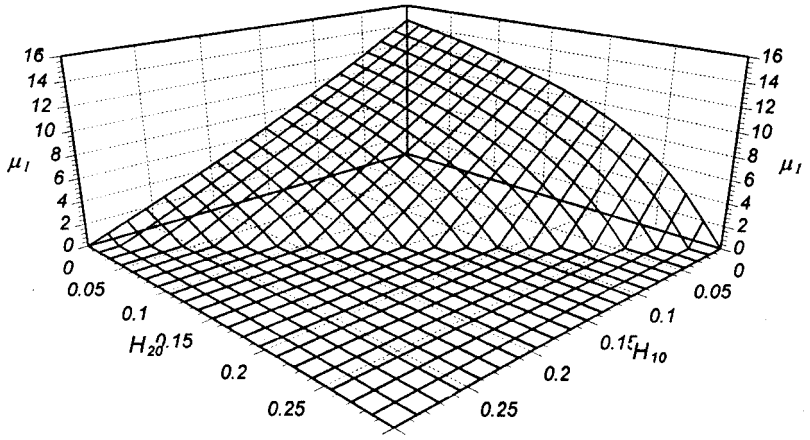


Fig. 7 Mean first-passage time of system (24) as function of H_{10} and H_{20}. $2D_3=0.1$, $2D_4=0.01$. The other parameters are the same as those in Fig. 2.

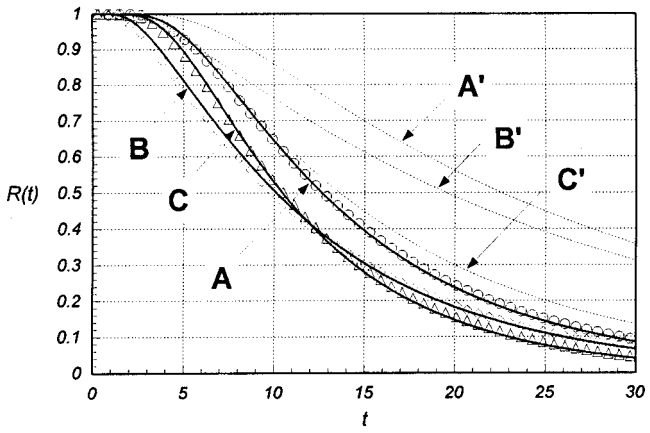


Fig. 8 Reliability function of system (52) for given initial condition. $\alpha_1=0.2$, $\alpha_2=0.1$, $\alpha_3=0.1$, $\beta_1=0.05$, $\omega=1.0$; $\alpha_4=0.4$, $\beta_2=0.1$, $k=2.0$, $2D_1=0.03$, $2D_2=0.01$, $H_c=0.3$. The other parameters are $2D_3=2D_4=0$, $H_{10}=H_{20}=0$ for A and A′; $2D_3=2D_4=0$, $H_{10}=0.04$, $H_{20}=0.02$ for B and B′; $2D_3=0.1$, $2D_4=0.05$, $H_{10}=H_{20}=0$ for C and C′. ——analytical result by using the present proposed procedure; – – – –analytical result by using the procedure proposed in [24]; ○ ◇ △ from digital simulation.

$$p(T|H_0)=\left.\frac{-\partial R(t|H_0)}{\partial t}\right|_{t=T}. \tag{47}$$

Similarly, the generalized Pontryagin equations for the moments of first-passage time of system (40) can be obtained as follows:

$$\frac{1}{2}b(H_0)\frac{\partial^2\mu_{l+1}}{\partial H_0^2}+a(H_0)\frac{\partial\mu_{l+1}}{\partial H_0}=-(l+1)\mu_l. \tag{48}$$

The boundary conditions are

$$\mu_{l+1}(H_c)=0 \tag{49}$$

$$\mu_{l+1}(0)=finite. \tag{50}$$

The qualitative condition (50) can be converted into quantitative one by using Eq. (48) and the limiting behavior of $a(H_0)$ and $b(H_0)$ near $H_0=0$. It is

$$\left[D_1+D_2-\frac{1}{6}(\beta_1+\beta_2)\left(\frac{1}{\omega_1^2}+\frac{1}{\omega_2^2}\right)\right]\frac{\partial\mu_{l+1}}{\partial H_0}=-(l+1)\mu_l. \tag{51}$$

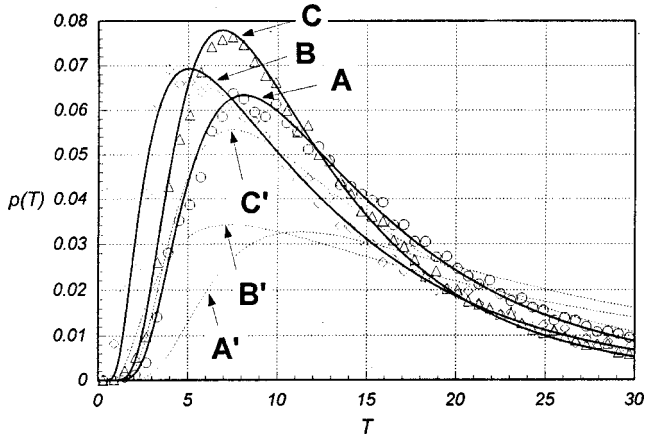


Fig. 9 Probability density of first-passage time of system (52) for given initial condition. The parameters and symbols are the same as those in Fig. 8.

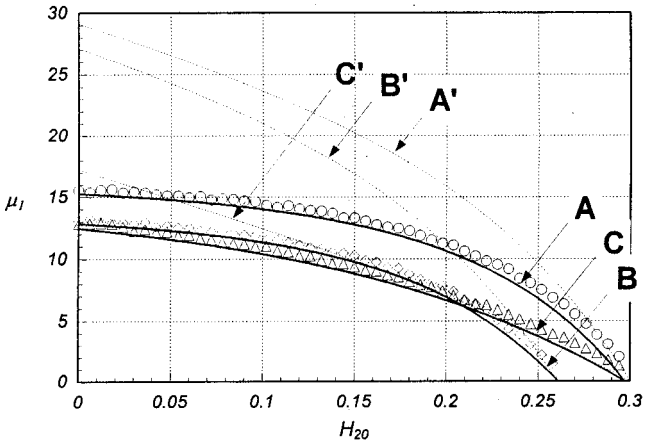


Fig. 10 Mean first-passage time of system (52) as function of H_{20} for given H_{10}. $2D_3=2D_4=0$, $H_{10}=0$ for A and A′; $2D_3=2D_4=0$, $H_{10}=0.04$ for B and B′; $2D_3=0.1$, $2D_4=0.05$, $H_{10}=0$ for C and C′. The other parameters and symbols are the same as those in Fig. 8.

The one-dimensional boundary value problem, Eqs. (48), (49), and (51), can be solved by using the Runge-Kutta method.

Obviously, for evaluating the statistics of the first-passage failure of system (24) the procedure for quasi-non-integrable Hamiltonian systems is much simpler than that for the quasi-integrable Hamiltonian system. However, the former generally yields inaccurate results as shown in Figs. 2–4. Our experience shows that it may yield good results in some very special cases, for example, the ratio of excitation intensity to damping coefficient for the first degree-of-freedom is the same as that for the second degree-of-freedom. In this case system (24) will behave like a quasi-non-integrable Hamiltonian system. On the other hand, the method proposed in this paper always yields good results for system (24) although the equations involved are more difficult to solve.

Example 2. Consider a van der Pol oscillator nonlinearly coupled with a Duffing oscillator subject to external and parametric excitations of Gaussian white noises. The equations of motion of the system are of the form

$$\begin{aligned}\ddot{X}_1+(-\beta_1+\alpha_1X_1^2+\alpha_2X_2^4+\alpha_3\dot{X}_2^2)\dot{X}_1+\omega^2X_1&=W_1(t)+X_1W_3(t)\\ \ddot{X}_2+(\beta_2+\alpha_4X_1^2)\dot{X}_2+kX_2^3&=W_2(t)+X_2W_4(t)\end{aligned} \tag{52}$$

where α_1, α_2, α_3, α_4, β_1, β_2, ω, k are constants; $W_k(t)(k=1,2,3,4)$ are independent Gaussian white noises with intensity $2D_k$. The response of system (52) with external excitations only has been studied by using the stochastic averaging method for quasi-integrable Hamiltonian systems ([19]). Let $X_1=Q_1$, $X_2=Q_2$, $\dot{X}_1=P_1$, $\dot{X}_2=P_2$, Eq. (52) can be rewritten as a quasi-Hamiltonian system of the form of Eq. (1), i.e.,

$$\begin{aligned}\dot{Q}_1&=P_1\\ \dot{Q}_2&=P_2\\ \dot{P}_1&=-\omega^2Q_1-(-\beta_1+\alpha_1Q_1^2+\alpha_2Q_2^4+\alpha_3P_2^2)P_1\\ &\quad+W_1(t)+Q_1W_3(t)\\ \dot{P}_2&=-kQ_2^3-(\beta_2+\alpha_4Q_1^2)P_2+W_2(t)+Q_2W_4(t).\end{aligned} \tag{53}$$

Equation (53) can be modeled as Itô equations. Since the Wong-Zakai correction terms for this example vanish, the modified Hamiltonian is the same as that associated with Eq. (53), i.e.,

$$H=H_1+H_2 \tag{54}$$

$$H_1=\tfrac{1}{2}(P_1^2+\omega^2Q_1^2) \tag{55}$$

$$H_2=\tfrac{1}{2}(P_2^2+\tfrac{1}{2}kQ_2^4). \tag{56}$$

Hamiltonian H is separable and so Eq. (53) governs a quasi-integrable Hamiltonian system. Suppose that the Hamiltonian system is nonresonant. The averaged Itô equations can be obtained from Eq. (53) by using the stochastic averaging method for quasi-integrable Hamiltonian systems ([19]). It is of the same form of Eq. (28) with the following drift and diffusion coefficients:

$$a_1=\beta_1H_1-\frac{\alpha_1}{2\omega^2}H_1^2-\frac{4\alpha_2}{3k}H_1H_2-\frac{4\alpha_3}{3}H_1H_2+D_1+\frac{H_1}{\omega^2}D_3$$

$$a_2=-\frac{4}{3}\beta_2H_2-\frac{4\alpha_4}{3\omega^2}H_1H_2+D_2+\frac{8\Gamma^2\left(\frac{7}{4}\right)}{9\Gamma^2\left(\frac{5}{4}\right)}\sqrt{\frac{H_2}{k}}D_4$$

$$b_{11}=2D_1H_1+\frac{H_1^2}{\omega^2}D_3 \tag{57}$$

$$b_{22}=\frac{8}{3}D_2H_2+\frac{64\Gamma^2\left(\frac{7}{4}\right)}{45\Gamma^2\left(\frac{5}{4}\right)}H_2\sqrt{\frac{H_2}{k}}D_4$$

$$b_{12}=b_{21}=0.$$

Since $H_i(i=1,2)$ vary from 0 to ∞ under the condition $k>0$, the safety domain of system (52) may be of the same form as that in Fig. 1. The backward Kolmogorov equation for the conditional reliability function, the generalized Pontryagin equations for the conditional moments of first-passage time, and their associated initial and boundary conditions for system (52) can be formulated and solved as for example 1. The only difference is that the drift and diffusion coefficients for this example are defined by Eq. (57) with H_1 and H_2 replaced by H_{10} and H_{20}, respectively.

The procedure for evaluating the statistics of first-passage failure of quasi-non-integrable Hamiltonian systems ([24]) can also be applied to applied to systems (52). The mathematical formulation is the same as that for example one, i.e., Eqs. (40)–(51), except the drift and diffusion coefficients. For this example, there coefficients are

$$a(H)=D_1+D_2+0.5484D_4\sqrt{\frac{H}{k}}+\frac{4}{7}\left(\beta_1-\beta_2+\frac{D_3}{\omega^2}\right)H$$
$$-\left(\frac{16}{17}\frac{\alpha_1}{k}+\frac{16}{17}\frac{\alpha_2}{\omega^2}+\frac{5}{77}\alpha_3+\frac{16}{17}\frac{\alpha_4}{\omega^2}\right)H^2 \tag{58}$$

$$b(H)=\overline{\sigma}^2(H)=0.4876D_4H\sqrt{\frac{H}{k}}+\frac{8}{7}(D_1+D_2)H+\frac{32}{17}\frac{D_3}{\omega^2}H^2.$$

Some numerical results for the conditional reliability function, the conditional probability density, and mean of first-passage time of system (52) are shown in Figs. 8–10. Some figures for this example similar to Figs. 5–7 are not given due to limited space. The same observations as those for example 1 can be made from these figures.

Conclusions

In the present paper a procedure for evaluating the statistics of the first passage failure, i.e., the conditional reliability function and the conditional probability density and moments of the first-passage time of quasi-integrable Hamiltonian systems has been proposed based on the stochastic averaging method for quasi-integrable Hamiltonian systems. Using the stochastic averaging method reduces the dimensions of the backward Kolmogorov equations governing the conditional reliability function and the generalized Pontryagin equations governing the conditional moments of first-passage time by a half when the associated Hamiltonian system is nonresonant. Furthermore, the backward Kolmogorov equation and generalized Pontryagin equations of an averaged system are nonsingular and much simpler than those for the original system. Applications of the proposed procedure to two examples show that the proposed procedure yields quite accurate results. Thus, the proposed procedure is promising and deserves further development and application.

The results for the two examples indicate that both the reliability and mean first-passage time are monotonously decreasing functions of initial energy of each degree-of-freedom of the system. This property will be used in the study of nonlinear stochastic optimal control of first-passage failure of quasi-integrable Hamiltonian systems.

The procedure for evaluating the statistics of the first-passage failure of quasi-non-integrable Hamiltonian systems has also been applied to the two examples. The numerical results showed that it generally yields inaccurate result for quasi-integrable Hamiltonian systems although it is much simpler than the procedure proposed in this paper. Experience shows that only in some very special cases it may yield good results.

It is remarked that the criteria for the failure considered in this paper are functions of the first integrals (energies) of the individual oscillators. The stochastic averaging method is the most effective for this kind of first-passage failure problem. If the failure criterion is given in terms of other physical quantity, such as the displacement, the first-passage failure problem will be much more difficult to solve. For such a kind of a first-passage failure problem of a single-degree-of-freedom quasi-Hamiltonian system, Roberts [31] developed an integral equation for evaluating the conditional transition probability density in the safety domain (the integral of which is the reliability function) by using the unconditional transition probability density obtained from solving the averaged FPK equation. Maybe this method can be extended to a multi-degree-of-freedom quasi-integrable Hamiltonian system but much more computational work is involved and some difficulties have to be solved. This will be the subject for our future research.

Acknowledgment

The work reported in this paper was supported by the National Natural Science Foundation of China under Grants No. 19972059 and 10002015 and the Cao Guang Biao Hi-Science-Technology Foundation of Zhejiang University.

References

[1] Bharucha-Reid, A. T., 1960, *Elements of Markov Processes and Their Applications*, McGraw-Hill, New York.
[2] Cox, D. R., and Miller, H. D., 1965, *The Theory of Stochastic Processes*, Chapman and Hall, New York.
[3] Bergman, L. A., and Heinrich, J. C., 1981, "On the Moments of Time to First Passage of the Linear Oscillator," Earthquake Eng. Struct. Dyn., **9**, pp. 197–204.
[4] Bergman, L. A., and Heinrich, J. C., 1982, "On the Reliability of the Linear Oscillator and Systems of Coupled Oscillators," Int. J. Numer. Methods Eng., **18**, pp. 1271–1295.
[5] Sun, J. Q., and Hsu, C. S., 1988, "First-Passage Time Probability of Non-Linear Stochastic Systems by Generalized Cell Mapping Method," J. Sound Vib., **124**, pp. 233–248.
[6] Sun, J. Q., and Hsu, C. S., 1990, "The Generalized Cell Mapping Method in Nonlinear Random Vibration Based Upon Short Time Gaussian Approximation," ASME J. Appl. Mech., **57**, pp. 1018–1025.
[7] Ariaratnam, S. T., and Pi, H. N., 1973, "On the First-Passage Time for Envelope Crossing for a Linear Oscillator," Int. J. Control, **18**, pp. 89–96.
[8] Lennox, W. C., and Fraser, D. A., 1974, "On the First Passage Distribution for the Envelope of a Non-stationary Narrow-Band Stochastic Process," ASME J. Appl. Mech., **41**, pp. 793–797.
[9] Ariaratnam, S. T., and Tam, D. S. F., 1979, "Random Vibration and Stability of a Linear Parametrically Excited Oscillator," Z. Angew. Math. Mech., **59**, pp. 79–84.
[10] Spanos, P. D., and Solomos, G. P., 1984, "Barrier Crossing due to Transient Excitation," J. Eng. Mech., **110**, pp. 20–36.

[11] Roberts, J. B., 1976, ''First Passage Probability for Nonlinear Oscillator,'' J. Eng. Mech., **102**, pp. 851–866.
[12] Roberts, J. B., 1978, ''First-Passage Time for Oscillator With Nonlinear Restoring Forces,'' J. Sound Vib., **56**, pp. 71–86.
[13] Roberts, J. B., 1986, ''Response of an Oscillator With Nonlinear Damping and a Softening Spring to Non-White Random Excitation,'' Probab. Eng. Mech., **1**, pp. 40–48.
[14] Roberts, J. B., 1986, ''First-Passage Time for Randomly Excited Nonlinear Oscillator,'' J. Sound Vib., **109**, pp. 33–50.
[15] Spanos, P. D., 1982, ''Survival Probability of Non-Linear Oscillators Subjected to Broad-Band Random Disturbance,'' Int. J. Non-Linear Mech., **17**, pp. 303–317.
[16] Zhu, W. Q., and Lei, Y., 1989, ''First Passage Time for State Transition of Randomly Excited Systems,'' *Proc. 47th Session of International Statistical Institute*, **LIII** (Invited Papers), Book 3, pp. 517–531.
[17] Cai, G. Q., and Lin, Y. K., 1994, ''On Statistics of First-Passage Failure,'' ASME J. Appl. Mech., **61**, pp. 93–99.
[18] Zhu, W. Q., and Yang, Y. Q., 1997, ''Stochastic Averaging of Quasi-Non-Integrable-Hamiltonian Systems,'' ASME J. Appl. Mech., **64**, pp. 157–164.
[19] Zhu, W. Q., Huang, Z. L., and Yang, Y. Q., 1997, ''Stochastic Averaging of Quasi-Integrable Hamiltonian Systems,'' ASME J. Appl. Mech., **64**, pp. 975–984.
[20] Zhu, W. Q., Huang, Z. L., and Suzuki, Y., 2002, ''Stochastic Averaging and Lyapunov Exponent of Quasi Partially Integrable Hamiltonian Systems,'' Int. J. Non-Linear Mech., **37**, pp. 419–437.
[21] Zhu, W. Q., and Huang, Z. L., 1998, ''Stochastic Stability of Quasi-Non-Integrable-Hamiltonian Systems,'' J. Sound Vib., **218**, pp. 769–789.
[22] Zhu, W. Q., and Huang, Z. L., 1999, ''Stochastic Hopf Bifurcation of Quasi-Non-Integrable-Hamiltonian Systems,'' Int. J. Non-Linear Mech., **34**, pp. 437–447.
[23] Zhu, W. Q., and Huang, Z. L., 1999, ''Lyapunov Exponents and Stochastic Stability of Quasi-Integrable-Hamiltonian Systems,'' ASME J. Appl. Mech., **66**, pp. 211–217.
[24] Gan, C. B., and Zhu, W. Q., 2001, ''First-Passage Failure of Quasi-Non-Integrable-Hamiltonian Systems,'' Int. J. Non-Linear Mech., **36**(2), pp. 209–220.
[25] Zhu, W. Q., and Ying, Z. G., 1999, ''Optimal Nonlinear Feedback Control of Quasi-Hamiltonian Systems,'' Sci. China, Ser. A: Math., Phys., Astron., **42**(11), pp. 1213–1219.
[26] Zhu, W. Q., Ying, Z. G., and Soong, T. T., 1999, ''Optimal Nonlinear Feedback Control of Structures Under Random Loading,'' *Stochastic Structural Dynamics*, B. F. Spencer, Jr., and E. A. Johnson, eds., Balkema, Rotterdam, pp. 141–148.
[27] Zhu, W. Q., Ying, Z. G., Ni, Y. Q., and Ko, J. M., 2000, ''Optimal Nonlinear Stochastic Control of Hysteretic Systems,'' J. Eng. Mech., **126**, pp. 1027–1032.
[28] Zhu, W. Q., Ying, Z. G., and Soong, T. T., 2001, ''An Optimal Nonlinear Feedback Control Strategy for Randomly Excited Structural Systems,'' Nonlinear Dyn., **24**, pp. 31–51.
[29] Zhu, W. Q., and Ying, Z. G., 2002, ''Nonlinear Stochastic Optimal Control of Partially Observable Linear Strucutres, '' Eng. Struct., **24**, pp. 333–342.
[30] Khasminskii, R. Z., 1968, ''On the Averaging Principle for Stochastic Differential Itô Equations,'' Kibernetica, **4**, pp. 260–279 (in Russian).
[31] Roberts, J. B., 1978, ''First Passage Time for Oscillators With Non-Linear Restoring Forces,'' J. Sound Vib., **56**, pp. 71–86.

Elastic Wave Propagation in Circumferential Direction in Anisotropic Cylindrical Curved Plates

S. Towfighi

T. Kundu
Fellow ASME

M. Ehsani
Department of Civil Engineering
and Engineering Mechanics,
University of Arizona,
Tucson, AZ 85721

Ultrasonic nondestructive inspection of large-diameter pipes is important for health monitoring of ailing infrastructure. Longitudinal stress-corrosion cracks are detected more efficiently by inducing circumferential waves; hence, the study of elastic wave propagation in the circumferential direction in a pipe wall is essential. The current state of knowledge lacks a complete solution of this problem. Only when the pipe material is isotropic a solution of the wave propagation problem in the circumferential direction exists. Ultrasonic inspections of reinforced concrete pipes and pipes retrofitted by fiber composites necessitate the development of a new theoretical solution for elastic wave propagation in anisotropic curved plates in the circumferential direction. Mathematical modeling of the problem to obtain dispersion curves for curved anisotropic plates leads to coupled differential equations. Unlike isotropic materials for which the Stokes-Helmholtz decomposition technique simplifies the problem, in anisotropic case no such general decomposition technique works. These coupled differential equations are solved in this paper. Dispersion curves for anisotropic curved plates of different curvatures have been computed and presented. Some numerical results computed by the new technique have been compared with those available in the literature. [DOI: 10.1115/1.1464872]

Introduction

Mathematical modeling of wave propagation in the axial direction of a cylinder has been studied extensively. However, for wave propagation in the circumferential direction, which is essential for nondestructive testing (NDT) of large diameter pipes, literature shows fewer investigations. Viktorov's work ([1]) establishes the fundamental mathematical modeling of the problem for isotropic material properties. He has introduced the angular wave number concept and has derived, decomposed and solved the governing differential equations. He has considered only one curved surface; in other words, he has found the solution for convex and concave cylindrical surfaces. In order to obtain the results for curved plates Qu et al. [2] have added the boundary conditions for the second surface and solved the problem of guided wave propagation in isotropic curved plates. Different aspects of the circumferential direction wave propagation along one or multiple curved surfaces have been analyzed by Grace and Goodman [3], Brekhovskikh [4], Cerv [5], Liu and Qu [6,7] and Valle, Qu, and Jacobs [8]. In all these works the material has been modeled as isotropic elastic material.

Many investigators have solved elastic wave propagation problem in homogeneous and multilayered anisotropic solids. However, all those works have been limited to the flat-plate case ([9]) or for waves propagating in the axial direction of a cylinder ([10]). Wave propagation in the circumferential direction of an anisotropic curved plate has not been analyzed earlier, and solved for the first time in this paper.

Unlike isotropic materials for which the Stokes-Helmholtz decomposition technique simplifies the problem, for anisotropic case no such general decomposition technique works. The differential equations remain coupled and require a more general solution technique.

The new technique, presented in this paper, solves coupled set of differential equations without attempting to decouple the equations. Hence it removes the obstacle arising from not being able to decouple the equations. Consequently it provides a systematic and unifying solution method, which is capable of solving a set of coupled differential equations, and can be utilized to solve a variety of wave propagation problems.

Fundamental Equations

The formulation presented here is for the wave propagation in a cylindrical curved plate in the direction of the curvature as shown in Fig. 1. We will interchangeably call the wave carrier a ''curved plate,'' ''cylinder,'' ''pipe segment,'' or simply ''pipe'' all meaning the same thing. What we are interested in is analyzing the dispersive waves in the curved plate for waves propagating from section T to R (see Fig. 1). This analysis does not include the reflected guided waves from the plate boundary. The problem geometry can be a segment of a cylinder or a complete cylinder.

Wave propagation in circumferential direction in pipes with isotropic material properties is usually modeled as a plane strain problem; i.e., the displacement component along the longitudinal axis of the pipe is set equal to zero. For a few other types of anisotropy this situation remains valid. However, for general anisotropy the longitudinal component of displacement must be considered in the mathematical modeling. The symmetry of both geometry and material properties is required for plane-strain idealization. In absence of such symmetry a three-dimensional mathematical modeling is necessary.

In cylindrical coordinates, strain components in terms of displacements can be written as

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, April 5, 2001; final revision, November 1, 2001. Associate Editor: A. K. Mal. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

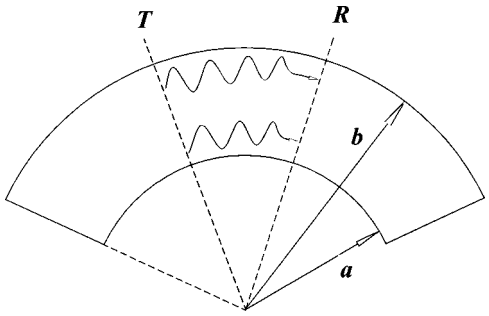


Fig. 1 Waves propagating from section *T* to *R* in a curved plate. Wave speed is proportional to radius of curvature.

$$e_{rr}=\frac{\partial u_r(r,\theta,t)}{\partial r}$$

$$e_{\theta\theta}=\frac{\partial u_\theta(r,\theta,t)}{r\partial\theta}+\frac{1}{r}u_r(r,\theta,t)$$

$$e_{zz}=\frac{\partial u_z(r,\theta,t)}{\partial z}$$

$$e_{r\theta}=\frac{1}{2}\left(\frac{\partial u_r(r,\theta,t)}{r\partial\theta}+\frac{\partial u_\theta(r,\theta,t)}{\partial r}-\frac{u_\theta(r,\theta,t)}{r}\right)$$

$$e_{\theta z}=\frac{1}{2}\left(\frac{\partial u_z(r,\theta,t)}{r\partial\theta}+\frac{\partial u_\theta(r,\theta,t)}{\partial z}\right)$$

$$e_{rz}=\frac{1}{2}\left(\frac{\partial u_r(r,\theta,t)}{\partial z}+\frac{\partial u_z(r,\theta,t)}{\partial r}\right). \tag{1}$$

The stress and displacement components are shown in Fig. 2. And constitutive matrix for general anisotropy contains 21 independent elastic constants:

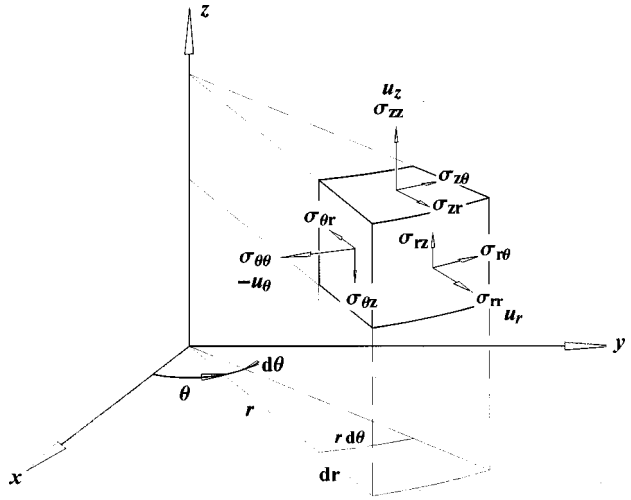


Fig. 2 Stress and displacement components in cylindrical coordinate system

$$\begin{pmatrix}\sigma_{\theta,\theta}\\ \sigma_{z,z}\\ \sigma_{r,r}\\ \sigma_{\theta,z}\\ \sigma_{r,\theta}\\ \sigma_{r,z}\end{pmatrix}=\begin{pmatrix}C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} & C_{1,6}\\ C_{1,2} & C_{2,2} & C_{2,3} & C_{2,4} & C_{2,5} & C_{2,6}\\ C_{1,3} & C_{2,3} & C_{3,3} & C_{3,4} & C_{3,5} & C_{3,6}\\ C_{1,4} & C_{2,4} & C_{3,4} & C_{4,4} & C_{4,5} & C_{4,6}\\ C_{1,5} & C_{2,5} & C_{3,5} & C_{4,5} & C_{5,5} & C_{5,6}\\ C_{1,6} & C_{2,6} & C_{3,6} & C_{4,6} & C_{5,6} & C_{6,6}\end{pmatrix}\begin{pmatrix}e_{\theta,\theta}\\ e_{z,z}\\ e_{r,r}\\ 2e_{\theta,z}\\ 2e_{r,\theta}\\ 2e_{r,z}\end{pmatrix}. \tag{2}$$

Equations of motion for three components of displacement in cylindrical coordinates are as follows:

$$\frac{\partial\sigma_{rr}}{\partial r}+\frac{\partial\sigma_{rz}}{\partial z}+\frac{\partial\sigma_{r\theta}}{r\partial\theta}+\frac{\sigma_{rr}-\sigma_{\theta\theta}}{r}-\rho\frac{\partial^2u_r(r,\theta,t)}{\partial t^2}=0$$

$$\frac{\partial\sigma_{r\theta}}{\partial r}+\frac{\partial\sigma_{\theta z}}{\partial z}+\frac{\partial\sigma_{\theta\theta}}{r\partial\theta}+\frac{2\sigma_{r\theta}}{r}-\rho\frac{\partial^2u_\theta(r,\theta,t)}{\partial t^2}=0$$

$$\frac{\partial\sigma_{rz}}{\partial r}+\frac{\partial\sigma_{zz}}{\partial z}+\frac{\partial\sigma_{\theta z}}{r\partial\theta}+\frac{\sigma_{rz}}{r}-\rho\frac{\partial^2u_z(r,\theta,t)}{\partial t^2}=0. \tag{3}$$

Stress components in the above equations can be substituted in terms of displacement components. Since displacement components are functions of wave forms, time dependency of waves must be established.

Wave Form

In cylindrical geometry the generation of surface waves in the circumferential direction with a plane wave front requires the circumferential wave speed to be a function of the radial distance. Viktorov [1] has introduced this concept and called it the angular wave number. Similar formulation has been adapted here:

$$u_r(r,\theta,t)=U_r(r)e^{i(p\theta-\omega t)}$$

$$u_\theta(r,\theta,t)=U_t(r)e^{i(p\theta-\omega t)}$$

$$u_z(r,\theta,t)=U_z(r)e^{i(p\theta-\omega t)} \tag{4}$$

where $U_r(r)$, $U_t(r)$, and $U_z(r)$ represent the amplitude of vibration in the radial, tangential, and axial directions, respectively. "i" is the imaginary number $\sqrt{-1}$. It should be noted here that the phase velocity is not a constant and changes with radius. As shown in Fig. 1 the phase velocity has to be proportional to the radius to have a plane wave front. Hence, if c_b is assumed to be the phase velocity at the outer surface with radius b; for other points having a radius r the phase velocity would be

$$v_{ph}(r)=c_br/b. \tag{4a}$$

For the flat-plate case wave number k is defined as ω/v_{ph} because curvature does not change. However, for a curved plate the same definition would be r dependent. Thus the angular wave number p, which is independent of r is defined as

$$p=\omega/(v_{ph}(r)/r)=\omega b/c_b. \tag{4b}$$

Governing Differential Equations

Subsequent substitution of Eqs. (4), (1), and (2) into Eq. (3) yields the following governing differential equations:

$$
\begin{aligned}
&-2C_{5,5}U_r(r)p^2-2C_{1,5}U_t(r)p^2-2C_{4,5}U_z(r)p^2-2iC_{1,1}U_t(r)p\\
&\quad-2iC_{5,5}U_t(r)p-2iC_{1,4}U_z(r)p+4irC_{3,5}U_r'(r)p\\
&\quad+2irC_{1,3}U_t'(r)p+2irC_{5,5}U_t'(r)p+2irC_{3,4}U_z'(r)p\\
&\quad+2irC_{5,6}U_z'(r)p+2r^2\rho\omega^2U_r(r)-2C_{1,1}U_r(r)\\
&\quad+2C_{1,5}U_t(r)+2rC_{3,3}U_r'(r)-2rC_{1,5}U_t'(r)-2rC_{1,6}U_z'(r)\\
&\quad+2rC_{3,6}U_z'(r)+2r^2C_{3,3}U_r''(r)2r^2C_{3,5}U_t''(r)\\
&\quad+2r^2C_{3,6}U_z''(r)=0\\
&-2C_{1,5}U_r(r)p^2-2C_{1,1}U_t(r)p^2-2C_{1,4}U_z(r)p^2+2iC_{1,1}U_r(r)p\\
&\quad+2iC_{5,5}U_r(r)p+2iC_{4,5}U_z(r)p+2irC_{1,3}U_r'(r)p\\
&\quad+2irC_{5,5}U_r'(r)p+4irC_{1,5}U_t'(r)p+2irC_{1,6}U_z'(r)p\\
&\quad+2irC_{4,5}U_z'(r)p+2C_{1,5}U_r(r)+2r^2\rho\omega^2U_t(r)\\
&\quad-2C_{5,5}U_t(r)+2rC_{1,5}U_r'(r)+4rC_{3,5}U_r'(r)+2rC_{5,5}U_t'\\
&\quad+4rC_{5,6}U_z'(r)+2r^2C_{3,5}U_r''(r)+2r^2C_{5,5}U_t''(r)\\
&\quad+2r^2C_{5,6}U_z''(r)=0\\
&-2C_{4,5}U_r(r)p^2-2C_{1,4}U_t(r)p^2-2C_{4,4}U_z(r)p^2+2iC_{1,4}U_r(r)p\\
&\quad-2iC_{4,5}U_t(r)p+2irC_{3,4}U_r'(r)p+2irC_{5,6}U_r'(r)p\\
&\quad+2irC_{1,6}U_t'(r)p+2irC_{4,5}U_t'(r)p+4irC_{4,6}U_z'(r)p\\
&\quad+2r^2\rho\omega^2U_z(r)+2rC_{1,6}U_r'(r)+2rC_{3,6}U_r'(r)\\
&\quad+2rC_{6,6}U_z'(r)+2r^2C_{3,6}U_r''(r)+2r^2C_{5,6}U_t''(r)\\
&\quad+2r^2C_{6,6}U_z''(r)=0. \qquad (5)
\end{aligned}
$$

Boundary Conditions

In order to obtain the dispersion curves, the traction-free boundary conditions (zero stress values on the inner and outer surfaces of the pipe) must be satisfied. Hence, at $r=a$ and $r=b$:

$$
\begin{aligned}
&C_{1,3}U_r(r)+ipC_{3,5}U_r(r)+ipC_{1,3}U_t(r)-C_{3,5}U_t(r)+ipC_{3,4}U_z(r)\\
&\quad+rC_{3,3}U_r'(r)+rC_{3,5}U_t'(r)+rC_{3,6}U_z'(r)=0\\
&C_{1,5}U_r(r)+ipC_{5,5}U_r(r)+ipC_{1,5}U_t(r)-C_{5,5}U_t(r)-ipC_{4,5}U_z(r)\\
&\quad+rC_{3,5}U_r'(r)+rC_{5,5}U_t'(r)+rC_{5,6}U_z'(r)=0\\
&C_{1,6}U_r(r)+ipC_{5,6}U_r(r)+ipC_{1,6}U_t(r)-C_{5,6}U_t(r)+ipC_{4,6}U_z(r)\\
&\quad+rC_{3,6}U_r'(r)+rC_{5,6}U_t'(r)+rC_{6,6}U_z'(r)=0. \qquad (6)
\end{aligned}
$$

Solution

It can be seen that all differential equations are functions of three displacement components and their derivatives. It should be also noted that $U_r(r)$, $U_t(r)$, and $U_z(r)$ are functions of the radius only and they appear in all equations. Therefore, there are three coupled differential equations and six boundary conditions that must be satisfied simultaneously.

To solve the equations, the unknown functions are expanded in Fourier series (FS). Substitution of FS expansions into the differential equations provides three algebraic equations that must be satisfied for the entire problem domain. To satisfy the equations for a given number of FS terms weighted residuals integration with a linear weight function has been utilized:

$$
R=\int_a^b wf(r,x_i)dr=0. \qquad (7)
$$

The radius corresponding to the peak value of the linear weight function can take any value between the inner and the outer radius, each resulting one independent equation. Hence from every differential equation any number of equations can be obtained.

On the other hand, it is known that the general solution is a linear combination of all solution functions that can be obtained. Therefore, the general solution should contain combinatorial parameters. The number of combinatorial parameters is the same as the number of individual solutions. These combinatorial parameters are necessary to satisfy the boundary conditions. Satisfaction of six boundary conditions requires six parameters and six equations. Therefore the necessary and sufficient number of combinatorial parameters is six and it indicates the existence of six independent solutions.

Substitution of solution functions into the differential equations leads to three equations, each containing all of the FS parameters. In other words, all FS parameters for the three amplitude functions appear in every equation. Because of this coupling, the values of parameters obtained for FS expansion of $U_r(r)$, $U_t(r)$, and $U_z(r)$ are not independent and a solution must yield all parameters as one set of results. Since the equations are linear and the results must be combined using combinatorial parameters only their relative values must be found. Therefore one of the FS parameters can be assumed equal to one. Then the relative values for other FS parameters can be calculated in terms of this unit value. Each set of the parameter values defines a set of dependent shapes for the above amplitude functions; these are called basic shapes. Since the number of equations must be equal to the number of unknowns a specific number of weight functions are required.

The FS expansion for $U_r(r)$ can be written as

$$
U_r(r)=x_0+\sum_{n=1}^{m}\left(\cos\left(\frac{n\pi r}{L}\right)x_n+\sin\left(\frac{n\pi r}{L}\right)y_n\right) \qquad (8)
$$

which contains $2m+1$ parameters or coefficients, x_n and y_n.

With two other expressions for $U_t(r)$ and $U_z(r)$ the number of unknowns increases to $6m+3$. Performing weighted residuals method, a set of linear equations results:

$$
\begin{pmatrix}
a_{1,1}x_1 & a_{1,2}x_2 & \cdots & a_{1,s}x_s & a_{1,s+1}x_{s+1} & a_{1,s+6}x_{s+6}\\
a_{2,1}x_1 & a_{2,2}x_2 & \cdots & a_{2,s}x_s & a_{2,s+1}x_{s+1} & a_{2,s+1}x_{s+6}\\
 & & \cdots & & & \\
 & & \cdots & & & \\
 & & \cdots & & & \\
a_{s,1}x_1 & a_{s,2}x_2 & \cdots & a_{s,s}x_s & a_{s,s+1}x_{s+1} & a_{s,s+6}x_{s+6}
\end{pmatrix}=\begin{pmatrix}0\\0\\ \\ \\ \\0\end{pmatrix} \qquad (9)
$$

where $x_{s+1}, x_{s+2}, \ldots, x_{s+6}$ represent the last sine and cosine terms of FS expansions. Assigning six independent unit vectors to the last six parameters as shown in Eq. (10),

$$\begin{pmatrix} x^1_{s+1} & x^2_{s+1} & x^3_{s+1} & x^4_{s+1} & x^5_{s+1} & x^6_{s+1} \\ x^1_{s+2} & x^2_{s+2} & x^3_{s+2} & x^4_{s+2} & x^5_{s+2} & x^6_{s+2} \\ x^1_{s+3} & x^2_{s+3} & x^3_{s+3} & x^4_{s+3} & x^5_{s+3} & x^6_{s+3} \\ x^1_{s+4} & x^2_{s+4} & x^3_{s+4} & x^4_{s+4} & x^5_{s+4} & x^6_{s+4} \\ x^1_{s+5} & x^2_{s+5} & x^3_{s+5} & x^4_{s+5} & x^5_{s+5} & x^6_{s+5} \\ x^1_{s+6} & x^2_{s+6} & x^3_{s+6} & x^4_{s+6} & x^5_{s+6} & x^6_{s+6} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{10}$$

yields six independent solutions. Therefore the number of equations has to be $s=6m-3$. Consequently, the general solution can be obtained as a linear combination of the above solutions:

$$A_1 \begin{pmatrix} {}^1_1 \\ x^1_2 \\ \\ \\ x^1_s \end{pmatrix} + A_2 \begin{pmatrix} x^2_1 \\ x^2_2 \\ \\ \\ x^2_s \end{pmatrix} + A_3 \begin{pmatrix} x^3_1 \\ x^3_2 \\ \\ \\ x^3_s \end{pmatrix} + A_4 \begin{pmatrix} x^4_1 \\ x^4_2 \\ \\ \\ x^4_s \end{pmatrix} + A_5 \begin{pmatrix} x^5_1 \\ x^5_2 \\ \\ \\ x^5_s \end{pmatrix} + A_6 \begin{pmatrix} x^6_1 \\ x^6_2 \\ \\ \\ x^6_s \end{pmatrix}. \tag{11}$$

The superscript for FS parameters shows the solution set number. Substitution of the obtained FS parameters into stress components on the inner and outer surfaces of the pipe leads to an eigenvalue problem. The determinant of the coefficients of A_i should be zero for any point located on the dispersion curves.

Numerical Results

Based on the proposed mathematical modeling a Mathematica program has been developed. To ensure the validity of the modeling and the computer program, its results are compared with the available dispersion curves for anisotropic flat plates by using small ratios of thickness to radius, when pipe geometry approaches flat plate geometry. Additionally, the results are compared with the published results for isotropic pipes ([2]). Since the exact input values have not been reported by Qu et al. [2], the comparison is done only qualitatively. The dispersion curves are also given for anisotropic pipes.

A Comparison With Available Data for Isotropic Flat Plate. Dispersion curves for a flat plate are given in Mal and Singh [11], see Fig. 3. Curves for the same plate thickness and material properties, but having an outer radius of 1 m, are generated by the proposed method and shown in Fig. 4.

A comparison of Figs. 3 and 4 shows a very good match between the two when only 20 terms are used in the FS expansions.

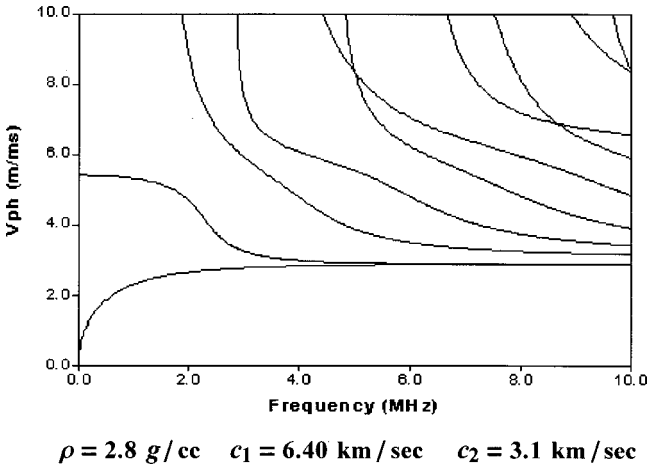


Fig. 3 Dispersion curves for isotropic flat plate ([11]). Plate thickness=1 mm.

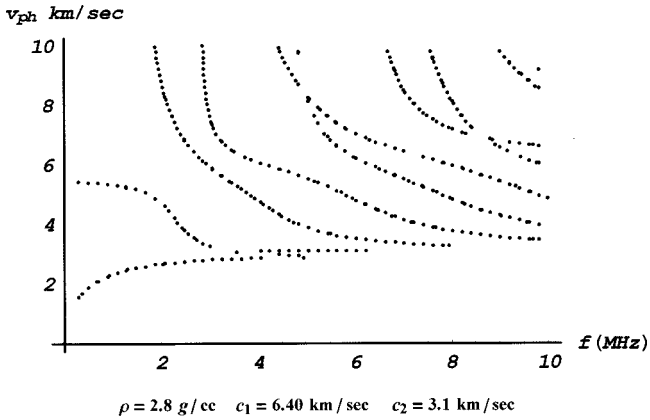


Fig. 4 Dispersion curves generated by the proposed method. Plate thickness=1 mm. Pipe outside radius=1.0 m.

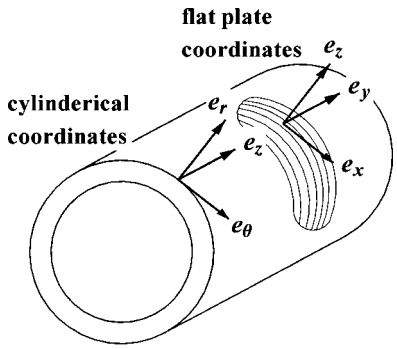


Fig. 5 Tangential direction of the fibers maintains the symmetry. Coordinate systems for flat-plate and pipe analyses are also shown.

B Comparison With Available Data for Anisotropic Flat Plate. Dispersion curves for anisotropic flat plates are available in the literature ([12,13]). In this section our results are compared with those given in Rose [13].

For the unidirectional composite plate or pipe with a zero-degree angle between the wave propagation direction and the fiber direction as shown in Fig. 5, the material and the geometric symmetry conditions are maintained; hence, the plain-strain formulation remains valid. Consequently the constitutive matrix reduces to the following form:

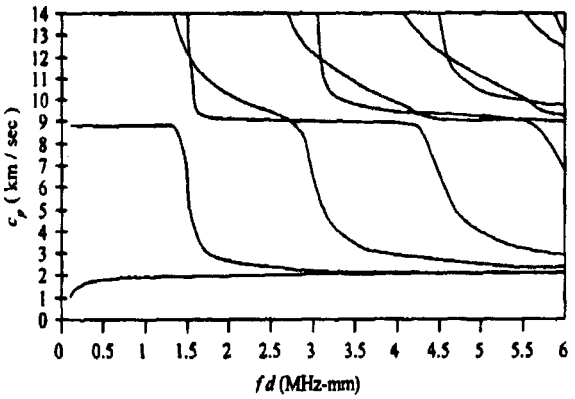


Fig. 6 Dispersion curves of a unidirectional composite plate for waves propagating in fiber direction (*x*-axis direction, 0 deg). Material properties are given in Eq. (12), $\rho = 1580$ kg/m^3 ([3]).

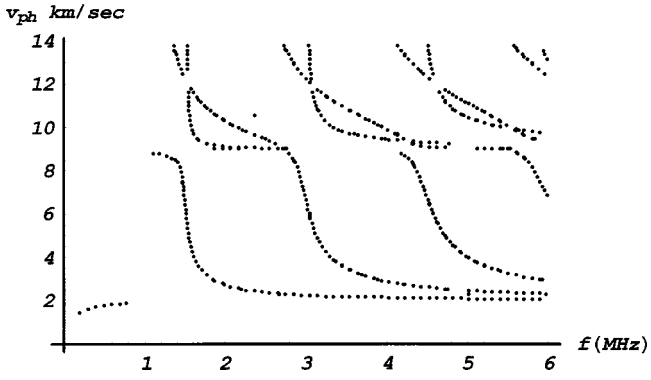


Fig. 7 Dispersion curves for a large-diameter pipe made of an anisotropic material. Material properties are given in Eq. (12). Pipe wall thickness = 1 mm. Pipe outer radius = 1000 mm, $m = 30$.

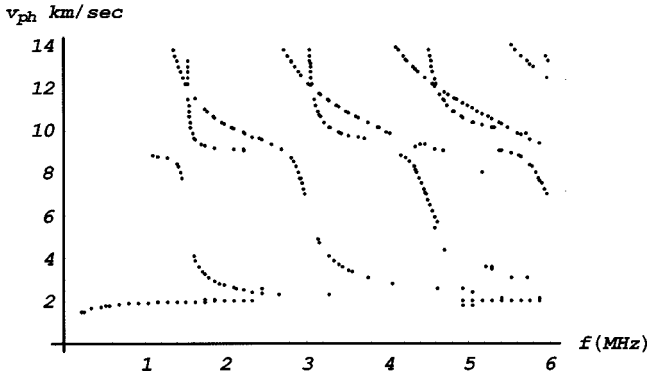


Fig. 8 Dispersion curves for the anisotropic pipe with $m = 20$. Pipe dimensions and material properties are same as in Fig. 7, only *m* is different.

$$\begin{pmatrix} \sigma_{\theta\theta} \\ \sigma_{zz} \\ \sigma_{rr} \\ \sigma_{r\theta} \end{pmatrix} = \begin{pmatrix} 128.2 & 6.9 & 6.9 & 0 \\ 6.9 & 14.95 & 7.33 & 0 \\ 6.9 & 7.33 & 14.95 & 0 \\ 0 & 0 & 0 & 6.73 \end{pmatrix} \begin{pmatrix} e_{\theta\theta} \\ 0 \\ e_{rr} \\ 2e_{r\theta} \end{pmatrix}. \quad (12)$$

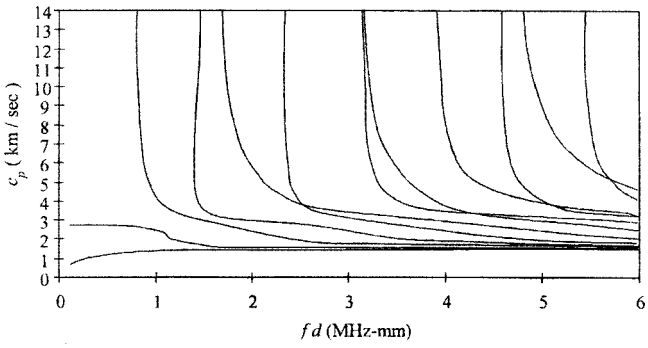


Fig. 9 Dispersion curves of unidirectional composite plate for waves propagating perpendicular to the fiber direction (*x*-axis direction, 90 deg). Material properties are given in Eq. (13). Plate thickness = 1 mm, $\rho = 1580$ kg/m^3 ([3]).

Stiffness values are given in GPa. Flat-plate results are shown in Fig. 6. Results for the curved plate are shown in Figs. 7 and 8.

The result of Fig. 7 is obtained using 30 terms ($m = 30$) in the Fourier series expansion. To show the effect of the number of terms (m) on the computed results the same dispersion curves are computed for $m = 20$ and shown in Fig. 8.

It is interesting to note that smaller value of m gives broken lines. Therefore the user can easily realize the need for a greater number of terms in the FS expansion when the lines in the dispersion curve plot are found broken. There are some missing parts of curves in Fig. 7 that can be obtained by increasing m. However, for $m = 30$ we get enough information for comparison with the results given by Rose [13].

For the same material with fibers going in the longitudinal direction of the pipe, the constitutive matrix changes to Eq. (13).

$$\begin{pmatrix} \sigma_{\theta\theta} \\ \sigma_{zz} \\ \sigma_{rr} \\ \sigma_{r\theta} \end{pmatrix} = \begin{pmatrix} 14.95 & 6.9 & 7.33 & 0 \\ 6.9 & 128.2 & 6.9 & 0 \\ 7.33 & 6.9 & 14.95 & 0 \\ 0 & 0 & 0 & 3.81 \end{pmatrix} \begin{pmatrix} e_{\theta\theta} \\ 0 \\ e_{rr} \\ 2e_{r\theta} \end{pmatrix}. \quad (13)$$

Obtained results for this case also match with the corresponding dispersion curves presented by Rose [13]; see Figs. 9 and 10.

For the case where fibers are oriented at 45 deg relative to the pipe axis, plane-strain assumptions are no longer valid. The constitutive matrix for this case is obtained by transformation of the coordinate system as shown in Eq. (14). See Figs. 11(*a*), 11(*b*) and 12 for comparison. This case also shows an excellent match between the available data and the obtained results.

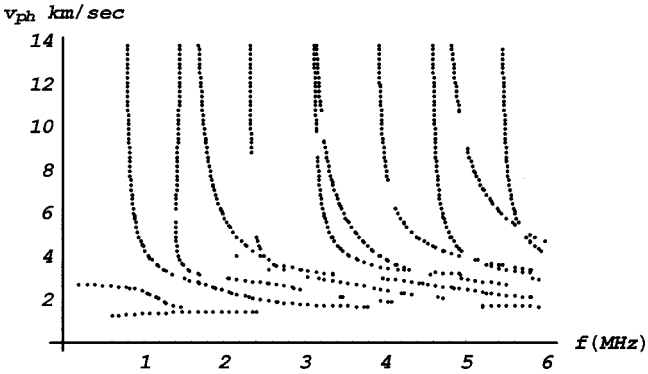


Fig. 10 Computed dispersion curves for an anisotropic large diameter pipe, when fiber and wave propagation directions are perpendicular to each other. Material properties are given in Eq. (13). Pipe wall thickness = 1 mm. Pipe outer radius = 1000 mm.

$$\begin{pmatrix}\sigma_{\theta,\theta}\\ \sigma_{z,z}\\ \sigma_{r,r}\\ \sigma_{z,r}\\ \sigma_{\theta,r}\\ \sigma_{\theta,z}\end{pmatrix}=\begin{pmatrix}45.9675 & 32.5075 & 7.115 & 0 & 0 & -28.3125\\ 32.5075 & 45.9675 & 7.115 & 0 & 0 & -28.3125\\ 7.115 & 7.115 & 14.95 & 0 & 0 & 0.215\\ 0 & 0 & 0 & 5.27 & -1.46 & 0\\ 0 & 0 & 0 & -1.46 & 5.27 & 0\\ -28.3125 & -28.3125 & 0.215 & 0 & 0 & 32.3375\end{pmatrix}\begin{pmatrix}e_{\theta,\theta}\\ e_{z,z}\\ e_{r,r}\\ 2e_{z,r}\\ 2e_{\theta,r}\\ 2e_{\theta,z}\end{pmatrix} \tag{14}$$

Since for the curved plate, the midplane is not the plane of symmetry, the dispersion curves cannot be grouped as symmetric and antisymmetric modes. That is why all modes are shown together in Fig. 12 for a large-diameter pipe.

C Comparison With Available Data for Isotropic Pipe As mentioned earlier, Qu et al. [2] have derived dispersion curves for aluminum pipes but the material properties have not been reported in their work. Hence, the quantitative comparison was not possible. However, curves presented here, Fig. 13, qualitatively look similar to those of Qu et al. [2], Fig. 14. Figures 13 and 14 show the obtained dispersion curves with non-dimensional $\bar{k}$ and $\bar{\omega}$ where $\bar{k}=k(b-a)$ and

$$\bar{\omega}=\omega(b-a)\sqrt{\frac{\rho}{\mu}}.$$

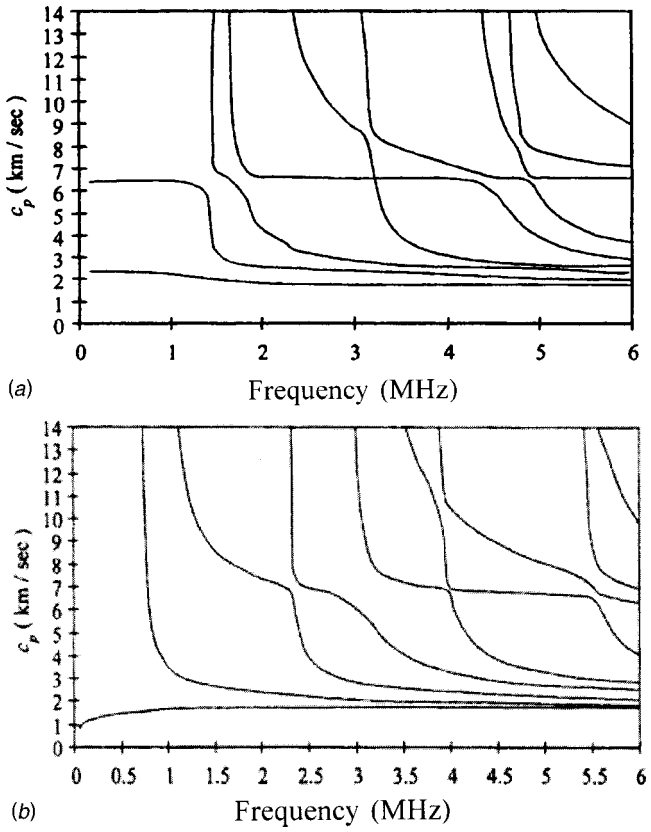


Fig. 11 (*a*) Dispersion curves for symmetric modes for a unidirectional composite plate for waves propagating in 45 deg to the fiber direction. Plate thickness=1 mm and ρ=1580 kg/m^3 ([3]). (*b*) Dispersion curves for antisymmetric modes for a unidirectional composite plate for waves propagating in 45 deg to the fiber direction. Plate thickness=1 mm and ρ=1580 kg/m^3 ([3]).

D Anisotropic Pipe of Small Radius of Curvature. To show the effect of the radius of curvature on the dispersion curves the pipe radius is varied from 1000 mm to 2.5 mm keeping the wall thickness and material properties same as those mentioned in the figure captions for Figs. 7 and 9. Dispersion curves obtained by the 30 terms FS expansion for $r=1000$, 10, 5, and 2.5 mm are shown in Figs. 15 and 16. Figure 15 shows dispersion curves for fibers going in the circumferential direction and Fig. 16 is for fibers going in the axial direction while the waves propagate in the circumferential directions in both cases.

From Figure 15 one can see that for fibers oriented in the circumferential direction the dispersion curves do not change significantly as the outer radius (r) is reduced from 1000 mm to 10 mm. However, as r is reduced further the deviation of the dispersion curves from the large radius case is no longer negligible. For

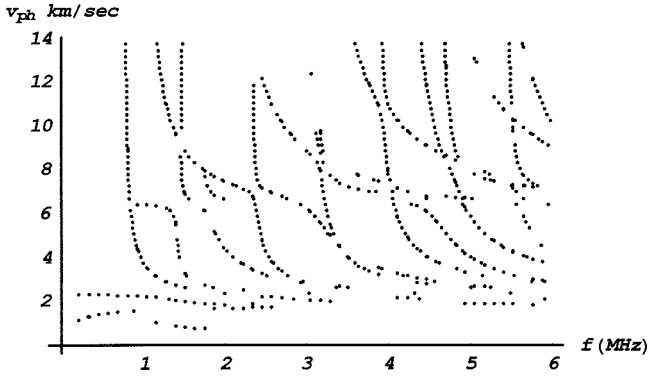


Fig. 12 Dispersion curves for a large diameter pipe made of an anisotropic material. Material properties are given in Eq. (14). Pipe wall thickness=1 mm. Pipe outer radius=1000 mm, m =25.

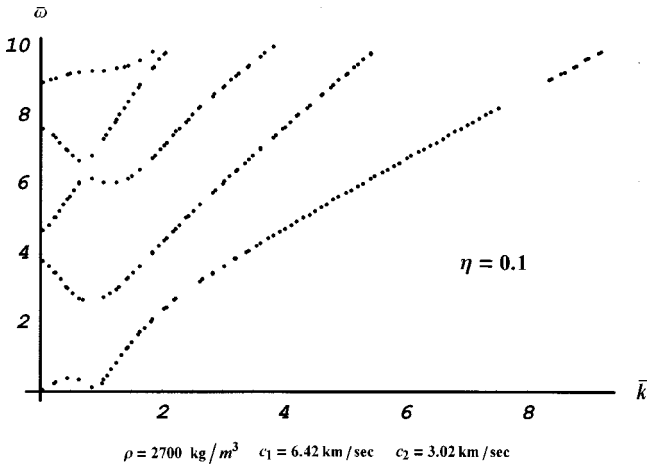


Fig. 13 Dispersion curves for aluminum pipe obtained by the proposed method. η (ratio of inner to outer radius)=0.1.

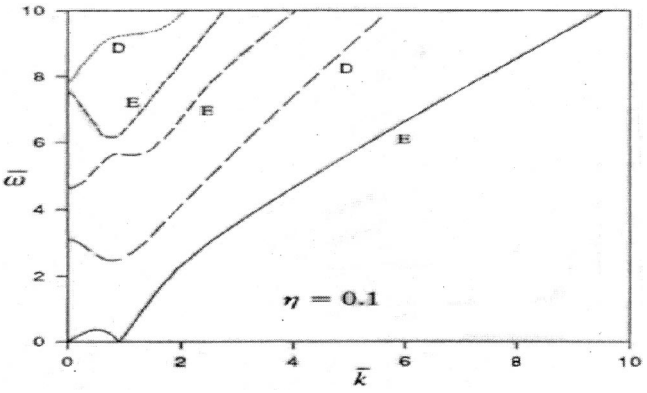


Fig. 14 Dispersion curves for aluminum pipe obtained by Qu et al. [2]. Material properties are not known.

fibers oriented in the axial direction (Fig. 16) the dispersion curves remain almost unchanged for $r = 1000$ mm down to 2.5 mm. For $r = 2.5$ mm the dispersion curves are obtained with $m = 45$ in FS expansion of amplitude functions. The computation with $m = 30$ gave too many broken lines in the dispersion curve plot for $r = 2.5$ mm.

In summary, a comparison between Figs. 15 and 16 shows that the effect of curvature is stronger when the fibers are oriented along the circumferential direction and hence when the fibers also have a curvature. When the fibers are oriented in the axial direction and hence don't have any curvature the flat-plate approximation can be extended to pipes of much lower radius.

Dispersion curves for the 5-mm outer radius pipe with fibers oriented along the 45 deg direction are shown in Fig. 17. This result is obtained for the material properties given in Eq. (14). In the frequency range smaller than 1 MHz, some vertical lines appeared due to the numerical errors when the number (m) of FS terms is 25. By increasing m to 35 those lines disappeared. The results for $m = 35$ are shown on the left side of Fig. 17.

Conclusion

A solution technique based on the Fourier series expansion of the unknown quantities has been introduced to solve the elastic wave propagation problem in anisotropic cylindrical plates in the circumferential direction. Accuracy of the technique has been verified by comparing the computed results for isotropic pipes with the published results. Since no published results are available for wave propagation in the circumferential direction in anisotropic cylindrical plates, the computed dispersion curves for anisotropic curved plates could not be compared with any results available in the literature. However, the Lamb wave dispersion curves for flat plates can be computed and those values are used to check the accuracy of the proposed technique. With the new technique, dispersion curves for cylindrical plates with large radius of curvature (outer radius of curvature to thickness ratio equal to 1000) have been computed and compared with the flat-plate results for both isotropic and anisotropic materials. Computed results for such low curvature plates matched very well with the flat-plate results. The effect on the dispersion curves as the curvature of the anisotropic plate increases has been also studied.

The solution technique used for this specific wave propagation

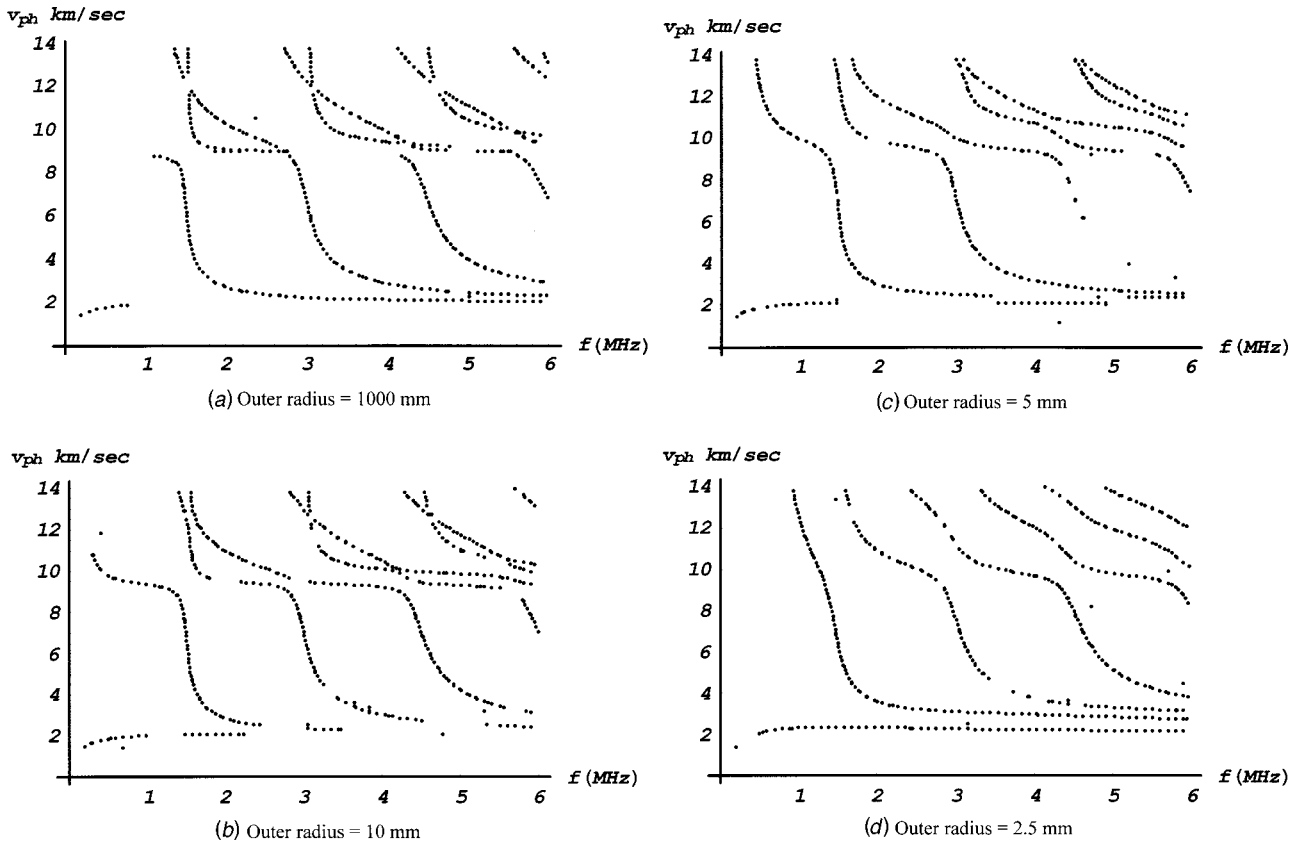


Fig. 15 Dispersion curves for circumferential direction wave propagation in fiber-reinforced cylindrical composite plates when fibers are oriented in the circumferential direction, outer radius of the pipe is (*a*) 1000 mm, (*b*) 10 mm, (*c*) 5 mm, and (*d*) 2.5 mm. Pipe wall thickness and material properties are same as those in Fig. 7.

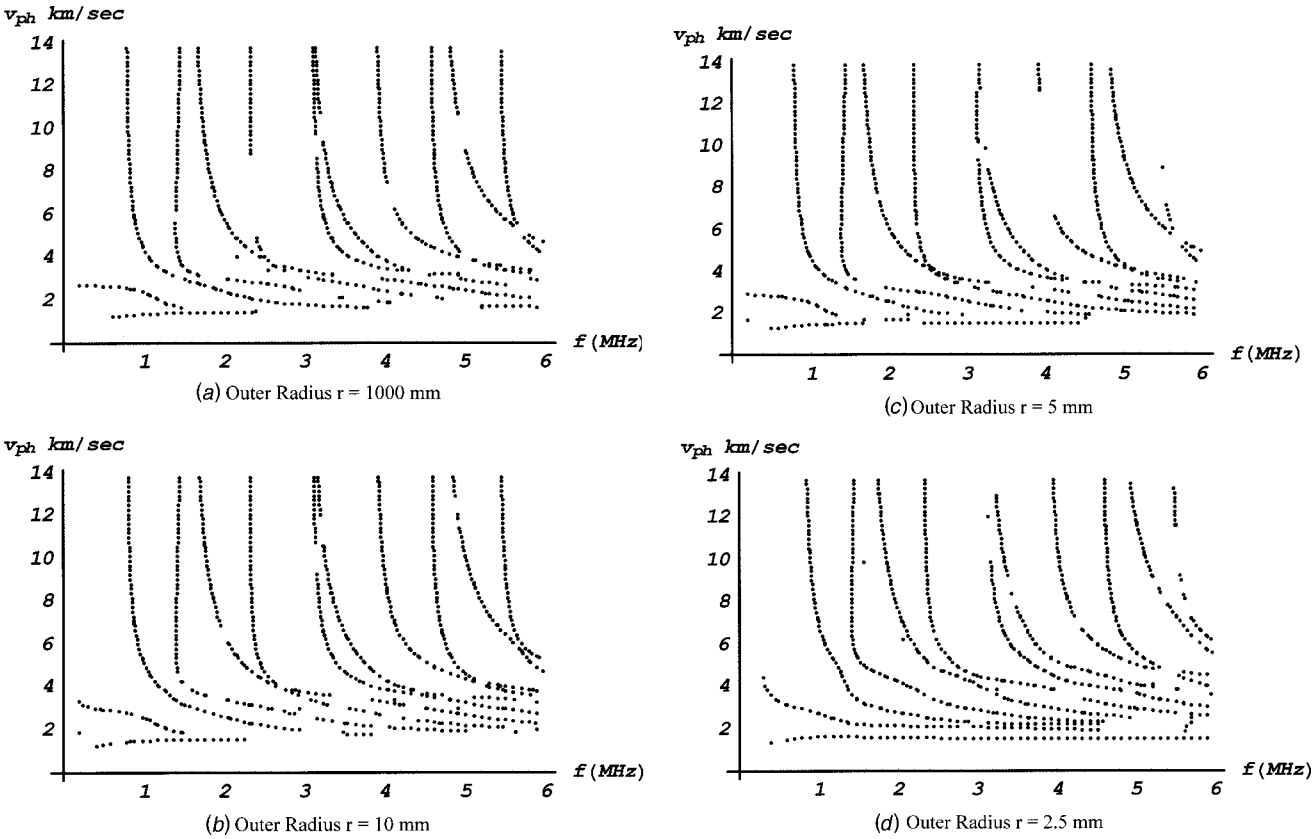


Fig. 16 Dispersion curves for circumferential direction wave propagation in fiber-reinforced composite cylindrical plates when fibers are oriented in the axial direction, outer radius of the pipe is (*a*) 1000 mm, (*b*) 10 mm, (*c*) 5 mm, and (*d*) 2.5 mm. Pipe wall thickness and material properties are same as those in Fig. 9.

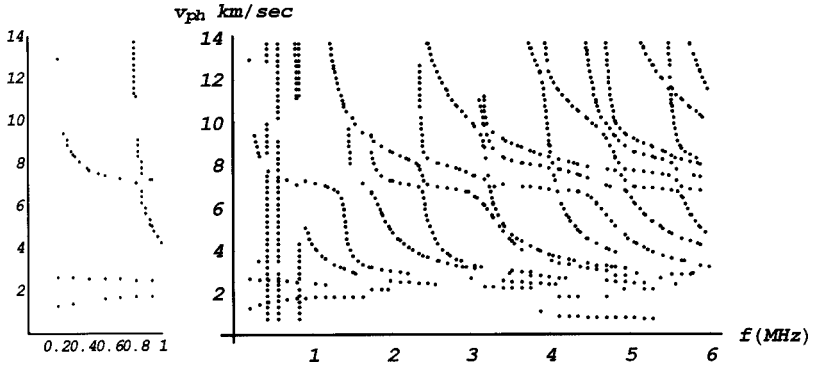


Fig. 17 Dispersion curves for the curved plate when the fibers are oriented in the 45 deg direction. Material properties are given in Eq. (14). Outer radius is 5 mm. Thickness=1 mm. Right figure is for $m=25$, and the left figure is for $m=35$. Frequency range for the left figure is 0 to 1 MHz and for the right figure it is 0 to 6 MHz.

problem is a newly developed general solution technique for solving a coupled partial differential equation set ([14]). Applicability of this technique to other wave propagation problems is currently under investigation.

Acknowledgment

This research was partially supported from the National Science Foundation grant CMS-9901221.

References

[1] Viktorov, I. A., 1958, ''Rayleigh-Type Waves on a Cylindrical Surface,'' Sov. Phys. Acoust., **4**, pp. 131–136.

[2] Qu. J., Berthelot, Y., and Li, Z., 1996, ''Dispersion of Guided Circumferential Waves in a Circular Annulus,'' *Review of Progress in Quantitative Nondestructive Evaluation*, D. O. Thompson and D. E. Chimenti, eds., Plenum, New York, **15**, pp. 169–176.

[3] Grace, O. D., and Goodman, R. R., 1966, ''Circumferential Waves on Solid Cyliners,'' J. Acoust. Soc. Am., **39**, pp. 173–174.

[4] Brekhovskikh, L. M., 1968, ''Surface Waves Confined to the Curvature of the Boundary in Solid,'' Sov. Phys. Acoust., **13**, pp. 462–472.

[5] Cerv, J., 1988, ''Dispersion of Elastic Waves and Rayleigh-Type Waves in a Thin Disc,'' Acta Tech. CSAV, **89**, pp. 89–99.
[6] Liu, G., and Qu, J., 1998, ''Guided Circumferential Waves in a Circular Annulus,'' ASME J. Appl. Mech., **65**, pp. 424–430.
[7] Liu, G., and Qu, J., 1998, ''Transient Wave Propagation in a Circular Annulus Subjected to Impulse Excitation on Its Outer Surface,'' J. Acoust. Soc. Am., **103**, pp. 1210–1220.
[8] Valle, C., Qu, J., and Jacobs, L. J., 1999, ''Guided Circumferential Waves in Layered Cylinders,'' Int. J. Eng. Sci., **37**, pp. 1369–1387.
[9] Nayfeh, A. H., 1995, *Wave Propagation in Layered Anisotropic Media With Application to Composites*, Elsevier, Amsterdam.
[10] Armenakas, A. E., and Reitz, E. S., 1973, ''Propagation of Harmonic Waves in Orthotropic, Circular Cylindrical Shell,'' ASME J. Appl. Mech., **40**, pp. 168–174.
[11] Mal, A. K., and Singh, S. J., 1991, *Deformation of Elastic Solids*, Prentice-Hall, Englewood Cliffs, NJ, p. 313.
[12] Karim, M. R., Mal, A. K., and Bar-Cohen, Y., 1990, ''Inversion of Leaky Lamb Wave Data by Simplex Algorithm,'' J. Acoust. Soc. Am., **88**, pp. 482–491.
[13] Rose, J. L., 1999, *Ultrasonic Waves in Solid Media*, Cambridge University Press, Cambridge, U.K., pp. 264–271.
[14] Towfighi, S., 2001, ''Elastic Wave Propagation in Circumferential Direction in Anisotropic Pipes,'' Ph.D. dissertation, The University of Arizona, Tucson, AZ.

The Effect of Debonding Angle on the Reduction of Effective Moduli of Particle and Fiber-Reinforced Composites

Y. H. Zhao[1]

G. J. Weng[2]
Fellow ASME
e-mail: weng@jove.rutgers.edu

Department of Mechanical and Aerospace Engineering,
Rutgers University,
New Brunswick, NJ 08903

In an effort to uncover the effect of interfacial partial debonding on the reduction of composite stiffness, a reduced moduli approach is proposed for the fictitious inclusions which are used to replace the original partially debonded inclusions. The fictitious inclusions are now perfectly bonded to the matrix and any micromechanical theory can be called upon to estimate the moduli of the composite. Using the volume of the inclusion directly beneath the interfacial cracks under the considered loading mode as a measure of damage, a set of anisotropic damage parameters is established in terms of the debonding angle, providing the reduced moduli for the fictitious inclusions. Specific considerations include debonding on the top and bottom of spheres and prolate inclusions, debonding on the lateral surface of spheres and oblate inclusions, and debonding on the top and bottom of circular fibers and elliptic cylinders. The reductions of the five transversely isotropic moduli for the partially debonded particle composites and the nine orthotropic moduli for the partially debonded fiber composites are examined as the debonding angle increases. The theory is also compared with some finite element results, and it suggests that the concept proposed to estimate the reduced moduli of the fictitious inclusions is a viable one. [DOI: 10.1115/1.1459068]

1 Introduction

The effective moduli of a partially debonded composite are known to be weaker than those of a perfectly bonded composite, but the extent of reduction depends on the debonding geometry and its determination is generally not a trivial problem. In an attempt to shed some light on the effect of partial debonding, Zhao and Weng [1] considered two kinds of debonding geometries: the first one involves debonding on the top and bottom of oblate inclusions and the second involves debonding on the lateral surface of the prolate inclusions so that, in each case, the overall property remains transversely isotropic. The approach was based on the concept of a fictitious inclusion whose property was determined from those of the original inclusion but with the additional assumption that the load-transfer ability of the inclusion in the debonded direction is lost. With this concept and with the help of Willis [2] and Mori-Tanaka [3] moduli for a perfectly bonded composite containing aligned ellipsoidal inclusions, the derived moduli for the partially debonded composite can still be cast in a simple, explicit form. The results reflect the significant loss of the overall moduli in the debonded direction, but not in the transverse direction, at all levels of inclusion shapes and concentrations.

What remained unclear was the effect of interfacial debonding angle on the reduction of the moduli. In order to answer this question, Zheng et al. [4] recently carried out a finite element investigation on the problem of a two-phase composite containing aligned fibers, and observed a continuous reduction of the effective modulus E_{33} in the debonding direction as the debonding angle ϕ increases. The nature of this reduction exhibits a deflection point at a sufficiently large ϕ, that is, in the E_{33} versus ϕ plot, the rate of reduction (slope) was initially small and then becomes large, and later turns to small again. Moreover, the ϕ angle at which such a deflection occurs also tends to decrease with increasing fiber volume concentration. They concluded that the debonding angle could play a key role in affecting the overall E_{33}.

The present study was in part motivated by this finite-element result, and in part by the desire to extend the concept suggested in Zhao and Weng [1] to include the debonding-angle dependence in the estimate of the effective moduli of a partially debonded composite. We shall again start out from the concept of fictitious inclusion, but shall not assume that the load transfer ability of the debonded inclusion in the debonding direction is completely lost while in the transverse direction it remains intact. Instead, partial lost in both directions will be taken, with a magnitude dependent upon the debonding angle and loading direction. The outcome will still be an explicit set of formulas for the effective moduli but now with a debonding-angle dependence.

2 Properties of the Fictitious Inclusion

We shall consider both particle and fiber composites in a rather broad sense, in that the particles may be spherical or aligned spheroidal inclusions and that fibers may be circular or elliptic cylinders. The schematic diagrams depicting the debonding locations (with exaggeration on the surface separation for clarity) for these two classes of inclusions are shown in Fig. 1(*a*)–(*d*) for particles and Fig. 1(*e*)–(*f*) for fibers. Figure 1(*a*) shows the debonding on the top and bottom of a prolate inclusion, and when the aspect ratio α—defined as the length-to-diameter ratio—is one, it turns into (*b*), a spherical inclusion. Figure 1(*c*) indicates the debonding on the lateral surface of an oblate inclusion and when the aspect ratio—the thickness-to-diameter ratio—is one, it turns into (*d*), a sphere. Note that the nature of debonding in (*b*) and (*d*) are different. In Fig. 1(*e*), debonding occurs on the top and bottom (the

[1]On leave from Shenyang Architectural and Civil Engineering Institute, Department of Civil Engineering, Shenyang, Liaoning 110015, P.R. China.

[2]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, September 27, 2000; final revision, November 22, 2001. Associate Editor: J. W. Ju. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

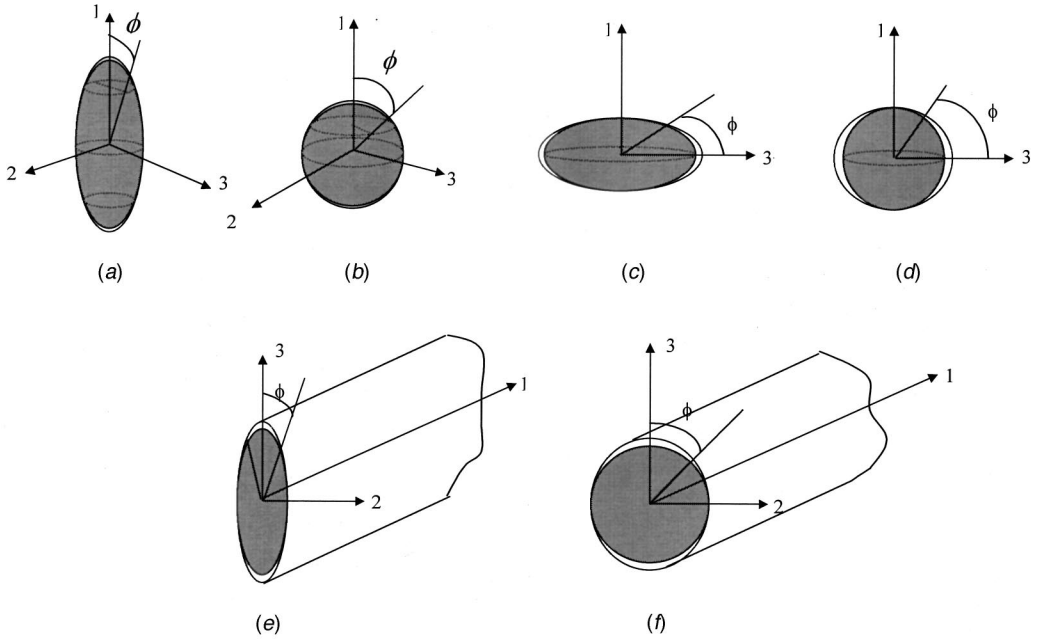


Fig. 1 Schematic diagrams for the interfacial partial debonding: (*a*) debonding on the top and bottom of prolate inclusions, (*b*) debonding on the top and bottom of spherical particles, (*c*) debonding on the lateral surface of oblate inclusions, (*d*) debonding on the lateral surface of spherical particles, (*e*) debonding on the top and bottom of elliptic cylinders, and (*f*) debonding on the top and bottom of circular fibers

narrower sides) of an elliptic fiber, and when the cross-sectional aspect ratio α—defined as the thickness-to-width ratio—is one, it turns into the traditional circular fiber in (f). The specific cases studied in Zhao and Weng [1] were for debonding on the top and bottom of Fig. 1(c) and on the lateral surface of Fig. 1(a); as both types of debonding were on the broad side of the spheroid it was considered sensible to assume that the load transfer ability of the inclusion was completely lost when the debonding angle was sufficiently wide. Cases (b) and (f) here are the types of debonding commonly studied in literature. We note in passing that the elastic field involving *one* interfacial arc crack for a circular fiber in an *infinitely* extended matrix has been derived by England [5] and Toya [6] and that the result has been used by Ju [7] to construct a damage model for a fiber-reinforced composite, but that issues involving *double* debonding in a *finite* matrix as considered here appear not to have been addressed before.

Now debonded particles or fibers of each kind are randomly placed in the matrix but with a fixed orientation, so that the composite as a whole is transversely isotropic in (a)–(d), and orthotropic in (e)–(f). Our objective is to derive a set of explicit formulas for the five or nine effective moduli that could provide some insights into the effect of debonding angle and at the same time remain potentially useful for design or application. The symmetric axis of the spheroidal particles in (a)–(c) is taken to be direction-1, whereas for the fibers in (e)–(f) the debonding direction is designated as direction-3. Angle ϕ represents one-half of the total debonding angle so that when $\phi=0$ the composite is perfectly bonded and when $\phi=90$ deg it is totally debonded. In the two-phase composite the inclusions will be referred to as phase 1 and the matrix as phase 0, with the volume concentration of the rth phase denoted by c_r. For simplicity both phases will be taken to be isotropic, with the Young's modulus E_r and Poisson's ratio ν_r. The concept to be presented, however, applies to anisotropic inclusions.

To determine the reduced moduli of the fictitious inclusion that is to be used to replace the debonded inclusion but now becomes perfectly bonded to the matrix, we shall account for the partial loss of load-transfer ability due to partial debonding in terms of the volume fraction of the inclusion directly between the debonding surfaces in accordance with the loading direction. An anisotropic damage parameter D_i will be used to denote this effect for loading direction-*i*. Out of the six figures in Fig. 1, there are three distinct cases: (a), (c), and (e). We now analyze these cases separately.

2.1a Aligned Prolate Inclusions With Debonding on the Top and Bottom Surfaces. For Fig. 1(a), the Young's modulus of the perfectly bonded fictitious inclusion in direction-1 is calculated on the basis of an effective stress σ_{11}^c such that $\sigma_{11}^e = \sigma_{11}/(1-D_1)$ under a pure tensile loading σ_{11}. This leads to a reduced Young's modulus

$$E_{11}^{(1)} = E_1(1-D_1), \tag{1}$$

in the debonding direction, where the superscript 1 signifies that it refers to the perfectly bonded fictitious inclusion. The damage parameter D_1 is calculated according to

$$D_1 = V_{d1}/V, \tag{2}$$

where V_{d1} is the volume of the "damaged" part of the inclusion directly between the two debonding surfaces, and V the volume of the inclusion. After some elementary calculus one finds that

$$D_1 = 1 - \left(\frac{\cos^2 \phi}{\alpha^2 \sin^2 \phi + \cos^2 \phi} \right)^{3/2}, \tag{3}$$

in terms of the debonding angle ϕ and aspect ratio α. When the prolate inclusion turns into a sphere as in Fig. 1(b), this damage parameter is simply

$$D_1 = 1 - \cos^3 \phi. \tag{4}$$

Due to symmetry the damage parameter D_2 and D_3 for tensile loading in the 2 and 3 directions are identical and their value can be evaluated as

$$D_2 = V_{d2}/V, \tag{5}$$

where V_{d2} is now the volume of the damaged part of the inclusion associated with the direction-2 loading; it is also the top and bottom parts of V_{d1}. It follows that

$$D_2 = 1 - \frac{\cos\phi(3\alpha^2 \sin^2\phi + 2\cos^2\phi)}{2(\alpha^2 \sin^2\phi + \cos^2\phi)^{3/2}}. \tag{6}$$

When the inclusions are spherical, it turns into

$$D_2 = 1 - \cos\phi\left(1 + \frac{1}{2}\sin^2\phi\right). \tag{7}$$

The Young's modulus of the fictitious inclusion in the transverse direction is taken to be

$$E_{22}^{(1)} = E_{33}^{(1)} = E_1(1 - D_2). \tag{8}$$

Following similar consideration for other loading directions, the compliance matrix of the fictitious inclusions can be written as

$$M_1 = \begin{bmatrix} \frac{1}{E_1(1-D_1)} & -\frac{\nu_1}{E_1} & -\frac{\nu_1}{E_1} & 0 & 0 & 0 \\ -\frac{\nu_1}{E_1} & \frac{1}{E_1(1-D_2)} & -\frac{\nu_1}{E_1(1-D_2)} & 0 & 0 & 0 \\ -\frac{\nu_1}{E_1} & -\frac{\nu_1}{E_1(1-D_2)} & \frac{1}{E_1(1-D_2)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_2)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_1)} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_1)} \end{bmatrix}. \tag{9}$$

2.1b Aligned Oblate Inclusions With Debonding on the Lateral Surface. When debonding occurs on the lateral surface of the oblate inclusion as depicted in Fig. 1(*c*), the compliance matrix of the fictitious inclusion may be written as

$$M_1 = \begin{bmatrix} \frac{1}{E_1(1-D_1')} & -\frac{\nu_1}{E_1} & -\frac{\nu_1}{E_1} & 0 & 0 & 0 \\ -\frac{\nu_1}{E_1} & \frac{1}{E_1(1-D_2')} & -\frac{\nu_1}{E_1(1-D_2')} & 0 & 0 & 0 \\ -\frac{\nu_1}{E_1} & -\frac{\nu_1}{E_1(1-D_2')} & \frac{1}{E_1(1-D_2')} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_2')} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_2')} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_2')} \end{bmatrix}, \tag{10}$$

where in terms of V_{d2}, V_{d1}, D_2, and D_1 of the previous case, the damage parameters are

$$D_2' = V_{d2}'/V = (V - V_{d2})/V = 1 - D_2,$$
$$D_1' = V_{d1}'/V = (V - V_{d1})/V = 1 - D_1. \tag{11}$$

Thus in terms of the current debonding angle ϕ which forms a 90 deg conjugate with the previous debonding angle

$$D_2' = \frac{\sin\phi(3\alpha^2\cos^2\phi + 2\sin^2\phi)}{2(\alpha^2\cos^2\phi + \sin^2\phi)^{3/2}},$$
$$D_1' = \left(\frac{\sin^2\phi}{\alpha^2\cos^2\phi + \sin^2\phi}\right)^{3/2}. \tag{12}$$

For spherical particles as depicted in Fig. 1(*d*), these damage parameters turn into

$$D_2' = \sin\phi\left(1+\frac{1}{2}\cos^2\phi\right), \quad D_1' = \sin^3\phi. \tag{11}$$

2.2 Aligned Elliptic Cylinders With Debonding on the Top and Bottom Surfaces. The anisotropic damage parameters in this case are measured by the area directly beneath the debonded portions according to the loading mode. For loading along direction-3, the debonding direction, direct integration of the area for the elliptic surface yields

$$D_3 = \frac{2}{\pi}\left[\frac{\alpha\sin\phi\cos\phi}{\alpha^2\sin^2\phi+\cos^2\phi} + \sin^{-1}\left(\frac{\alpha\sin\phi}{(\alpha^2\sin^2\phi+\cos^2\phi)^{1/2}}\right)\right]. \tag{12}$$

For circular fibers ($\alpha=1$) it is simply

$$D_3 = \frac{2}{\pi}(\sin\phi\cos\phi+\phi). \tag{13}$$

For direction-2 along the thickness, it is

$$D_2 = \frac{2}{\pi}\frac{\alpha\sin\phi}{(\alpha^2\sin^2\phi+\cos^2\phi)^{1/2}}\left[\sin^{-1}\left(\frac{\alpha\sin\phi}{(\alpha^2\sin^2\phi+\cos^2\phi)^{1/2}}\right) - \frac{\cos\phi}{(\alpha^2\sin^2\phi+\cos^2\phi)^{1/2}}\right]. \tag{14}$$

With circular fibers it has the simple expression

$$D_2 = \frac{2}{\pi}\sin\phi(\phi-\cos\phi). \tag{15}$$

The compliance matrix of the fictitious elliptic fiber takes the form

$$M_1 = \begin{bmatrix} \frac{1}{E_1} & -\frac{\nu_1}{E_1} & -\frac{\nu_1}{E_1} & 0 & 0 & 0 \\ -\frac{\nu_1}{E_1} & \frac{1}{E_1(1-D_2)} & -\frac{\nu_1}{E_1} & 0 & 0 & 0 \\ -\frac{\nu_1}{E_1} & -\frac{\nu_1}{E_1} & \frac{1}{E_1(1-D_3)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_3)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_3)} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1+\nu_1}{E_1(1-D_2)} \end{bmatrix}. \tag{16}$$

3 Effective Moduli of the Partially Debonded Composites

Once the elastic compliances of the fictitious inclusions are established, one may invoke any micromechanical theories for perfectly bonded composites to evaluate the effective moduli. For the problem with aligned ellipsoidal inclusions the simplest one is likely to be Willis' [2] approach, which gives an identical result to the Mori-Tanaka [3] approach for the aligned case. This model inherently takes the distribution function of the inclusions to be identical to the inclusion shape (see Weng [8] for proof). The established compliances M_1 turn into the moduli L_1 for the fictitious inclusions with $L_1 = M_1^{-1}$. The effective moduli tensor of the debonded composite then follows as

$$L = (c_1L_1A_1 + c_0L_0)(c_1A_1 + c_0I)^{-1}, \tag{17}$$

where I is the fourth-order identity tensor, and A_1 the strain concentration tensor of a single fictitious inclusion embedded in the infinitely extended matrix. In terms of Eshelby's [9] S-tensor for an ellipsoidal inclusion, it is given by

$$A_1 = [I + SL_0^{-1}(L_1 - L_0)]^{-1}. \tag{18}$$

The components of the S-tensor for a spheroidal inclusion and an elliptic cylinder can be found in Mura [10].

We now examine the results based on the proposed approach. Numerical calculations were performed for the silicon-carbide/aluminum system, with the properties ([11,12]):

$$\text{Silicon carbide: } E_1 = 490\text{ GPa}, \quad \nu_1 = 0.17,$$

$$\text{Aluminum: } E_0 = 68.3\text{ GPa}, \quad \nu_0 = 0.33. \tag{19}$$

3.1a Aligned Prolate Inclusions With Debonding on the Top and Bottom Surfaces. With spherical particles ($\alpha=1$) the five transversely isotropic moduli of the composite whose partial debonding is depicted in Fig. 1(*b*) are shown in Fig. 2. Here E_{11}, E_{22}, ν_{12}, μ_{12}, and μ_{23} are, respectively, the longitudinal Young's modulus, transverse Young's modulus, major Poisson's ratio, axial shear modulus, and transverse shear modulus of the partially debonded composite. The plots are displayed as a function of debonding angle ϕ at three selected particle concentrations: 0.1, 0.2, and 0.3. We did not go beyond 0.3 as the Willis-Mori-Tanaka formulas are essentially a low concentration theory. A quick glance over these five figures indicates that E_{11}, ν_{12} and μ_{12} are the moduli that are sensitive to the debonding angle. Even for these three moduli there appears to have an incubation period before the debonding angle begins to show its effect, that is, for spherical particles as shown here, when the debonding angle is small, the reduction of the longitudinal Young's modulus does not

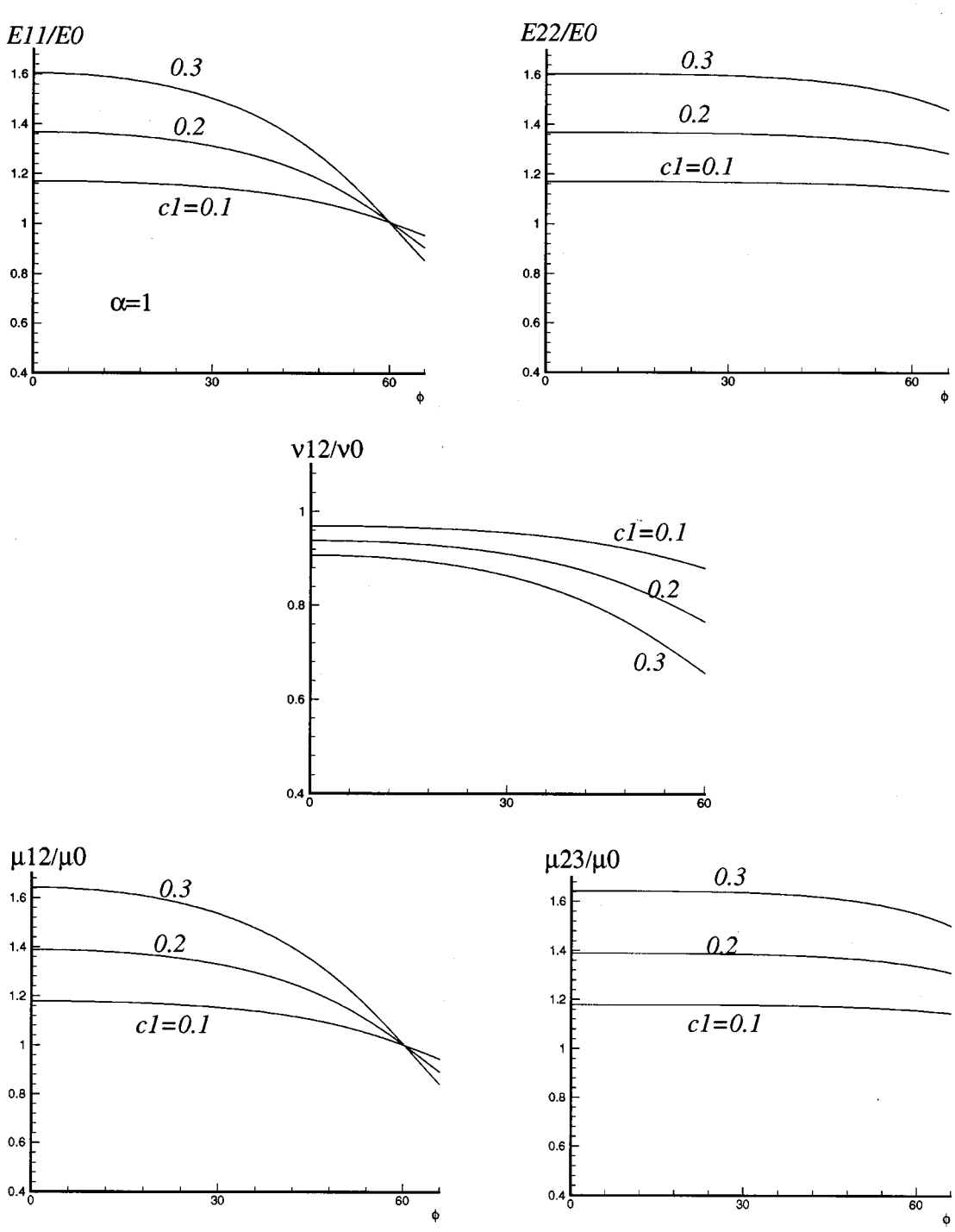


Fig. 2 The influence of debonding angle on the five effective moduli of the composite whose spherical particles debonded on the top and bottom

begin to show up until about 30 deg at $c_1=0.1$ and about 15 deg at $c_1=0.3$. But further increase in the debonding angle will lead to a visible reduction. At $\phi=60$ deg, the reduction of E_{11} with $c_1=0.1$ is about 15% and with $c_1=0.3$ it is about 38%. There is also a cross over around this ϕ as a higher concentration of debonded particles becomes more detrimental to the strengthening effect. Similar features are also observed for μ_{12}, but the crossover event is not a feature of the major Poisson's ratio ν_{12} which exhibits a reversed trend in terms of the particle concentration.

When the prolate inclusions take another shapes, the reductions of the five moduli are shown in Fig. 3, at $c_1=0.1$. The shape-dependence for E_{11} and ν_{12} are seen to be not monotonic. While there is a significant incubation period for $\alpha=1$, the drop of the modulus for $\alpha=10$ is almost immediate. This is due to the fact that, at this α, the major body of the inclusion with a large debonding angle is directly under the cracks and it is not functioning as a fully effective medium to carry the load. Even though under the perfectly bonded condition ($\phi=0$) its E_{11} was greater than

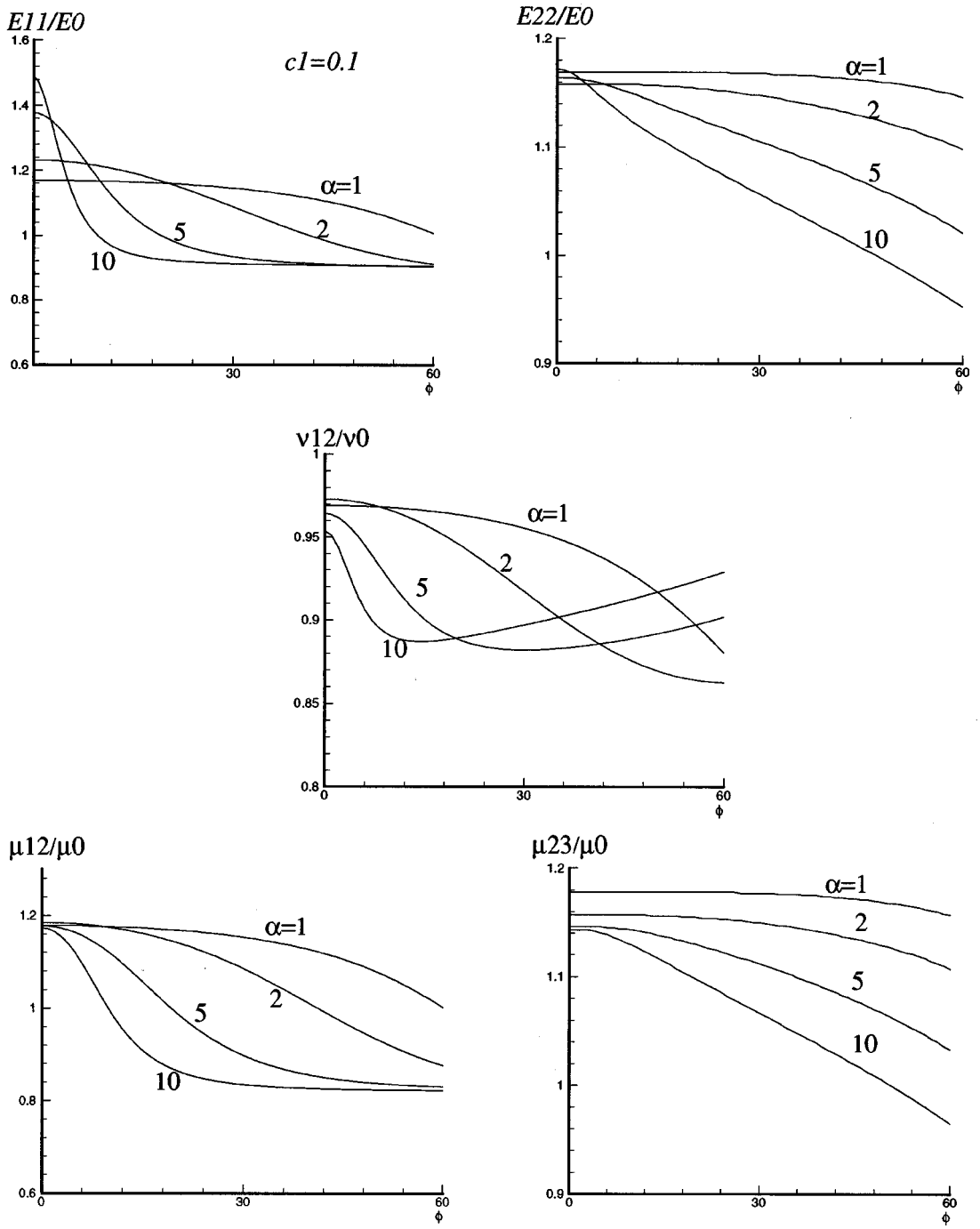


Fig. 3 The influence of debonding angle on the five effective moduli of the composite whose prolate inclusions debonded on the top and bottom

that of $\alpha = 1$, it quickly lost the ground and rendered the composite weak when the debonding angle increases. This is the reason for the crossover effect in E_{11}. The existence of a deflection point as mentioned in the Introduction—though not evident in $\alpha = 1$ up to 60 deg—is now visible in every curve. In the transverse direction, such as E_{22} and μ_{23}, the reduction of the moduli also becomes significant as α increases, so unlike the spherical case the moduli reduction in the transverse direction is generally not negligible for prolate particles as the interfacial cracks will spread quickly to the surface normal to the 2-3 directions.

3.1b Aligned Oblate Inclusions With Debonding on the Lateral Surface. When debonding occurs on the lateral surface of the spherical particles as depicted in Fig. 1(*d*), the angle-dependence of the reduction for the five moduli is displayed in Fig. 4. In this case the transverse properties—E_{22} and μ_{23}—are the figures of merit. The nature of the reduction in E_{22} is to be compared with that of E_{11} in Fig. 2, and μ_{23} with μ_{12}. It is found that the incubation period—even for the low concentration $c_1 = 0.1$—is shorter in this case. The reduction in both E_{22} and μ_{23} is almost immediate as the debonding angle increases, leading to a

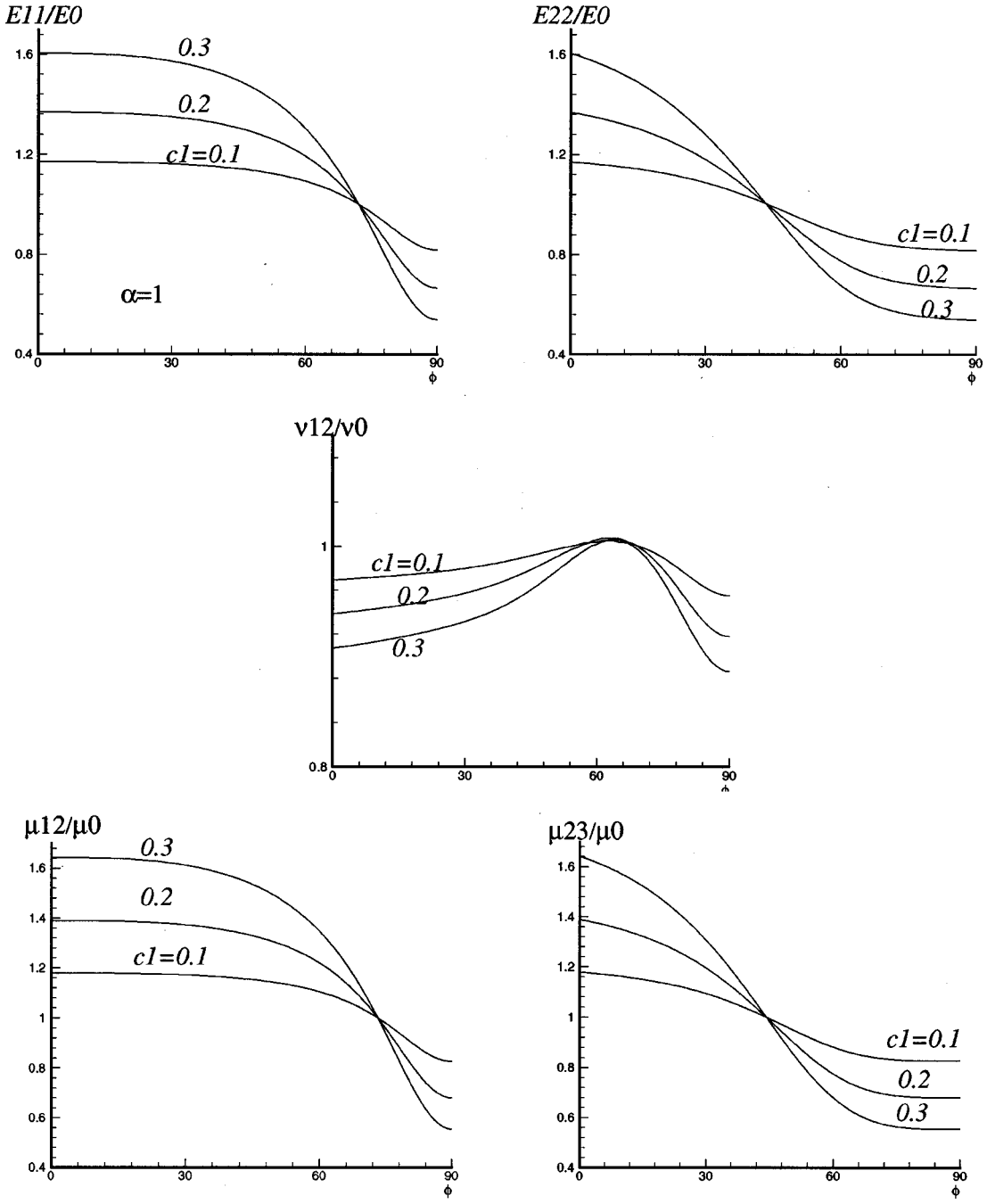


Fig. 4 The influence of debonding angle on the five effective moduli of the composite whose spherical particles debonded on the lateral surface

crossover. The off-axis moduli E_{11} and μ_{12} in this case also show a clear dependence on ϕ following an incubation period, a phenomenon not present in E_{22} and μ_{23} of Fig. 2. The crossover of the moduli seems to occur at a fixed debonding angle for each modulus, regardless of the particle concentration. Such an angle is smaller for the two in-plane moduli (around 45 deg), but greater for the two axial moduli E_{11} and μ_{12} (about 70 deg).

The shape-dependence of the reduction with the lateral interfacial debonding is shown in Fig. 5, for $c_1=0.1$. As the inclusions take the more oblate shape, say with $\alpha=0.1$, the reduction in E_{22} and μ_{23} is quite drastic. Partial debonding in this case quickly spreads to the flat parts of the inclusion on the top and bottom, rendering the entire inclusion almost useless. The asymptotic parts of the moduli in E_{22} and μ_{23} are essentially those of a porous material. Due to such a sharp reduction at a small α, there is a crossover effect in these two moduli. While no crossover is observed for E_{11} and μ_{12}, significant influence by ϕ is evident when the inclusion is flat.

3.2 Aligned Elliptic Cylinders With Debonding on the Top and Bottom Surfaces. With circular fibers as depicted in Fig. 1(f), the reduction of the nine orthotropic moduli as a function of debonding angle ϕ is shown in Fig. 6. Here, as in the previous case, Poisson's ratio ν_{ij} is defined as the ratio of strain shrinkage in the j-direction due to a tensile stress in the i-direction. Keeping in mind that fibers are aligned in direction-1 and debonding is along direction 3, it is clear that E_{11}, μ_{12}, and ν_{12} are not affected

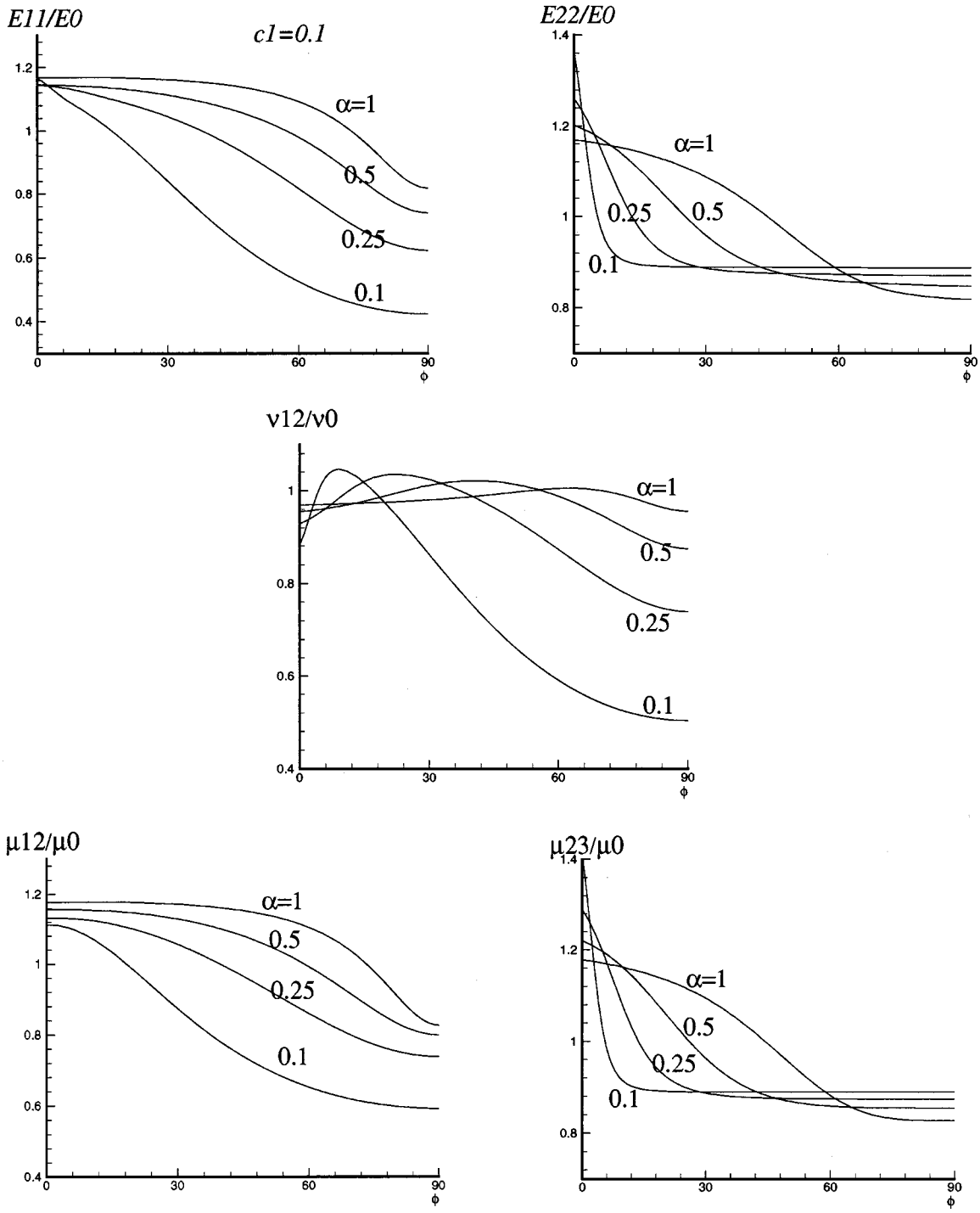


Fig. 5 The influence of debonding angle on the five effective moduli of the composite whose oblate inclusions debonded on the lateral surface

by ϕ. But the figures of merit in this problem are E_{33} and μ_{23}, and it is found that their reductions are faster than their counter parts in the spherical case (E_{11} and μ_{12} in Fig. 2), for more volume is affected by the interface cracks in a cylinder. Both E_{33} and μ_{23} show a cross over at about 50 deg, in contrast to the 60 deg for spherical particles.

The shape-dependence of the moduli reduction for the fiber composite is displayed in Fig. 7, at $c_1=0.1$. As expected, E_{11} is still not affected by ϕ and is virtually independent of the cross-sectional shape of the fibers. But the two previous insensitive moduli—μ_{12} and ν_{12}—are now exhibiting a strong dependence on the debonding angle especially when the elliptic cylinder takes the form of a thin ribbon ($\alpha=0.1$). The two major moduli—E_{33} and μ_{23}—clearly are very sensitive to the debonding angle as the cross-sectional aspect ratio diminishes. The counterparts of these nine moduli for a perfectly bonded composite can be found in [13].

Finally, in order to provide some perspective on the quantitative

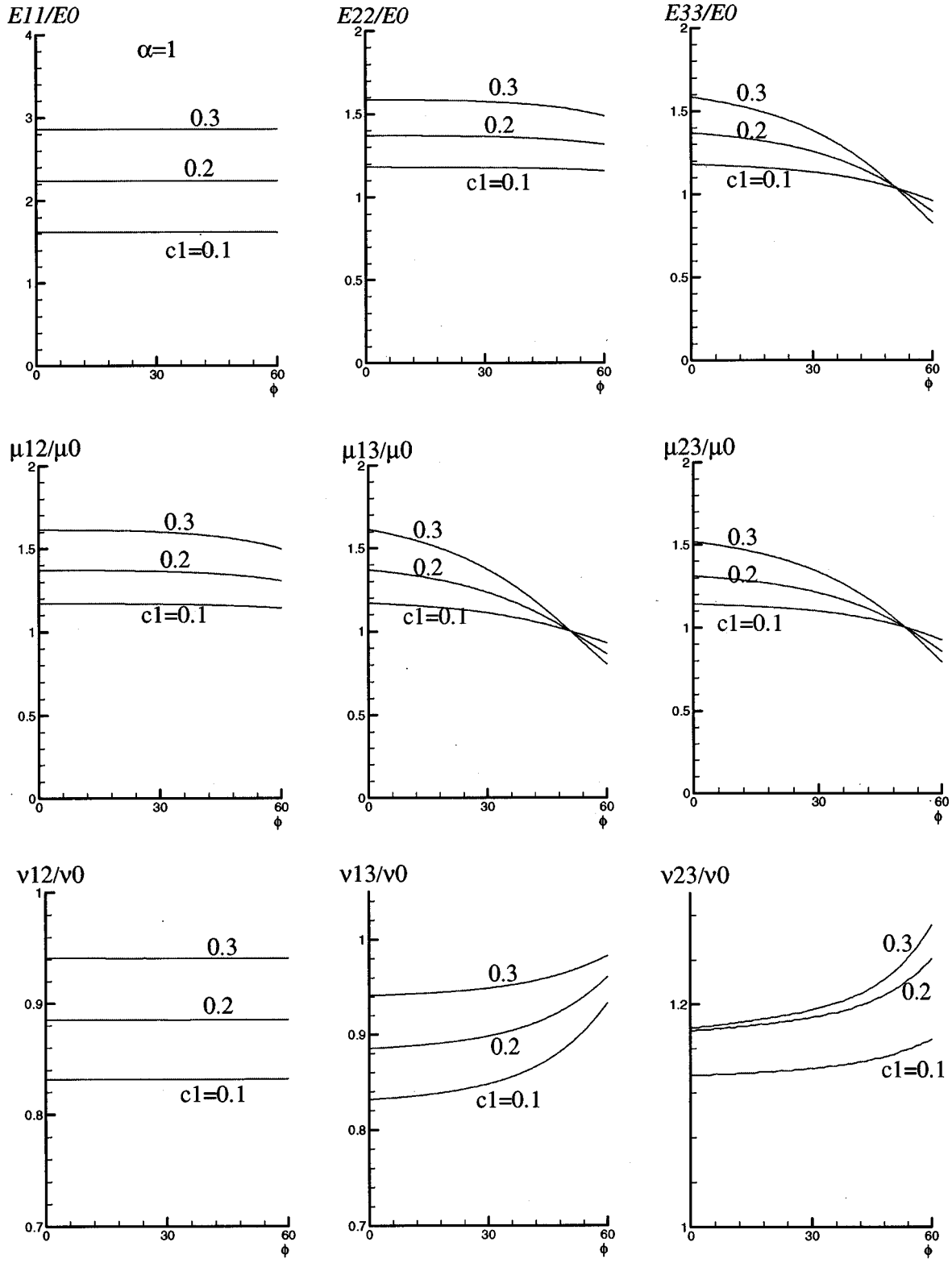


Fig. 6 The influence of debonding angle on the nine effective moduli of the composite whose circular fibers debonded on the top and bottom

accuracy of the proposed concept, a comparison for the case of circular fiber is made with the finite element results of Zheng et al. [4]. This comparison is given in Fig. 8 for E_{33} at $c_1 = 0.1$ and 0.3. The finite element calculations also made use of the same silicon carbide/aluminum system, and the fiber debonding is also on the top and bottom as depicted in Fig. 1(*f*). Their results, using a traction-prescribed boundary condition, are reproduced in Fig. 8 as dashed curves, and the results from the present development are given as solid lines. The finite element results also exhibit certain incubation period and, at $c_1 = 0.3$, develop a deflection

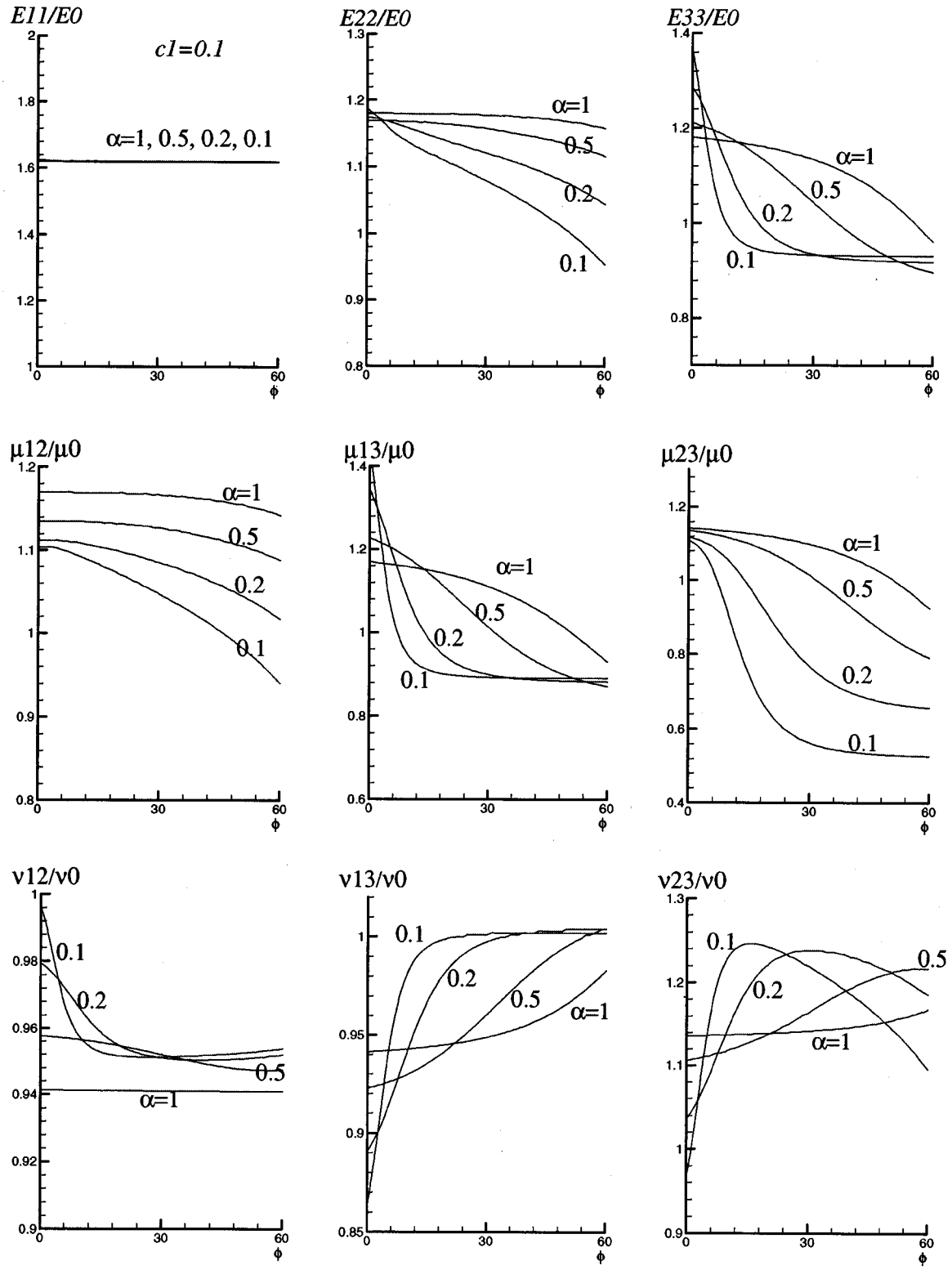


Fig. 7 The influence of debonding angle on the nine effective moduli of the composite whose elliptic cylinders debonded on the top and bottom

point in the curve at around $\phi = 35$ deg. Comparison between these two sets of curves indicates that the theory agrees well with the finite element result at $c_1 = 0.1$, but not so well at $c_1 = 0.3$. As the effective-medium theory of Willis-Mori-Tanaka is accurate at low concentration, any error in $c_1 = 0.1$ is a direct result of the new concept proposed in Section 2. Such an error—as shown in Fig. 8—is small, suggesting that using the volume between the interface cracks as a measure of damage parameter is a reasonable approach. The significant departure between the two at a large ϕ for $c_1 = 0.3$ may be attributed to two factors: (i) the boundary-traction approach adopted in the finite element tends to provide a lower-bound value, and (ii) the Willis-Mori-Tanaka approach is

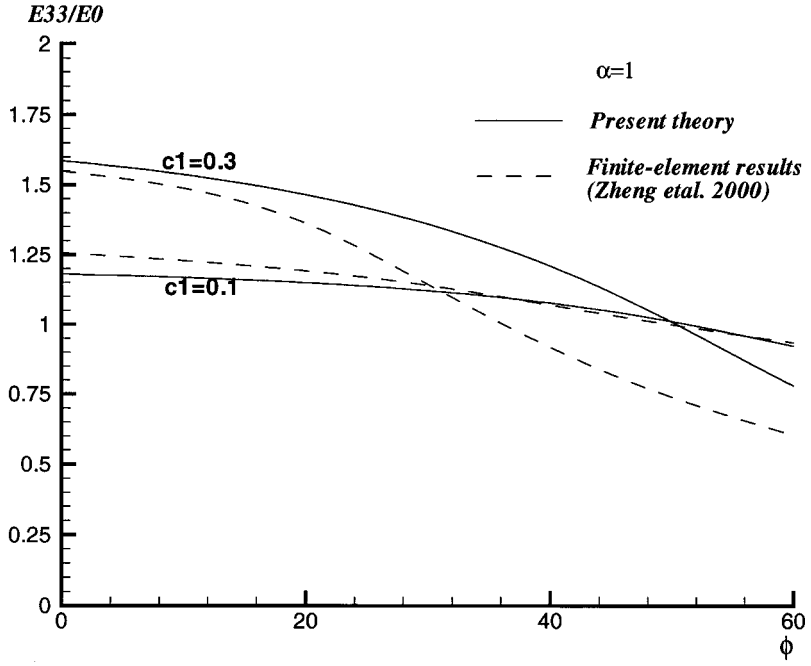


Fig. 8 Comparison between the proposed theory and the finite element result

also a lower-bound one before debonding but can turn into an upper-bound one when the debonding angle is wide (the inclusions would become softer than the matrix). As no calculation was made using the displacement-boundary condition in finite element and no high-concentration effective medium theory is known to exist for the present problem, the precise cause of the discrepancy at $c_1 = 0.3$ remains somewhat difficult to ascertain.

4 Concluding Remarks

In order to account for the effect of debonding angle on the overall elastic moduli of a partially debonded composite, we have proposed to use the volume of the inclusion directly beneath the interface cracks as a measure of damage parameter. This volume not only depends on the the debonding angle ϕ, but also on the loading direction. As such, a set of anisotropic damage parameters have been constructed. This in turn provides the properties for the fictitious inclusions which are used to replace the debonded inclusions but now are perfectly bonded to the matrix with reduced elastic moduli. Debonding on the top and bottom of prolate inclusions, on the lateral surface of oblate inclusions, and on the top and bottom surfaces of elliptic cylinders, have been considered.

With these reduced moduli for the fictitious inclusions, the overall elastic moduli of the partially debonded composite then can be determined using any micromechanical theories. For simplicity we have used the Willis-Mori-Tanaka theory to compute the effective moduli. The results for the five transversely isotropic moduli of two kinds of particulate composites and the nine orthotropic moduli for the fibrous composites are given. A comparison with the finite element result for the fiber case suggests that the concept proposed for the determination of the reduced moduli of the fictitious inclusions could be a viable one.

Acknowledgment

This work was supported in part by the China Scholarship Council and in part by the National Science Foundation, under Grants CMS-9625304 and CMS-0114801.

References

[1] Zhao, Y. H., and Weng, G. J., 1997, ''Transversely Isotropic Moduli of Two Partially Debonded Composites,'' Int. J. Solids Struct., **34**, pp. 493–507.
[2] Willis, J. R., 1977, ''Bounds and Self-Consistent Estimates for the Overall Properties of Anisotropic Composites,'' J. Mech. Phys. Solids, **25**, pp. 185–202.
[3] Mori, T., and Tanaka, K., 1973, ''Average Stress in the Matrix and Average Elastic Energy of Materials With Misfitting Inclusions,'' Acta Metall., **21**, pp. 571–574.
[4] Zheng, S. F., Denda, M., and Weng, G. J., 2000, ''Interfacial Partial Debonding and Its Influence on the Elasticity of a Two-Phase Fiber-Reinforced Composite,'' Mech. Mater., pp. 695–709.
[5] England, A. H., 1966, ''An Arc Crack Around a Circular Elastic Inclusion,'' ASME J. Appl. Mech., **33**, pp. 637–640.
[6] Toya, M., 1974, ''A Crack Along the Interface of a Circular Inclusion Embedded in an Infinite Solid,'' J. Mech. Phys. Solids, **22**, pp. 325–348.
[7] Ju, J. W., 1991, ''A Micromechanical Damage Model for Uniaxially Reinforced Composites Weakened by Interfacial Arc Microcracks,'' ASME J. Appl. Mech., **58**, pp. 923–930.
[8] Weng, G. J., 1992, ''Explicit Evaluation of Willis' Bounds With Ellipsoidal Inclusions,'' Int. J. Eng. Sci., **30**, pp. 83–92.
[9] Eshelby, J. D., 1957, ''The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems,'' Proc. R. Soc. London, Ser. A, **241**, pp. 376–396.
[10] Mura, T., 1987, *Micromechanics of Defects in Solids*, 2nd Ed., Martinus Nihoff, Dordrecht.
[11] Arsenault, R. J., 1984, ''The Strengthening of Aluminum Alloy 6061 by Fiber and Platelet Silicon Carbide,'' Mater. Sci. Eng., **64**, pp. 171–181.
[12] Nieh, T. G., and Chellman, D. J., 1984, ''Modulus Measurements in Discontinuous Reinforced Aluminum Composites,'' Scr. Metall., **8**, pp. 925–928.
[13] Zhao, Y. H., and Weng, G. J., 1990, ''Effective Elastic Moduli of Ribbon Reinforced Composite,'' ASME J. Appl. Mech., **57**, pp. 158–167.

Torsional Impact Response of a Penny-Shaped Interface Crack in Bonded Materials With a Graded Material Interlayer

C. Li[1]
Department of Architectural Engineering,
Shijiazhuang Railway Institute,
Shijiazhuang 050043, P. R. China

Z. Duan
Laboratory of Nonlinear Mechanics,
Institute of Mechanics,
Chinese Academy of Sciences,
Beijing 100080, P. R. China

Z. Zou
Department of Communications Engineering,
Shijiazhuang Railway Institute,
Shijiazhuang 050043, P. R. China

In this paper, the dynamic response of a penny-shaped interface crack in bonded dissimilar homogeneous half-spaces is studied. It is assumed that the two materials are bonded together with such a inhomogeneous interlayer that makes the elastic modulus in the direction perpendicular to the crack surface is continuous throughout the space. The crack surfaces are assumed to be subjected to torsional impact loading. Laplace and Hankel integral transforms are applied combining with a dislocation density function to reduce the mixed boundary value problem into a singular integral equation with a generalized Cauchy kernel in Laplace domain. By solving the singular integral equation numerically and using a numerical Laplace inversion technique, the dynamic stress intensity factors are obtained. The influences of material properties and interlayer thickness on the dynamic stress intensity factor are investigated. [DOI: 10.1115/1.1459066]

1 Introduction

Interface crack problems of composite structures have been the important topic of fracture mechanics in recent decades. There are a large number of solutions in the technical literature for isotropic, orthotropic, and anisotropic bonded materials containing interface cracks. Some typical studies that should be mentioned are that the asymptotic analysis of the elastic fields (Williams [1]), the standard interface crack solutions (Erdogan [2], Rice and Sih [3], Willis [4] and Qu and Bassani [5]), the crack-tip contact model (Comninou [6] Achenbach et al. [7] and Rice [8]), the elastic-plastic analysis (Shih and Asaro [9]) and so on. Hutchinson and Suo [10] once gave an extensive overview on the static behavior of interface cracks. On the other hand, there are also a number of papers devoted to the dynamic fracture mechanics of interface cracks. Sih and Chen [11] studied several dynamic responses of composite materials with interface cracks, such as antiplane shear of interface rectangular cracks in layered orthotropic dissimilar materials, orthotropic layered composite debonded over a penny-shaped region subjected to sudden shear, diffraction of time-harmonic waves by interface cracks in dissimilar media. Takei and co-workers [12] and Li and Tai [13] considered the elastodynamic response of a composite with an interface crack under antiplane shear loading. Ueda and co-workers [14] reported the torsional impact response of a penny-shaped crack on a bimaterial interface. Beyond these, considerable experimental works on the dynamics of interface cracks (Lambros and Rosakis [15] and Singh, Lambros, and Rosakis [16]) and numerical simulations of dynamic interfacial crack growth (Xu and Needleman [17] and Needleman and Rosakis [18]) were also carried out. Rosakis and Ravichandran [19] recently made a rather comprehensive review on dynamic failure mechanics.

The researches mentioned above usually assumed that the dissimilar materials were bonded directly (bimaterials) or with a thin homogeneous layer which properties different from that of bonded materials. However, recent studies have indicated that in many cases an inhomogeneous interlayer exists between the bonded materials (Subramanian and Crasto [20]). This kind of interlayer may be developed as a result of certain processing techniques (Lugscheider [21] and Shiau et al. [22]) or results from intentional grading of the material composition (Kurihara et al. [23] and Jager et al. [24]). For the static problems of fracture mechanics about the inhomogeneous interlayer, there have been many theoretical studies (Delale and Erdogan [25], Ozturk and Erdogan [26], Wang et al. [27] and Fildis and Yahsi [28]). In their studies, two kind of inhomogeneous interlayer models have been proposed. One of them is the exponential function model and another is a so-called generalized interlayer model, which is a power function. These models have physical background and make the problem of stress oscillatory singularity (Williams [1]) overcome. However, as for dynamic fracture mechanic of interface cracks, there are few studies considered the effect of an inhomogeneous interlayer.

In this paper, we examine the torsional impact response of a penny-shaped interface crack in a layered composite. Although this problem is rather a theoretical problem, it also has the engineering background, such as the sudden appearance of a penny-shaped interface crack in a component under torsional loading. The main difference between our present paper and literature (Ueda, Shindo, and Astumi [14]) is that a graded material interlayer is introduced. Our main objective is to investigate whether the graded material interlayer is helpful in reducing the dynamic stress intensity factor of an interface crack in a bonded materials and how the material inhomogeneity and interlayer thickness influence the dynamic stress intensity factor. The methods used in our paper are the Laplace and Hankel integral transforms and the singular integral equation technique.

2 Formulation of the Problem

As shown in Fig. 1, consider two dissimilar half-spaces (Material-1 and Material-3) to be bonded with an inhomogeneous interlayer, which denoted as Material-2. The material properties of Material-1 and Material-3 are constant and denoted as ρ_1, μ_1 and ρ_3, μ_3 respectively, where ρ is the mass density and μ is the shear modulus.

As we have known, there are two material parameters involved in the dynamic torsional problems. They are the shear modulus μ

[1]Current address: Department of Mechanical Engineering, University of Delaware, Newark, DE 19716.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, June 23, 1999; final revision, June 22, 2000. Editor: A. Needleman. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

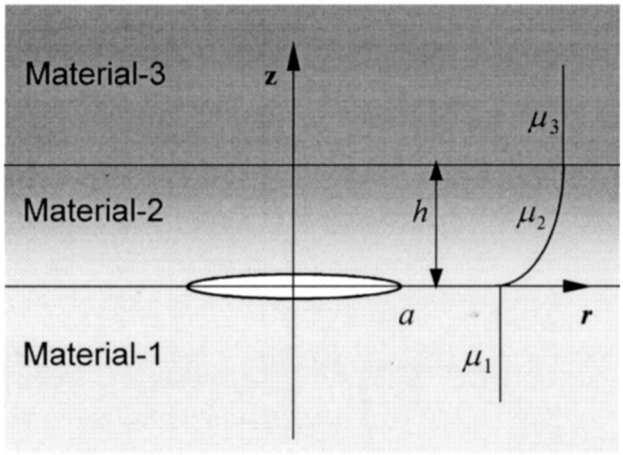


Fig. 1 A penny-shaped crack on the interface of a graded material interlayer and a homogeneous material

and the mass density ρ. For the inhomogeneous interlayer, due to the mathematical complexity introduced by the inertia term, it is necessary to assume that the shear modulus and the mass density can vary independently. Such an idealization can offer considerable simplifications to the analysis. After compared the several models for expressing the variation of the shear modulus, such as the exponential form $\mu(z)=\mu_1\exp(\alpha z)$ (Delale and Erdogan [25]), and the power form $\mu_2(z)=\mu_1(1+\alpha z)^k$ (Wang et al. [27]), we found that the variations

$$\mu_2=\mu_1(1+\alpha z)^2, \tag{1}$$

$$\rho_2=(\rho_1+\rho_3)/2, \tag{2}$$

are mathematically tractable, and still physically representative enough. In Eq. (1), the parameter α can be determined by the continuity condition of the shear modulus $\mu_2(0)=\mu_1$ and $\mu_2(h)=\mu_3$, that is $\alpha=(\sqrt{\mu_3/\mu_1}-1)/h$.

Assume a penny-shaped crack of diameter $2a$ is located at the interface of Material-1 and Material-2 and subjected to a torsional impact loading $P(r)$. For the present problem, in the cylindrical polar coordinates (r,θ,z), only the displacement $(u_\theta)_i=w_i(r,z,t)$ nonvanishes, where subscripts $i=1,2,3$ refer to materials 1, 2, and 3, and where t is the time. The nonvanishing stress components $\tau_{\theta z}$ and $\tau_{r\theta}$ are as follows:

$$(\tau_{\theta z})_i=\mu_i\frac{\partial w_i}{\partial z},\quad (\tau_{r\theta})=\mu_i\left(\frac{\partial w_i}{\partial r}-\frac{w_i}{r}\right),\quad i=1,2,3. \tag{3}$$

The governing equation of motion gives

$$\frac{\partial^2 w_i}{\partial r^2}+\frac{1}{r}\frac{\partial w_i}{\partial r}-\frac{w_i}{r^2}+\frac{\partial^2 w_i}{\partial z^2}=\frac{\rho_i}{\mu_i}\frac{\partial^2 w_i}{\partial t^2},\quad i=1,3 \tag{4}$$

$$\frac{\partial^2 w_2}{\partial r^2}+\frac{1}{r}\frac{\partial w_2}{\partial r}-\frac{w_2}{r^2}+\frac{\partial^2 w_2}{\partial z^2}+\frac{\mu_2'(z)}{\mu_2(z)}\frac{\partial w_2}{\partial z}=\frac{\rho_2}{\mu_2(z)}\frac{\partial^2 w_2}{\partial t^2} \tag{5}$$

where $\mu_2'(z)$ is the derivative of $\mu_2(z)$ with respect to z.

The boundary conditions are given as follows:

$$(\tau_{\theta z})_1(r,0^-,t)=(\tau_{\theta z})_2(r,0^+,t)=P(r)H(t),\quad 0\leqslant r<a, \tag{6}$$

$$w_1(r,0^-,t)=w_2(r,0^+,t),\quad r\geqslant a, \tag{7}$$

where $H(t)$ is the Heaviside unit step function. The continuity conditions of the displacement and the shear stress across the interfaces give

$$(\tau_{\theta z})_1(r,0^-,t)=(\tau_{\theta z})_2(r,0^+,t),\quad r\geqslant a, \tag{8}$$

$$w_2(r,h^-,t)=w_3(r,h^+,t),\quad 0\leqslant r<\infty, \tag{9}$$

$$(\tau_{\theta z})_2(r,h^-,t)=(\tau_{\theta z})_3(r,h^+,t),\quad 0\leqslant r<\infty. \tag{10}$$

Note that the standard Laplace transform on $f(t)$ is

$$f^*(p)=\int_0^\infty f(t)e^{-pt}dt \tag{11}$$

whose inversion is

$$f(t)=\frac{1}{2\pi i}\int_{\mathrm{Br}} f^*(p)e^{pt}dp \tag{12}$$

and Br denotes the Bromwich path of integration. Applying the transform (11) to Eqs. (4) and (5) results in the transformed equations

$$\frac{\partial^2 w_i^*}{\partial r^2}+\frac{1}{r}\frac{\partial w_i^*}{\partial r}-\frac{w_i^*}{r^2}+\frac{\partial^2 w_i^*}{\partial z^2}=\frac{\rho_i p^2}{\mu_i}w_i^*,\quad i=1,3 \tag{13}$$

$$\frac{\partial^2 w_2^*}{\partial r^2}+\frac{1}{r}\frac{\partial w_2^*}{\partial r}-\frac{w_2^*}{r^2}+\frac{\partial^2 w_2^*}{\partial z^2}+\frac{\mu_2'(z)}{\mu_2(z)}\frac{\partial w_2^*}{\partial z}=\frac{\rho_2 p^2}{\mu_2(z)}w_2^*. \tag{14}$$

Moreover, introducing the pair of Hankel transforms of the first order,

$$V_i(s,z,p)=\int_0^\infty w_i^*(r,z,p)J_1(sr)rdr, \tag{15}$$

$$w_i^*(r,z,p)=\int_0^\infty V_i(s,z,p)J_1(sr)sds, \tag{16}$$

where $J_1(\ )$ is the Bessel function of the first kind, then applying Eq. (15) to the Eqs. (13) and (14) yields

$$\frac{\partial^2 V_i(s,z,p)}{\partial z^2}-\left[s^2+\frac{\rho_i p^2}{\mu_i}\right]V_i(s,z,p)=0,\quad i=1,3 \tag{17}$$

$$\frac{\partial^2 V_2(s,z,p)}{\partial z^2}+\frac{2\alpha}{1+\alpha z}\frac{\partial V_2(s,z,p)}{\partial z}-\left[s^2+\frac{\rho_2 p^2}{\mu_1(1+\alpha z)^2}\right]V_2(s,z,p)=0. \tag{18}$$

Considering the displacement conditions that w_1 and w_2 vanish at $|z|\to\infty$, the solutions of Eqs. (17) and (18) can be expressed as

$$V_1(s,z,p)=A_1(s,p)\exp(\gamma_1 z) \tag{19}$$

$$V_3(s,z,p)=A_4(s,p)\exp(-\gamma_3 z) \tag{20}$$

$$V_2(s,z,p)=A_2(s,p)(1+\alpha z)^{-1/2}I_\beta\left[(1+\alpha z)\frac{s}{|\alpha|}\right]+A_3(s,p)(1+\alpha z)^{-1/2}K_\beta\left[(1+\alpha z)\frac{s}{|\alpha|}\right], \tag{21}$$

where

$$\gamma_1=\sqrt{s^2+\frac{\rho_1 p^2}{\mu_1}},\quad \gamma_3=\sqrt{s^2+\frac{\rho_3 p^2}{\mu_3}},\quad \beta=\sqrt{\frac{1}{4}+\frac{\rho_2 p^2}{\mu_1\alpha^2}} \tag{22}$$

and $I_\beta(\ )$, $K_\beta(\ )$ are the modified Bessel function of the first kind and the second kind, respectively.

From Eq. (16), we can obtain the displacements in the Laplace domain. Subsequently, the shear stresses in the Laplace transform domain $\tau_{\theta z}^*$ and $\tau_{r\theta}^*$ can be obtained from Eq. (3). Then the unknown functions A_1, A_2, A_3, A_4 can be determined from the boundary and the continuity conditions.

3 Derivation of the Singular Integral Equation

In Laplace domain, the boundary conditions become

$$(\tau^*_{\theta z})_1(r,0^-,p)=(\tau^*_{\theta z})_2(r,0^+,p)=\frac{P(r)}{p}, \quad 0\leqslant r<a, \tag{23}$$

$$w^*_1(r,0^-,p)=w^*_2(r,0^+,p), \quad r\geqslant a, \tag{24}$$

and the continuity conditions across the interfaces become

$$(\tau^*_{\theta z})_1(r,0^-,p)=(\tau^*_{\theta z})_2(r,0^+,p), \quad r\geqslant a, \tag{25}$$

$$w^*_2(r,h^-,p)=w^*_3(r,h^+,p), \quad 0\leqslant r<\infty, \tag{26}$$

$$(\tau^*_{\theta z})_2(r,h^-,p)=(\tau^*_{\theta z})_3(r,h^+,p), \quad 0\leqslant r<\infty. \tag{27}$$

To reduce the mixed boundary conditions (23) and (24) into an integral equation, we first define the following dislocation density function on the interface of Material-1 and Material-2:

$$g(r,p)=\frac{1}{r}\frac{\partial}{\partial r}[rw^*_2(r,0^+,p)-rw^*_1(r,0^-,p)]. \tag{28}$$

From the continuity conditions and the dislocation density function, we can obtain

$$(\tau^*_{\theta z})_2(r,0,p)=\mu_2(0)\int_0^a R(u,r,p)g(u,p)u\,du \tag{29}$$

where

$$R(u,r,p)=\int_0^\infty D(s,p)J_1(sr)J_0(su)s\,ds \tag{30}$$

and

$$D(s,p)=\frac{d_{21}(sd_{32}+d_{42})-d_{22}(sd_{31}+d_{41})}{(d_{11}-d_{21})(sd_{32}+d_{42})-(sd_{31}+d_{41})(d_{12}-d_{22})}. \tag{31}$$

The coefficients d_{ij} in Eq. (31) are as follows:

$$\begin{aligned}
d_{11}&=sI_\beta\left(\frac{s}{|\alpha|}\right), \quad d_{12}=sK_\beta\left(\frac{s}{|\alpha|}\right),\\
d_{21}&=-\left(\frac{1}{2}+\beta\right)\alpha I_\beta\left(\frac{s}{|\alpha|}\right)I_{\beta-1}\left(\frac{s}{|\alpha|}\right)\frac{s\alpha}{|\alpha|},\\
d_{22}&=-\left(\frac{1}{2}+\beta\right)\alpha K_\beta\left(\frac{s}{|\alpha|}\right)-K_{\beta-1}\left(\frac{s}{|\alpha|}\right)\frac{s\alpha}{|\alpha|},\\
d_{31}&=(1+\alpha h)^{-1/2}I_\beta\left((1+\alpha h)\frac{s}{|\alpha|}\right),\\
d_{32}&=(1+\alpha h)^{-1/2}K_\beta\left((1+\alpha h)\frac{s}{|\alpha|}\right),\\
d_{41}&=-\left(\frac{1}{2}+\beta\right)\alpha(1+\alpha h)^{-3/2}I_\beta\left((1+\alpha h)\frac{s}{|\alpha|}\right)\\
&\quad+(1+\alpha h)^{-1/2}I_{\beta-1}\left((1+\alpha h)\frac{s}{|\alpha|}\right)\frac{s\alpha}{|\alpha|},\\
d_{42}&=-\left(\frac{1}{2}+\beta\right)\alpha(1+\alpha h)^{-3/2}K_\beta\left((1+\alpha h)\frac{s}{|\alpha|}\right)\\
&\quad-(1+\alpha h)^{-1/2}K_{\beta-1}\left((1+\alpha h)\frac{s}{|\alpha|}\right)\frac{s\alpha}{|\alpha|}.
\end{aligned} \tag{32}$$

Note that

$$\lambda=\lim_{s\to\infty}D(s,p)=-\tfrac{1}{2}. \tag{33}$$

$R(u,r,p)$ can be further expressed as

$$R(u,r,p)=R_n(u,r,p)+R_s(u,r,p) \tag{34}$$

where

$$R_n(u,r,p)=\int_0^\infty[D(s,p)-\lambda]J_1(sr)J_0(su)s\,ds, \tag{35}$$

$$\begin{aligned}
R_s(u,r,p)&=\lambda\int_0^\infty J_1(sr)J_0(su)s\,ds\\
&=-\frac{\lambda}{\pi}\left[\frac{1}{u(u-r)}+\frac{-u-r+2rM(u,r)}{u(u^2-r^2)}\right],
\end{aligned} \tag{36}$$

and

$$M(u,r)=\begin{cases}\dfrac{u}{r}E\left(\dfrac{u}{r}\right), & u<r,\\[2ex] \dfrac{u^2}{r^2}E\left(\dfrac{r}{u}\right)-\dfrac{u^2-r^2}{r^2}K\left(\dfrac{r}{u}\right), & u>r.\end{cases} \tag{37}$$

$E(\ )$ and $K(\ )$ are complete elliptic integrals of the second and first kind, respectively. From the boundary condition (23), we obtain a singular integral equation with a generalized Cauchy kernel,

$$\int_0^a\left[-\frac{\lambda}{\pi}\frac{1}{u-r}+R_0(u,r,p)\right]g(u,p)\,du=\frac{P(r)}{\mu_2(0)p}, \quad 0<r<a, \tag{38}$$

where

$$R_0(u,r,p)=uR_n(u,r,p)+\frac{\lambda}{\pi}\frac{u+r-2rM(u,r)}{u^2-r^2}. \tag{39}$$

The single-valued condition can be given from the definition of $g(u,p)$,

$$\int_0^a ug(u,p)\,du=0. \tag{40}$$

4 Dynamic Stress Intensity Factor

Normalized the interval by the following transformation of variables:

$$u=\frac{a}{2}(1+\xi), \quad r=\frac{a}{2}(1+\eta). \tag{41}$$

The integral Eqs. (38) and (40) can be rewritten as

$$\int_{-1}^1\left[-\frac{\lambda}{\pi}\frac{1}{\xi-\eta}+\overline{R}_0(\xi,\eta,p)\right]G(\xi,p)\,d\xi=\frac{\overline{P}(\eta)}{\mu_2(0)p}, \tag{42}$$

$$\int_{-1}^1(1+\xi)G(\xi,p)\,d\xi=0, \tag{43}$$

where

$$\overline{R}_0(\xi,\eta,p)=\frac{a}{2}R_0\left[\frac{a}{2}(1+\xi),\frac{a}{2}(1+\eta),p\right], \tag{44}$$

$$G(\xi,p)=g\left[\frac{a}{2}(1+\xi),p\right], \tag{45}$$

$$\overline{P}(\eta)=P\left[\frac{a}{2}(1+\eta)\right]. \tag{46}$$

Considering the singularity at the crack tip, we assume that

$$G(\xi,p)=\frac{\overline{G}(\xi,p)}{p}\frac{1}{\sqrt{1-\xi^2}}. \tag{47}$$

Following the numerical method developed by Erdogan for singular integral equations (Erdogan [29]), expanding $\overline{G}(\xi,p)$ in forms of Chebeshev polynomials

$$\overline{G}(\xi,p)=\sum_{n=0}^{\infty} B_n T_n(\xi), \tag{48}$$

we can obtain a system of equations,

$$\sum_{i=1}^{n}\left[\frac{-\lambda}{\xi_i-\eta_j}+\pi\overline{R}_0(\xi_i,\eta_j,p)\right]\frac{\overline{G}(\xi_i,p)}{n}=\frac{\overline{P}(\eta_j)}{\mu_2(0)}, \tag{49}$$

$$\sum_{i=1}^{n}\frac{(1+\xi_i)}{n}\overline{G}(\xi_i,p)=0, \quad j=1,2,\ldots,n-1, \tag{50}$$

where ξ_i, η_j are the roots of Chebeshev polynomial of the first kind and the second kind, respectively,

$$\xi_i=\cos\left(\frac{2i-1}{2n}\pi\right), \quad i=1,2,\ldots,n,$$

$$\eta_j=\cos\left(\frac{j}{n}\pi\right), \quad j=1,2,\ldots,n-1. \tag{51}$$

Solving the system of linear algebraic Eqs. (49) and (50), the unknown function $\overline{G}(\xi,p)$ can be obtained.

If the mode III stress intensity factor in Laplace domain is defined by

$$K^*_{\mathrm{III}}(p)=\lim_{r\to a^+}\sqrt{2(r-a)}(\tau^*_{\theta z})_2(r,0,p), \tag{52}$$

then by using the properties of Chebeshev polynomials, we obtain

$$K^*_{\mathrm{III}}(p)=\lambda\mu_2(0)\sqrt{\frac{a}{2}}\frac{\overline{G}(1,p)}{p}. \tag{53}$$

The dynamic stress intensity factor in time domain can be obtained by

$$K_{\mathrm{III}}(t)=\lambda\mu_2(0)\sqrt{\frac{a}{2}}\frac{1}{2\pi i}\int_{\mathrm{Br}}\frac{\overline{G}(1,p)}{p}e^{pt}dp. \tag{54}$$

5 Results and Discussion

Suppose that the crack surface torsional loading is $P(r)=-\tau_0 r/a$. In this problem, the variables are μ_3/μ_1, h/a, and ρ_3/ρ_1. To investigate the influences of these parameters on the dynamic stress intensity factor, we analyzed some real composite materials, such as Al_2O_3/Ni, TiC/C, SiO_2/Ni, SiC/C, and so on, and found that the parameter μ_3/μ_1 may vary in a wide range but the parameter ρ_3/ρ_1 may vary in a relatively narrow range. Finally, we chose the following combinations for the analysis: $\mu_3/\mu_1=1/12,1/3,3,12$; $\rho_3/\rho_1=0.5,1.0,2.0,4.0$; $h/a=0.2,0.5,1.0,2.0$.

Solving Eqs. (49) and (50), and accomplishing the Laplace inversion (54) by the numerical procedure developed by Miller and Guy [30], the mode III dynamic stress intensity factors in different cases are obtained. The results of the normalized dynamic stress intensity factor $K_{\mathrm{III}}(t)/\tau_0\sqrt{a}$ as a function of $c_{21}t/a$ are shown in Figs. 2–4, where $c_{21}=\sqrt{\mu_1/\rho_1}$ is the shear wave velocity in material-1. A general feature of the curves is observed to be that the stress intensity factors rise rapidly and reach a peak, then oscillate about their static values with decreasing magnification. This general feature has been reported for homogeneous materials and layered composite materials.

Figure 2 shows the variations of the normalized dynamic stress intensity factor with time for various ratios of the shear modulus μ_3/μ_1 while $\rho_3/\rho_1=1.0$ and $h/a=1.0$. It can be seen that the $K_{\mathrm{III}}(t)$ factor tends to monotonically decrease with the increasing of μ_3/μ_1. The differences between the peak values of curves and the static values also decrease with increasing μ_3/μ_1. This ten-

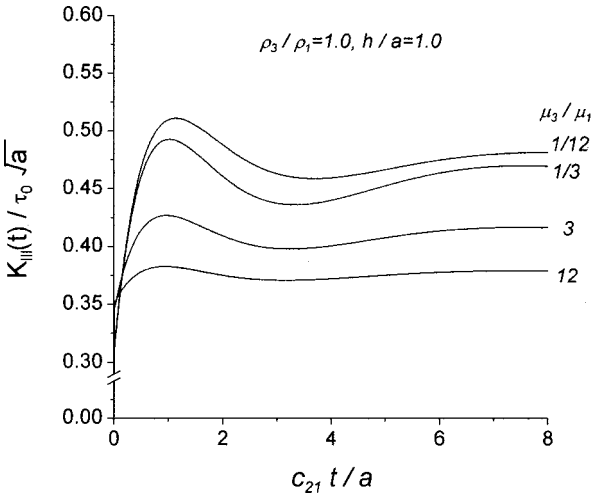


Fig. 2 The effect of the ratio of shear modulus on the normalized dynamic stress intensity factor

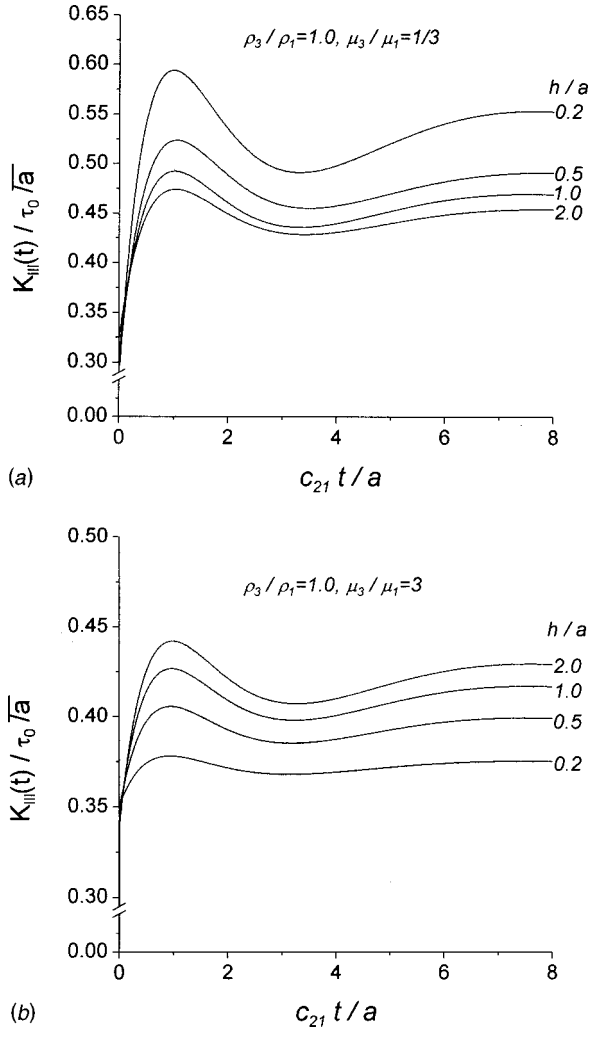


Fig. 3 The effect of the interlayer thickness on the normalized dynamic stress intensity factor

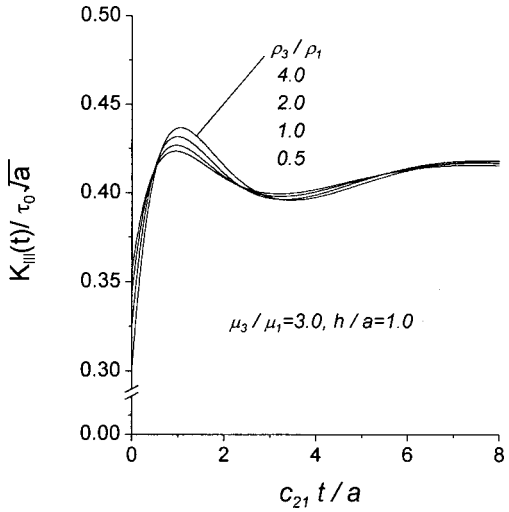


Fig. 4 The effect of the ratio of mass density on the normalized dynamic stress intensity factor

dency is somewhat different from that of two dissimilar materials bonded directly without an interlayer (Ueda, Shindo, and Astumi [14]). In the latter problem the peak values of $K_{III}(t)$ factor decrease with the increasing of μ_3/μ_1, but the intersections exist during the oscillating procedure.

Figures 3(*a*) and 3(*b*) display that the $K_{III}(t)$ factor is also affected by the ratio of interlayer thickness to crack radius h/a. For $\mu_3/\mu_1<1$, the dynamic stress intensity factors decrease with increasing h/a. The larger h/a is, the more the peak value goes beyond its corresponding static value. This phenomenon is depicted in Fig. 3(*a*) for $\mu_3/\mu_1=1/3$. For $\mu_3/\mu_1>1$, the opposite phenomenon can be observed from Fig. 3(*b*) for $\mu_3/\mu_1=3$ that the dynamic stress intensity factors increase with increasing h/a.

The effect of the mass density ratio ρ_3/ρ_1 on the variation of the dynamic stress intensity factor is shown in Fig. 4. This effect has not been reported before for layered composite materials. It is observed that the peak value of $K_{III}(t)$ factor increases when the ratio ρ_3/ρ_1 increases. This phenomenon can be observed for an arbitrary μ_3/μ_1 and different ratios h/a, although these results are not given here as the space of the paper is limited.

As explained in Section 2, in this paper we only use the form $\mu_2(z)=\mu_1(1+\alpha z)^2$ to obtain the solution. A different choice of $\mu_2(z)$ may change the numerical values, but they should not lead to any change in the general trends of the results. We believe it can be verified in our future works by using numerical methods, such as the finite element method.

6 Conclusions

This paper presents the dynamic stress intensity factors for a penny-shaped interface crack in bonded dissimilar homogeneous half-spaces sandwiching an inhomogeneous interlayer. It is assumed that the shear modulus in the direction perpendicular to the crack surface is continuous throughout the space and the crack surfaces are subjected to torsional impact loading. A special model for describing material inhomogeneity parameter is introduced. Laplace and Hankel transforms are applied to reduce the mixed boundary value problem into a singular integral equation with a generalized Cauchy kernel. The results reveal that the dynamic stress intensity factors are affected not only by the stiffness ratio but also by the interlayer thickness and the mass density ratio. It is observed that the influences of the stiffness ratio and the interlayer thickness are stronger than the influences of the mass density ratio.

Acknowledgment

C. Li would like to thank Prof. G. J. Weng for providing him the opportunity to work at Rutgers University and giving him many suggestions in the revised manuscript. The work of C. Li was supported by the National Science Foundation of USA under CMS-9625304. Z. Duan was supported by the National Natural Science Fund of China under the Key Project No. 19891180. Z. Zou was supported by the National Natural Science Fund of China under the project No. 19772029.

References

[1] Williams, M. L., 1959, ''The Stresses Around a Fault or Crack in Dissimilar Media,'' Bull. Seismol. Soc. Am., **49**, pp. 199–204.
[2] Erdogan, F., 1965, ''Stress Distribution in Bonded Dissimilar Materials With Cracks,'' ASME J. Appl. Mech., **32**, pp. 403–410.
[3] Rice, J. R., and Sih, G. C., 1965, ''Plane Problems of Cracks in Dissimilar Media,'' ASME J. Appl. Mech., **32**, pp. 418–423.
[4] Willis, J. R., 1971, ''Fracture Mechanics of Interfacial Crack,'' J. Mech. Phys. Solids, **19**, pp. 353–368.
[5] Qu, J., and Bassani, J. L., 1993, ''Interfacial Fracture Mechanics for Anisotropic Bimaterials,'' ASME J. Appl. Mech., **60**, pp. 422–431.
[6] Comninou, M., 1977, ''The Interface Crack,'' ASME J. Appl. Mech., **44**, pp. 631–636.
[7] Achenbach, J., Keer, L., Khetan, R., and Chen, S., 1979, ''Loss of Adhesion at the Tip of an Interface Crack,'' J. Elast., **9**, pp. 397–424.
[8] Rice, J., 1988, ''Elastic Fracture Mechanics Concepts for Interface Cracks,'' ASME J. Appl. Mech., **55**, pp. 98–103.
[9] Shih, C., and Asaro, R., 1988, ''Elastic-Plastic Analysis of Cracks on Bimaterials Interfaces. Part I: Small Scale Yielding,'' ASME J. Appl. Mech., **55**, pp. 299–316.
[10] Hutchinson, J. W., and Suo, Z., 1992, ''Mixed Mode Cracking in Layered Materials,'' Adv. Appl. Mech., **29**, pp. 63–191.
[11] Sih, G. C., and Chen, E. P., 1981, *Mechanics of Fracture 6: Cracks in Composite Materials*, Martinus Nijhoff, The Hague.
[12] Takei, M., Shindo, Y., and Astumi, A., 1982, ''Diffraction of Transient Horizontal Shear Waves by a Finite Crack at the Interface of Two Bonded Dissimilar Elastic Solids,'' Eng. Fract. Mech., **16**, pp. 799–807.
[13] Li, D. H., and Tai, W. H., 1991, ''Elastodynamic Response of an Interface Crack in a Layered Composite Under Antiplane Shear Impact Load,'' Eng. Fract. Mech., **39**, pp. 687–693.
[14] Ueda, S., Shindo, Y., and Astumi, A., 1983, ''Torsional Impact Response of a Penny-Shaped Crack on a Bimaterial Interface,'' Eng. Fract. Mech., **18**, pp. 1059–1066.
[15] Lambros, J., and Rosakis, A. J., 1995, ''Shear Dominated Transonic Interfacial Crack Growth in a Bimaterial—I Experimental Observations,'' J. Mech. Phys. Solids, **43**, pp. 169–188.
[16] Singh, R. P., Lambros, J., and Rosakis, A. J., 1997, ''Investigation of the Mechanics of Intersonic Crack Propagation Along a Bimaterial Interface Using Coherent Gradient Sensing and Photoelasticity,'' Proc. R. Soc. London, Ser. A, **A453**, pp. 2649–2667.
[17] Xu, X. P., and Needleman, A., 1996, ''Numerical Simulations of Dynamic Crack Growth Along an Interface,'' Int. J. Fract., **74**, pp. 289–324.
[18] Needleman, A., and Rosakis, A. J., 1999, ''The Effect of Bond Strength and Loading Rate on the Conditions Governing the Attainment of Intersonic Crack Growth Along Interfaces,'' J. Mech. Phys. Solids, **47**, pp. 2411–2449.
[19] Rosakis, A. J., and Ravichandran, G., 2000, ''Dynamic Failure Mechanics,'' Int. J. Solids Struct., **37**, pp. 331–348.
[20] Subramanian, R. V., and Crasto, A. S., 1986, ''Electrodeposition of a Polymer Interphase in Carbon-Fiber Composites,'' Polym. Compos., **7**, pp. 201–218.
[21] Lugscheider, E., 1987, ''Plasma Spraying for Wear Applications,'' *Thermal Spray: Advances in Coating Technology*. Proceedings of the National Thermal Spray Conference. D. L. Houck, ed., ASM International, Materials Park, OH.
[22] Shiau, F. Y., Zuo, Y., Zeng, X. Y., Lin, J. C., and Chang Y. A., 1988, ''Interfacial Reactions Between CO and GaAs,'' *Adhesion in Solids, Material Research Society Symposium*, Proc. Vol. 119, D. M. Mattox, J. E. Baglin, R. J. Gottshail, and C. D. Batich, eds., Materials Research Society, Pittsburgh, PA, pp. 171–176.
[23] Kurihara, K., Sasaki, K., and Kawarada, M., 1990, ''Adhesion Improvement of diamond films,'' *FGM'90-Proc. of the First International Symposium on Functionally Graded Materials*, M. Yamanouchi, M. Koizumi, T. Hirai, and I. Shiota eds., Functionally Graded Material Forum, Sendai, Japan, pp. 65–69.
[24] Jager, D. A., Stover, D. and H. G. Schutz, 1991, ''Plasma Spraying of Graded Composites,'' *Thermal Spray Coatings: Properties, Processes and Applications*. Proceedings of the National Thermal Spray Conference. T. F. Bernecki, ed., ASM International, Materials Park, OH.
[25] Delale, F., and Erdogan, F., 1988, ''On the Mechanical Modeling of the Interfacial Region in Bonded Half Planes,'' ASME J. Appl. Mech., **55**, pp. 317–324.
[26] Ozturk, M., and Erdogan, F., 1995, ''An Axisymmetric Crack in Bonded Materials With an Inhomogeneous Interfacial Zone Under Torsion,'' ASME J. Appl. Mech., **62**, pp. 116–125.
[27] Wang, X. Y., Zou, Z. Z., and Wang, D., 1996, ''On the Griffith Crack in a

Inhomogeneous Interlayer of Adjoining Two Different Elastic Materials,'' Int. J. Fract., **79**, pp. R51–R56.
[28] Fildis, H., and Yahsi, O. S., 1996, ''The Axisymmetric Crack Problem is a Non-homogeneous Interfacial Region Between Homogeneous Half Spaces,'' Int. J. Fract., **78**, pp. 139–164.
[29] Erdogan, F., 1975, *Complex Function Technique, In Continuum Physics*, Vol. II, Academic Press, San Diego, CA, pp. 523–603.
[30] Miller, M. K., and Guy, W. T., 1966, ''Numerical inversion of the Laplace transform by use of Jacobi polynomials,'' SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal., **3**, pp. 624–635.

Elastic-Inelastic Self-Consistent Model for Polycrystals

A. Abdul-Latif[1]
ERBEM/GIM,
Université Paris8,
IUT de Tremblay,
93290 Tremblay-en-France, France
e-mail: aabdul@iut-tremblay.univ-paris8.fr

J. P. Dingli

K. Saanouni

GSM/LASMIS,
Université de Technologie de Troyes,
B. P. 2060,
10010 Troyes cedex, France

Based on a well-established nonincremental interaction law for fully anisotropic and compressible elastic-inelastic behavior of polycrystals, tangent formulation-based and simplified interaction laws, of softened nature, are derived to describe the nonlinear elastic-inelastic behavior of fcc polycrystals under different loading paths. Within the framework of small strain hypothesis, the elastic behavior, which is defined at granular level, is assumed to be isotropic, uniform, and compressible neglecting the grain rotation. The heterogeneous inelastic deformation is microscopically determined using the slip theory. In addition, the granular elastic behavior and its heterogeneous distribution from grain to grain within a polycrystal are taken into account. Comparisons between these two approaches show that the simplified one is more suitable to describe the overall responses of polycrystals notably under multiaxial loading paths. Nonlinear stress-strain behavior of polycrystals under complex loading, especially a cyclic one, is of particular interest in proposed modeling. The simplified model describes fairly well the yield surface evolution after a certain inelastic prestraining and the principle cyclic features such as Bauschinger effect, additional hardening, etc. [DOI: 10.1115/1.1427693]

1 Introduction

Since the first reported study in the area of the self-consistent approach originated by Sachs [1], then Cox and Sompmith [2], and Taylor [3], this area has been a topic of increasing interest in the field of polycrystal modeling. For small and large deformation conditions, many research efforts have been theoretically developed. These concern purely elastic case (see, for example, Hershey [4] and Kröner [5]) or viscoelastic behavior (see, for example, Laws and McLaughlin [6] and Kouddane et al. [7]) as well as inelastic behavior, i.e., plastic or viscoplastic (Brown [8], Rice [9,10], Hutchinson [11], Molinari et al. [12], Weng [13], and Lebensohn and Tomé [14,15]) or elastic-inelastic behavior of polycrystals (Lin [16], Kröner [17], Budianski and Wu [18], Hill [19], Hutchinson [20], Berveiller and Zaoui [21], Weng [22], Iwakuma and Nemat-Nasser [23], Nemat-Nasser and Obata [24], Lipinski et al. [25,26], Kouddane et al. [7], Rougier et al. [27], Molinari et al. [28], Schmitt et al. [29], Abdul-Latif et al. [30], and others).

The polycrystal is usually viewed as an aggregate of numerous (single or polyphase crystal) grains with different orientations with respect to the loading axes. The number, orientation, and morphology of grains play an important role on the predicted result. The case of the random crystal distribution in an aggregate of grains is considered here according to the macroscopic isotropic behavior of the aggregate. The emphasis is placed here on the single-phase fcc polycrystals and the properties of each grain are identical with respect to the crystallographic reference system. The granular heterogeneity comes, in general, from the differences in the orientation of the grains and the single-crystal intrinsic anisotropy since the morphology and spatial distribution of the grains are not taken into account.

For the nonlinear elastic-inelastic behavior of polycrystals, the developed approaches up to now give, in general, approximate solutions. Whatever the self-consistent approach, the inelastic strain in each grain is calculated based on the glide on the crystallographic slip system (css) level. The resolved shear stresses τ^s are also determined on all slip systems by the well-known Schmid's law ([31]). Hence, the modeling of the single crystal, for such approaches, is almost standard and has indisputable treatments. Since the overall behavior of polycrystals is strongly influenced by grain/matrix interaction law, therefore the heterogeneous stress and strain fields throughout the matrix (aggregate) necessitate obviously the grain interaction consideration. Moreover, the type of the interaction (hard or soft) of the grain with its matrix gives an appropriate estimation about the accommodated plastic strain and its repartition between the grain and the surrounding. A trivial way to theoretically obtain this grain interaction is the well-known self-consistent relations. Thus, the interaction law represents an extremely important key factor for this type of modeling. Some contributions (for example, Kouddane et al. [7], Weng [13], and Molinari et al. [28]) have developed approximate solutions taking into account partly the viscous character of the intergranular interactions. A recent work ([30]) was devoted to mainly describe the overall mechanical cyclic behavior of polycrystals under complex loading paths. This represents the first simplified version of a self-consistent interaction law proposed by Kouddane et al. [7] in the case of incompressible elastic properties.

In this work, based on one-site nonincremental interaction law for anisotropic and compressible elastic-inelastic behavior of an inclusion embedded in infinite matrix ([12]), tangent formulation-based and simplified interaction laws are determined. The obtained models allow to describe the elastic-inelastic behavior under different loading paths (monotonic and cyclic) for fcc polycrystals. The theoretical basis of rate-dependent inelastic strain is examined at the css level. Contrary to the Abdul-Latif et al.'s hypothesis ([30]) which cannot fulfill the physical requirement entirely, the granular elastic part of the strain is now assumed to be compressible, uniform, and isotropic. Further, the overall kinematic hardening can be naturally described by the obtained interaction law of the grain with the surrounding. This is due to the existence of the granular elastic behavior in the interaction law since the intergranular accommodation has an elastic nature ([27]). In the case of simplified law, a phenomenological parameter is introduced in order to reproduce the elastic-inelastic behavior of polycrystals under multiaxial cyclic loadings as a particular interest of this model.

In view of the importance of the fatigue rupture of polycrystals, a self-consistent model describing the plastic fatigue behavior of

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Jan. 25, 2000; final revision, Apr. 30, 2001. Associate Editor: J. W. Ju. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

polycrystals has been already proposed (Abdul-Latif and Saanouni [32–34], Saanouni and Abdul-Latif [35], Abdul-Latif [36], and Abdul-Latif et al. [37]). It was based on the self-consistent model proposed by Cailletaud [38] as an initial model of elastic-viscoplastic behavior of fcc metals. Note that the actual developed elastic-inelastic model uses less number of internal variables and material coefficients in comparison with Cailletaud's model ([38]). Hence, it seems to be a judicious choice when the present self-consistent model will be coupled with damage as a future work predicting the low-cycle fatigue life of metallic polycrystals.

2 Self-Consistent Modeling

2.1 Grain/Matrix Interaction Relation. A simplified interaction law is examined and gives the granular stress fields due to the difference between the granular strain rate and that of the aggregate (matrix). According to the advantage that the nonincremental formulation is shown to be softer grain/matrix interactions than the incremental model, the emphasis is placed on the first version made by Kouddane et al. [7] and then by Molinari et al. [28]. Concerning Kouddane et al.'s work ([7]), an elastic-viscoplastic nonincremental self-consistent model for incompressible polycrystals has been proposed. Motivated by a Maxwell-type law, the interaction law for a spherical isotropic and incompressible inclusion with its matrix is given by

$$-\frac{1}{3\mu^o}(\dot{\mathbf{s}}^g-\dot{\mathbf{S}})-\frac{1}{3\eta^o}(\mathbf{s}^g-\mathbf{S})=(\dot{\boldsymbol{\varepsilon}}^g-\dot{\mathbf{E}}) \tag{1}$$

where $\dot{\boldsymbol{\varepsilon}}^g$ and $\dot{\mathbf{E}}$ are, respectively, the total granular and overall strain rates. $\dot{\mathbf{S}}$ and $\dot{\mathbf{s}}^g$ are the global and granular rates of deviatoric parts of Cauchy stress tensor, respectively. η^o is the scalar macroscopic viscous tangent modulus depending on the deformation history, and μ^o is the the uniform macroscopic shear modulus. Note that the viscous relaxation is taken into account through the term $(-1/3\eta^o)$. In a linear elastic case, the interaction law (1) can take into account the instantaneous elastic response, i.e., when the viscosity tends to infinity $(\eta^o\to\infty)$, the total strain rate becomes fully elastic $\dot{\mathbf{E}}\to\dot{\mathbf{E}}_e$. This leads to the same form of Kröner's interaction relation for an elastic inclusion

$$\dot{\mathbf{s}}^g-\dot{\mathbf{S}}=-3\mu^o(\dot{\boldsymbol{\varepsilon}}_e^g-\dot{\mathbf{E}}_e). \tag{2}$$

Note that this relation (2) remains valid under a large jump of the strain rate where the elastic response dominates.

In the case where a constant strain rate is applied, a viscous relaxation, at steady state condition, will dominate, i.e., we obtain $(\dot{\mathbf{s}}^g-\dot{\mathbf{S}})\to 0$ for $t\to\infty$. Therefore, the interaction law (1) becomes

$$\mathbf{s}^g-\mathbf{S}=-3\,\eta^o(\dot{\boldsymbol{\varepsilon}}_{in}^g-\dot{\mathbf{E}}_{in}). \tag{3}$$

In order to satisfy the self-consistency conditions, η^o has to be adjusted at each instant (Kouddane et al. [7]). Numerically, this task, for uniaxial cyclic loading conditions, is relatively reasonable (Dingli [39]). However, if complex biaxial loading paths are applied, such conditions will not be respected by the adjustment of the scalar η^o, since it depends always on one strain rate direction and not on two or more directions in the same time. Hence, a symmetric fourth-rank tangent viscous modulus tensor can be used instead of the scalar η^o. In this case, the determination of all the viscous parameters of such operator becomes a very difficult task particularly from algorithmic point of view. In addition, the use of nonisotropic tangent viscous modulus introduces a considerable difficulty in the derivation of the interaction law.

A simplified solution of the grain/matrix interaction law given by (1) has been performed by the authors ([30]) to get

$$\dot{\mathbf{s}}^g-\dot{\mathbf{S}}+\beta(\mathbf{s}^g-\mathbf{S})=-3\mu^o(\dot{\boldsymbol{\varepsilon}}^g-\dot{\mathbf{E}}). \tag{4}$$

This simple modification is shown to be enough to simulate accurately the polycrystal behavior under different cyclic loading paths when the parameter β is carefully determined ([30]).

Recently, a generalization of the Kouddane et al.'s model ([7]) for fully anisotropic and compressible elastic-viscoplastic behavior with small strain hypothesis was developed by Molinari et al. [28]. Based on the tangent formulation, the local and global behaviors are also presented here by elastic-viscoplastic relation but for compressible elastic properties. In order to simplify the significant difficulty of the resolution of this problem, Molinari et al. [28] proposed an approximate solution keeping the same structure of the incompressible interaction law (Kouddane et al. [7]), i.e., with elastic and viscoplastic parts. The generalized elastic-viscoplastic interaction law of Molinari et al. [28] is expressed as follows:

$$(\mathfrak{J}^{s^{-1}}+\mathbf{C})^{-1}:(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})+(\mathfrak{J}'^{s^{-1}}+\mathbf{A})^{-1}:(\mathbf{s}^g-\mathbf{S})=(\dot{\boldsymbol{\varepsilon}}^g-\dot{\mathbf{E}}) \tag{5}$$

where $\mathfrak{J}^s$ and $\mathfrak{J}'^s$ are respectively fourth rank tensors which have to be computed by Green function and integral methods using $\mathbf{A}$ and $\mathbf{C}$ ([28]). $\mathbf{A}$ is the macroscopic tangent viscoplastic modulus and $\mathbf{C}$ being the global stiffness tensor.

In the case where the elastic response dominates the viscoplastic term $(\mathfrak{J}'^{s^{-1}}+\mathbf{A})^{-1}$ becomes negligible with respect to the elastic part, and the interaction law can be written by

$$(\mathfrak{J}^{s^{-1}}+\mathbf{C})^{-1}:(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})=(\dot{\boldsymbol{\varepsilon}}_e^g-\dot{\mathbf{E}}_e). \tag{6}$$

It corresponds to the solution of the elasticity problem for heterogeneous materials.

For a spherical inclusion embedded in an infinite homogenous isotropic matrix having elastic properties defined by μ and λ (the classical Lame's constants), the fourth-order interaction tensor is determined ([40]) as

$$\mathfrak{J}^s_{ijkl}=\frac{1}{15\mu(3\lambda+6\mu)}[(3\lambda+3\mu)\delta_{ij}\delta_{kl}-3(3\lambda+8\mu)I_{ijkl}] \tag{7}$$

with

$$I_{ijkl}=\frac{1}{2}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}). \tag{8}$$

As a consequence, the elastic part of the interaction law can be expressed by

$$2A(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})+\mathrm{Btr}(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})\mathbf{1}=(\dot{\boldsymbol{\varepsilon}}_e^g-\dot{\mathbf{E}}_e) \tag{9}$$

where the constants A and B are defined as

$$A=\frac{-(8\mu+3\lambda)}{2\mu(14\mu+9\lambda)} \tag{10}$$

$$B=\frac{(6\mu+\lambda)(3\lambda+8\mu)}{\mu(448\mu^2+456\mu\lambda+108\lambda^2)} \tag{11}$$

For a fully viscoplastic behavior dominating at stationary state (in the long range response), the term $(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})$ is practically vanished. Therefore, Eq. (5) can be written approximately as

$$(\mathfrak{J}'^{s^{-1}}+\mathbf{A})^{-1}:(\mathbf{s}^g-\mathbf{S})=(\dot{\boldsymbol{\varepsilon}}_{in}^g-\dot{\mathbf{E}}_{in}). \tag{12}$$

This equation represents a self-consistent approach developed by Molinari et al. [12] describing the viscoplastic behavior of polycrystals under large deformation condition.

In the case where the viscoplastic behavior of the matrix is supposed isotropic and incompressible, the tangent modulus can be approximated by

$$A_{ijkl}=2\,\eta^o I_{ijkl}. \tag{13}$$

For spherical inclusion, the interaction tensor $\mathfrak{J}'^s$ is given by

$$\mathfrak{J}'^s_{ijkl}=-\frac{1}{5\,\eta^o}I_{ijkl}. \tag{14}$$

The viscoplastic part of the general interaction law (5) is therefore deduced as

$$-\frac{1}{3\eta^o}(\mathbf{s}^g-\mathbf{S})=(\dot{\boldsymbol{\varepsilon}}^g_{in}-\dot{\mathbf{E}}_{in}). \tag{15}$$

The general interaction law for the elastic-viscoplastic behavior of polycrystal (Eq. (5)) is equal to the sum of the two approximate parts (elastic: Eq. (9) and viscoplastic: Eq. (15)). Thus, we get

$$2A(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})+\mathrm{Btr}(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})\mathbf{1}-\frac{1}{3\eta^o}(\mathbf{s}^g-\mathbf{S})=(\dot{\boldsymbol{\varepsilon}}^g-\dot{\mathbf{E}}). \tag{16}$$

The same problem of η^o adjustment, as in the incompressible case (Eq. (4)), came across in order to satisfy the self-consistency conditions especially under multiaxial loading path. Hence, the term $(1/3\eta^o)$ is replaced by a phenomenological parameter $(\alpha>0)$ and the simplified interaction law can be written as

$$2A(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})+\mathrm{Btr}(\dot{\boldsymbol{\sigma}}^g-\dot{\boldsymbol{\Sigma}})\mathbf{1}-\alpha(\mathbf{s}^g-\mathbf{S})=(\dot{\boldsymbol{\varepsilon}}^g-\dot{\mathbf{E}}). \tag{17}$$

2.2 Single-Crystal Constitutive Relations. At this level, it is assumed that slip is the dominant deformation mechanism and other mechanisms like twinning, grain boundary sliding, etc., are neglected. The constitutive equations of the inelastic strain are examined at the css scale in the case of fcc structure. The resolved shear stresses τ^s are determined by the twice-contracted tensorial product between $\boldsymbol{\sigma}^g$ and the Schmid factor tensor $\mathbf{m}^s$:

$$\tau^s=\boldsymbol{\sigma}^g:\mathbf{m}^s \tag{18}$$

$$\mathbf{m}^s=\frac{1}{2}[\mathbf{n}^s\otimes\mathbf{b}^s+\mathbf{n}^s\otimes\mathbf{b}^s] \tag{19}$$

where $\mathbf{b}^s$ is the unit vector in the slip direction and $\mathbf{n}^s$ is the vector normal to the slip plane.

For each slip system, only the transgranular isotropic hardening is modeled and represented by the couple of the internal variables (q^s,R^s) describing the expansion of the elastic domain on the system s. Throughout this paper, the index $s\in\{1,2,\ldots,n\}$ is associated to the system rank, with n being the maximum number of octahedral systems in the grain ($n=12$ for fcc).

The elastic and inelastic parts of the granular specific free energy ψ^g can be written as

$$\psi^g=\psi^g_e+\psi^g_{in}. \tag{20}$$

As an internal variable, the granular elastic part $\boldsymbol{\varepsilon}^g_e$ is assumed to be uniform, isotropic, and compressible and its associated variable is thermodynamically represented by a granular stress tensor $\boldsymbol{\sigma}^g$.

$$\rho\psi^g_e(\boldsymbol{\varepsilon}^g_e)=\frac{1}{2}\lambda(\mathrm{tr}\,\boldsymbol{\varepsilon}^g_e)^2+\mu\mathrm{tr}(\boldsymbol{\varepsilon}^g_e)^2 \tag{21}$$

where λ and μ are the classical Lame's constants of the grain, and ρ is the material density.

The granular stress tensor $\boldsymbol{\sigma}^g$ can be deduced as

$$\boldsymbol{\sigma}^g=\rho\frac{\partial\psi^g_e}{\partial\boldsymbol{\varepsilon}^g_e}=2\mu\boldsymbol{\varepsilon}^g_e+\lambda(\mathrm{tr}\,\boldsymbol{\varepsilon}^g_e)\mathbf{I} \tag{22}$$

where $\mathbf{I}$ is the second-order unit tensor.

At a constant temperature, the granular coefficients (λ and μ) remain always constants, the time derivative of Eq. (22) gives

$$\dot{\boldsymbol{\sigma}}^g=2\mu\dot{\boldsymbol{\varepsilon}}^g_e+\lambda(\mathrm{tr}\,\dot{\boldsymbol{\varepsilon}}^g_e)\mathbf{I} \tag{23}$$

and

$$\dot{\boldsymbol{\varepsilon}}^g_e=\frac{\dot{\boldsymbol{\sigma}}^g}{2\mu}-\frac{\lambda}{2\mu(2\mu+3\lambda)}\mathrm{tr}(\dot{\boldsymbol{\sigma}}^g)\mathbf{I}. \tag{24}$$

According to the small strain hypothesis, the total granular strain $\boldsymbol{\varepsilon}^g$ is partitioned into elastic $\boldsymbol{\varepsilon}^g_e$ and inelastic $\boldsymbol{\varepsilon}^g_{in}$ parts:

$$\boldsymbol{\varepsilon}^g=\boldsymbol{\varepsilon}^g_e+\boldsymbol{\varepsilon}^g_{in}. \tag{25}$$

The granular inelastic part of the state potential $\rho\psi^g_{in}$ is expressed (Saanouni and Abdul-Latif [35]) as follows:

$$\rho\psi^g_{in}=\frac{1}{2}\sum_{r=1}^{n}\sum_{s=1}^{n}H_{rs}Q^sq^rq^s \tag{26}$$

where Q^s is the transgranular isotropic hardening modulus of the css. The hardening interaction matrix H_{rs} is supposed to describe dislocation-dislocation interactions (Kocks and Brown [41], Jackson and Basinski [42], and Franciosi [43]). The dual variable R^s (state law) can be derived from (26) as follows:

$$R^s=\rho\frac{\partial\psi^g_{in}}{\partial q^s}=Q^s\sum_{r=1}^{n}H_{rs}q^r. \tag{27}$$

The slip rate can be determined as long as the shear stress and the hardening variables are known. Hence, the evolution rates can be obtained by introducing a yield function f^s together with a dissipation potential F^s for each slip system according to the non associated plasticity framework.

$$f^s=|\tau^s|-R^s-k^s_o \tag{28}$$

where k^s_o is the initial value of the critical resolved shear stress (friction stress).

The transgranular inelastic dissipation potential is similarly written as initially proposed in Saanouni and Abdul-Latif [35].

$$F^s=f^s+b^sq^sR^s \tag{29}$$

where b^s characterizes the nonlinearity of the local isotropic hardening.

Now, the evolution laws of the granular inelastic strain and the internal variable of the transgranular isotropic hardening are given by generalized normality rule:

At the granular level

$$\dot{\boldsymbol{\varepsilon}}^g_{in}=\sum_{s=1}^{n}\dot{\lambda}^s\frac{\partial F^s}{\partial\boldsymbol{\sigma}^g}=\sum_{s=1}^{n}\dot{\lambda}^s\,\mathrm{sign}(\tau^s)\mathbf{m}^s=\sum_{s=1}^{n}\dot{\gamma}^s\mathbf{m}^s \tag{30}$$

with

$$\dot{\gamma}^s=\dot{\lambda}^s\,\mathrm{sign}(\tau^s) \tag{31}$$

where $\dot{\gamma}^s$ is the slip rate on the css.

At the css level, the rate of the isotropic hardening variable evolution is expressed by

$$\dot{q}^s=\sum_{r=1}^{n}\dot{\lambda}^r\frac{\partial F^r}{\partial R^s}=\dot{\lambda}^s(1-b^sq^s). \tag{32}$$

In the framework of viscoplasticity, the value of "pseudo-multiplier" $\dot{\lambda}^s$ for each css is a power function of the distance to the yield point defined by the criterion f^s:

$$\dot{\lambda}^s=\left\langle\frac{f^s}{K^s}\right\rangle^{z^s}=\left\langle\frac{|\tau^s|-R^s-k_o}{K^s}\right\rangle^{z^s} \tag{33}$$

where K^s and z^s are material constants describing the local viscous effect of the material.

It is worth noting that the Eq. (33) together with (31) represent generalization to hardening case of the classical power-law micro-constitutive equation frequently used in the literature.

The intrinsic dissipation at the granular level is given by

$$\boldsymbol{\sigma}^g:\dot{\boldsymbol{\varepsilon}}^g_{in}-\sum_{s=1}^{n}R^s\dot{q}^s\geqslant 0. \tag{34}$$

By using (27), (28), (30), and (31), it is easy to show the thermodynamical admissibility of the single-crystal model. This can be achieved by showing that the inequality (34) is satisfied as long as the parameters Q^s and k^s_o are positive.

Table 1 identified model constants ([44])

λ (MPa)	μ (MPa)	α	k_0 (MPa)	z	K	Q (MPa)	b	h_1	h_2	h_3	h_4	h_5	h_6
144780	81440	5×10^{-7}	240	12	38	256	12.6	1	0.75	0.74	1.48	1.05	3.5

2.3 Polycrystal Constitutive Relations. After determining the granular inelastic strain rate as the sum of the contribution from all activated slip systems, the transition from the single to polycrystal response is performed by the well-known averaging procedures depending on the granular (elastic and inelastic) rates.

The overall total strain (elastic and inelastic) rate is therefore calculated as follows

$$\dot{\mathbf{E}}=\dot{\mathbf{E}}^e+\dot{\mathbf{E}}^{in} \tag{35}$$

$$\dot{\mathbf{E}}^e=\sum_{g=1}^{Ng} v^g\dot{\boldsymbol{\varepsilon}}_e^g \tag{36}$$

$$\dot{\mathbf{E}}^{in}=\sum_{g=1}^{Ng} v^g\dot{\boldsymbol{\varepsilon}}_{in}^g \tag{37}$$

where v^g represents the volume fraction of the same oriented grains. $\dot{\boldsymbol{\varepsilon}}_e^g$ and $\dot{\boldsymbol{\varepsilon}}_{in}^g$ are given by Eqs. (24) and (30), respectively.

An attempt is performed to compare two formulations:

(i) The tangent formulation-based interaction law given by Eq. (16), itself based on the following macroscopic combination of Hooke's law and tangent viscoplastic relation.

$$\dot{\boldsymbol{\Sigma}}=2\mu\dot{\mathbf{E}}+\lambda(\mathrm{tr}\,\dot{\mathbf{E}})\mathbf{I}-\frac{\mu^o}{\eta^o}(\mathbf{S}-\mathbf{S}^o) \tag{38}$$

where $\mathbf{S}^o$ is the macroscopic back stress.

(ii) The proposed simplified interaction law (Eq. (17)) together with the macroscopic fully elastic behavior given by

$$\dot{\boldsymbol{\Sigma}}=2\mu\dot{\mathbf{E}}_e+\lambda(\mathrm{tr}\,\dot{\mathbf{E}}_e)\mathbf{I}. \tag{39}$$

Some numerical applications are conducted using both models in order to study and to compare their overall responses under uniaxial loading situations.

For the tangent formulation-based model, the adjustment process of η^o in order to fulfill the self-consistency conditions can be performed as follows:

In strain-controlled conditions

$$\dot{\mathbf{E}}=\dot{\mathbf{E}}' \tag{40}$$

where $\dot{\mathbf{E}}'$ is the imposed macroscopic total strain rate.

While, in stress-controlled conditions

$$\dot{\boldsymbol{\Sigma}}=\dot{\boldsymbol{\Sigma}}' \tag{41}$$

where $\dot{\boldsymbol{\Sigma}}'$ represents the imposed macroscopic stress rate.

The exact satisfaction of the self-consistency condition (Eq. (40) or Eq. (41)) is not always evident due to the fact that the adjustment process is performed only on the scalar η^o. Thus, η^o is calculated in order to minimize the relative error (Re) defined by in strain-controlled condition:

$$\mathrm{Re}=\|\dot{\mathbf{E}}-\dot{\mathbf{E}}'\|/\|\dot{\mathbf{E}}'\| \tag{42}$$

and, in stress-controlled condition:

$$\mathrm{Re}=\|\dot{\boldsymbol{\Sigma}}-\dot{\boldsymbol{\Sigma}}'\|/\|\dot{\boldsymbol{\Sigma}}'\| \tag{43}$$

where the norm $\|\mathbf{x}\|$ is defined as $\|\mathbf{x}\|=\sqrt{a\mathbf{x}:\mathbf{x}}$ with $a=2/3$ for a strain-controlled situation and $a=3/2$ for a stress-controlled condition.

3 Comparison Between the Two Models and Parametric Study

The simplified self-consistent model is tested using an aggregate of 200 grains ([30]). For simplicity, it is assumed that all the grains and the slip systems have the same material properties. Consequently, all the grains have the same value of the elastic constants λ and μ and all the octahedral slip systems have the same constants z, K, k_o, Q, and b. The identified constants (Dingli et al. [44]) are listed in Table 1.

Some numerical comparisons between the tangent formulation-based and the simplified model are carried out. The first comparison is conducted employing monotonic tensile tests. Three uniaxial macroscopic strain rates are used ($\dot{E}_{11}=0.1$, 0.01 and 0.001/s). It is important to note that, for the tangent formulation-based model (Eqs. (16, 38)), the value of η^o, which satisfies the self-consistency condition, remains almost constant for each controlled strain rate for any time $t>0$, since there is no change neither in the strain rate nor in its direction. Nevertheless, for the simplified model, the self-consistency conditions are systematically satisfied. Examination of Fig. 1 shows the overall responses of both models using the above strain rates. For a given strain rate, both models give practically the same overall responses showing their sensitivity to the strain rate.

The effect of an abrupt change of uniaxial loading direction on the overall polycrystal behavior is studied. One cycle of tension-compression with strain-controlled condition $\Delta E_{11}=1$ percent is used (Fig. 2(*a*)). Obviously, both models describe appropriately the overall stress evolution notably in the inelastic zone, i.e., hardening evolution. Moreover, the similarity of the responses is well captured by these models. For the tangent formulation-based model, the evolution of η^o is recorded and analyzed during this uniaxial cyclic loading. Figure 2(*b*) reveals that its value remains constant when a constant strain rate is applied (tensile phase) with which the consistency condition is almost respected. However, as soon as a rapid change in the strain rate direction is taken place ($t=10$ seconds), an extremely high jump of η^o value is consequently recorded. This evolution can be interpreted by the fact that, to have an instantaneous overall elastic response, i.e., $\dot{E}_{11}$

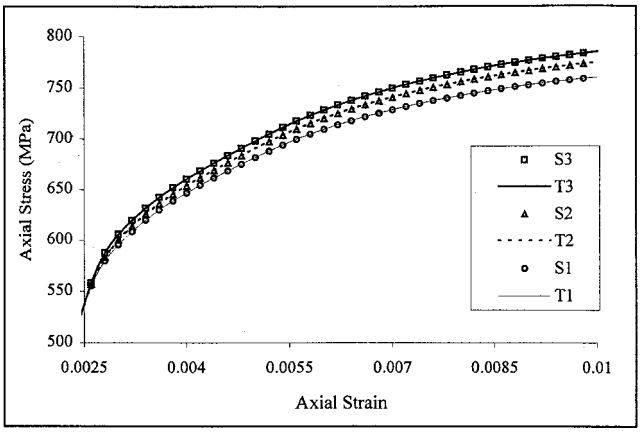


Fig. 1 Plot of comparison between tangent formulation-based (T) and simplified (S) models showing the overall stress evolution versus the strain for uniaxial tensile test under three strain rate values (1:0.1/s, 2: 0.01/s and 3:0.001/s)

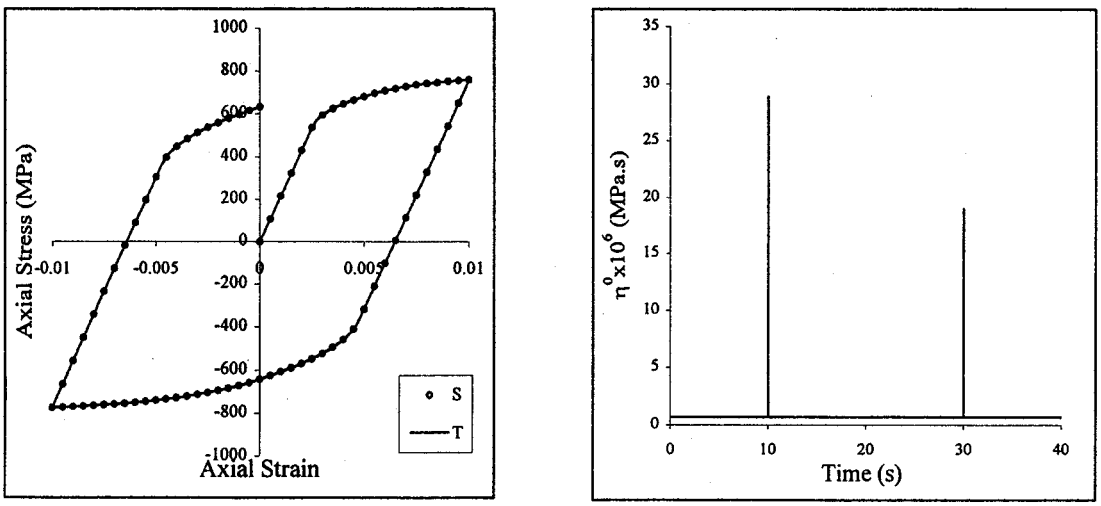


Fig. 2 Plots showing (*a*) a comparison between tangent formulation-based (T) and simplified (S) models for tension-compression loading, (*b*) the evolution of η^o versus time during tension-compression loading test

$\rightarrow \dot{E}^e_{11}$, the term $(\mu^o/\eta^o)(S_{11}-S^o_{11})$ (Eq. (38)) has to tend to zero by the high increase of η^o value. Just after this sudden change of direction, a constant compressive strain rate is imposed by the load. The second change of direction (at $t=30$ seconds) provokes another considerable increasing of η^o as in the tensile phase (Figure 2(*b*)). Consequently, the tangent formulation-based model needs more calculation time in comparison with the simplified model due to the adjustment process of the macroscopic viscous tangent modulus η^o at each time step. In conclusion, for uniaxial loading paths, the simplified self-consistent model is more suitable than the tangent formulation-based model from the calculation time point of view. Furthermore, it has been demonstrated by Dingli et al. [44] that the simplified model is more suitable for multiaxial loading paths for which the tangent formulation-based model cannot work.

To understand the role of the parameter α on the kinematic hardening evolution, the simplified interaction law (Eq. (17)) is used as a basis of interpretation. In fact, this parameter has to be determined in such a manner that the instantaneous elastic effect and the viscoplastic relaxation at steady state are appropriately ensured. Hence, three different values of α are chosen ($\alpha = 5.10^{-7}$, 5.10^{-6}, and 10^{-4}) with which two extreme and an intermediate responses vis à vis the overall hardening evolution are recorded. Using a monotonic tensile load for simplicity, the overall, granular, and microscopical responses are recorded and analyzed for each value of α (Fig. 3). Evolution of activated slip systems during the straining is pointed out in Fig. 4. Recorded at the end of loading, elastic and inelastic strains at the granular level are illustrated in Figs. 5 and 6, respectively. For a greatest value of α ($\alpha = 10^{-4}$), a high heterogeneity of the granular inelastic strain is clearly recorded (Fig. 6). When the value of α is relatively high, this provokes stiff inelastic interactions. According to Eq. (17), the term $(\alpha(s^g - S))$ dominates with respect to the other elastic terms in the left-hand side of this interaction law. The instantaneous

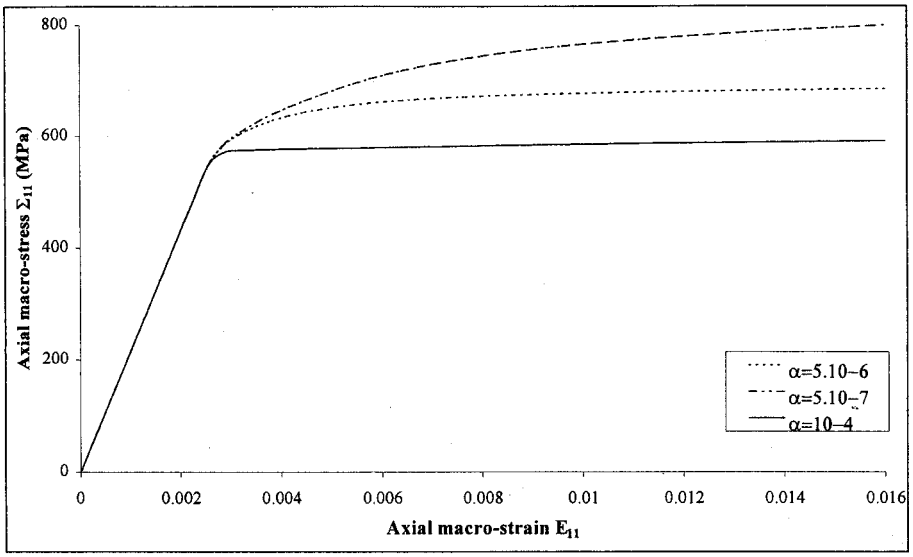


Fig. 3 Overall stress-strain simplified model response showing its sensitivity to the phenomenological parameter (α) under monotonic tensile test

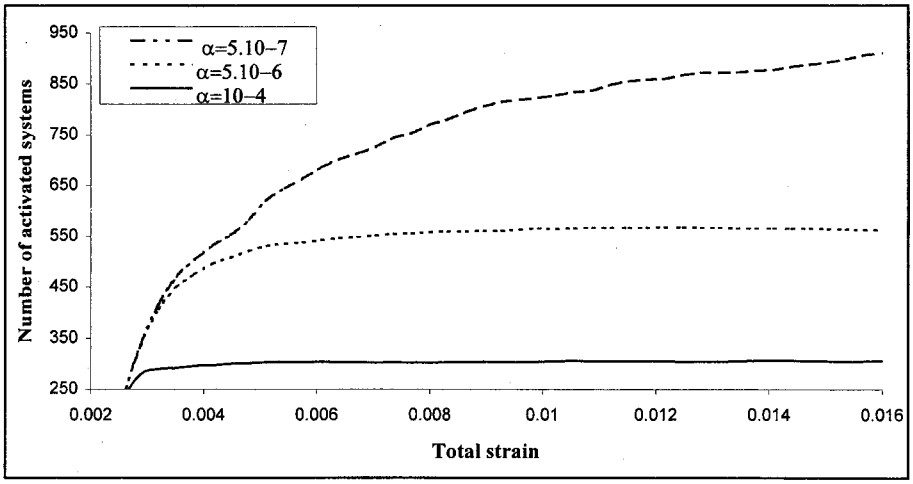


Fig. 4 Effect of the parameter (α) on the evolution of the activated slip systems during monotonic tensile load

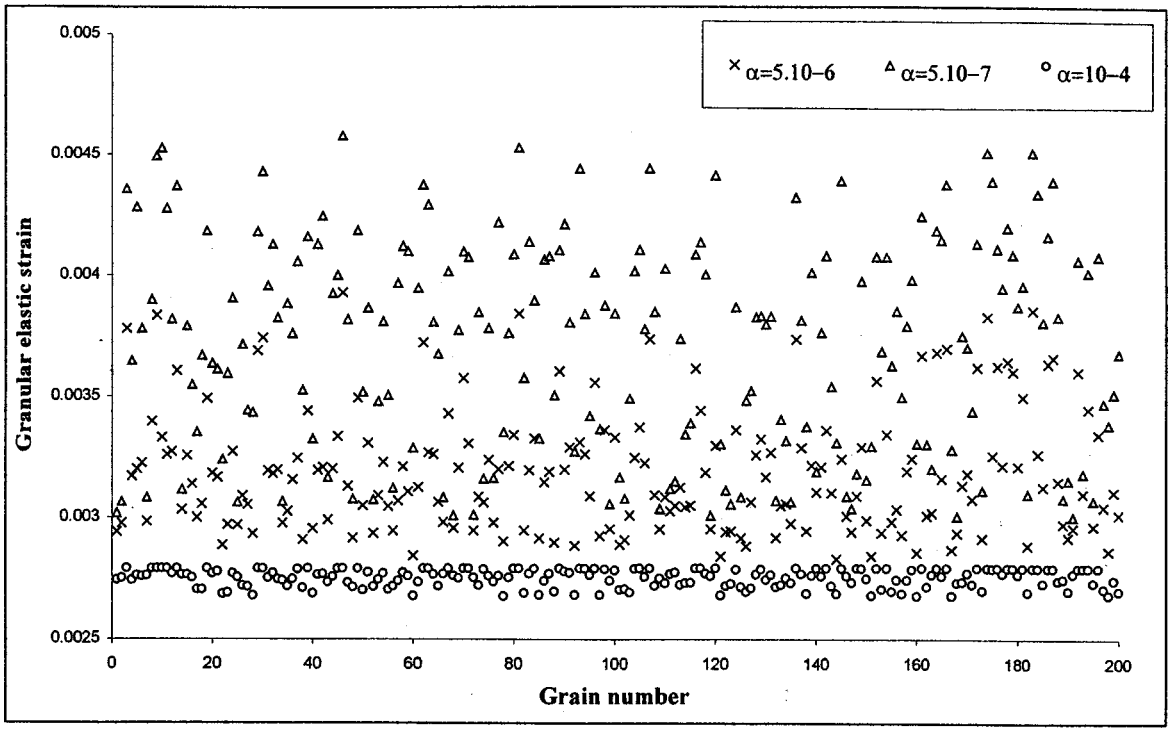


Fig. 5 Effect of the parameter (α) on the heterogeneity of the elastic strain (ε_g^e) in 11 direction of each grain within the used aggregate at the end of the monotonic tensile load

elastic effect becomes consequently negligible giving a considerable heterogeneity of granular inelastic behavior (Fig. 6). However, the granular elastic strain distributions are almost homogenous (Fig. 5) and their values are relatively small leading consequently to few instantaneous activated slip systems (Fig. 4). In the case where $\alpha = 5.10^{-7}$, it can be noticed that the term $(\alpha(s^g - S))$ gives a certain equilibrium between the first two elastic terms in the left-hand side of Eq. (17). The heterogeneity of the instantaneous granular elastic effect giving the progressive increasing of the activated slip systems. Therefore, the granular viscoplastic relaxation at steady-state becomes heterogeneous (Figs. 5 and 6) leading relatively to smooth interactions between the grains and their matrix. As a matter of fact, the existence of the heterogeneous granular elastic behavior in the interaction law leads to a natural description of the global kinematics hardening effect. This can be interpreted by the fact that the intergranular accommodation has an elastic nature (Rougier et al. [27]). In conclusion, the phenomenological parameter α has a great influence on the hardening evolution. To appropriately describe the effect of the hardening on strain-stress behavior of polycrystals, all the qualitative simulations were therefore conducted using $\alpha = 5.10^{-7}$.

4 Remarks and Conclusions

This study is based on the generalized nonincremental interaction law for fully anisotropic and compressible elastic-inelastic

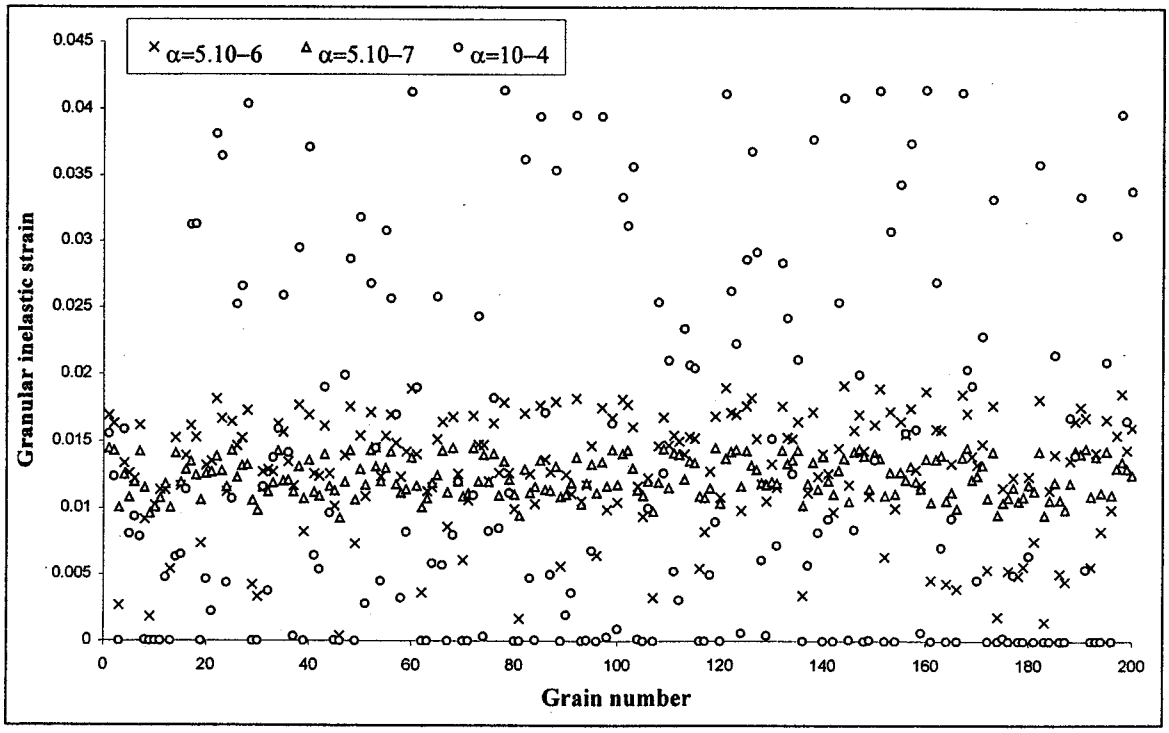


Fig. 6 Effect of the parameter (α) on the heterogeneity of the inelastic strain (ε_g^{in}) in 11 direction of each grain within the used aggregate at the end of the monotonic tensile load

behavior of an isotropic inclusion embedded in an isotropic matrix proposed by Molinari and co-workers [28]. Under small strain condition, the obtained tangent formulation-based and simplified models are numerically compared. It was found that both models give practically the same elastic-inelastic responses of fcc polycrystal under uniaxial loading paths. However, the tangent formulation-based model needs evidently more calculation time in comparison with the simplified model. This is due the fact that macroscopic viscous tangent modulus η^o needs to be adjusted at each time step in order to satisfy the self-consistency condition, while this condition is systematically respected in the case of simplified model.

It is obvious that the tangent formulation-based model is incapable of reproducing appropriately the elastic-inelastic behavior of the polycrystal under multiaxial loading paths. This is due the impossible adjustment of the coefficient η^o since this latter depends only on one strain rate direction and not on two or more directions at the same time. However, the simplified model can successfully describe the polycrystal behavior under multiaxial loadings. From the standpoint of self-consistent approaches, this capability represents a primordial factor.

In this simplified model, a local isotropic hardening variable is modeled and represented on the system level to evaluate its evolution. The interactions between activated slip systems in the same grain can be also taken into account. The intergranular and transgranular kinematic hardening effects can be globally and naturally described (without introducing internal variables) by the simplified interaction law. Moreover, independent on the strain history, the proposed phenomenological constant α in the interaction law has a considerable role in the kinematic hardening evolution.

Under multiaxial loading conditions, the developed model highlights the advantages to describe the additional hardening phenomenon related to the complexity of the loading path. In fact, the activated slip systems can be multiplied as the loading path complexity increases. Qualitative and quantitative results concerning elastic-inelastic behaviors of the polycrystal under several complex loading paths were largely discussed by the authors ([44]).

References

[1] Sachs, G., 1928, "Zur Ableitung einer Fliessbedingung," Z. Ver. Dent. Ing., **72**, p. 734.

[2] Cox, H. L., and Sompmith, D. E., 1937, "Effect of Orientation on Stresses in Single Crystals and of Random Orientation on Strength of Polycrystalline Aggregates," Proc. Phys. Soc. London, **49**, p. 134.

[3] Taylor, G. I., 1938, "Plastic Strain in Metals," J. Inst. Met., **62**, p. 307.

[4] Hershey, A., 1954, "The Elasticity of an Isotropic Aggregate of Anisotropic Cubic Crystals," ASME J. Appl. Mech., **21**, p. 236.

[5] Kröner, E., 1958, "Berechnung der Elastischen Konstanten des Vielkristalls aus den Konstanten des Einkristalls," Z. Phys., **151**, p. 504.

[6] Laws, N., and McLaughlin, R., 1978, "Self-Consistent Estimates for the Viscoelastic Creep Compliance of Composite Materials," Proc. R. Soc. London, Ser. A, **A359**, p. 251.

[7] Kouddane, R., Molinari, A., and Canova, G. R., 1993, "Self-Consistent Modeling of Heterogeneous Viscoelastic and Elasto-Viscoplastic Materials," *Large Plastic Deformation: Fundamentals and Applications to Metal Forming*, C. Teodosiu, J. L. Raphanel, and F. Sidoroff, eds., Mecamat 91, A. A. Balkema, Rotterdam, p. 129.

[8] Brown, G. M., 1970, "A Self-Consistent Polycrystalline Model for Creep Under Combined Stress States," J. Mech. Phys. Solids, **18**, p. 367.

[9] Rice, J. R., 1970, "On the Structure of Stress-Strain Relations for Time Dependent Plastic Deformation in Metals," ASME J. Appl. Mech., **37**, p. 728.

[10] Rice, J. R., 1971, "Inelastic Constitutive Relations for Solids: An Internal-Variable Theory and Its Application to Metal Plasticity," Trans. J. Mech. Phys. Solids, **19**, p. 433.

[11] Hutchinson, J. W., 1976, "Bounds and Self-Consistent Estimate for Creep of Polycrystalline Materials," Proc. R. Soc. London, Ser. A, **A348**, p. 101.

[12] Molinari, A., Canova, G. R., and Ahzi, S., 1987, "A Self-Consistent Approach of the Large Deformation Viscoplasticity," Acta Metall., **35**, p. 2983.

[13] Weng, G. J., 1993, "A Self-Consistent Relation for the Time-Dependent Creep of Polycrystals," Int. J. Plast., **9**, p. 181.

[14] Lebensohn, R. A., and Tomé, C. N., 1993, "A Self-Consistent Anisotropic Approach for the Simulation of Plastic Deformation and Texture Development of Polycrystals: Application to Zirconium Alloys," Acta Metall. Mater., **41**, p. 2611.

[15] Lebensohn, R. A., Tomé, C. N., 1994, "A Self-Consistent Viscoplastic Model: Prediction of Rolling Texture of Anisotropic Polycrystals," Mater. Sci. Eng., A, **A175**, p. 71.

[16] Lin, T. H., 1957, "Analysis of Elastic and Plastic Strain of FCC Crystal," J. Mech. Solids, **5**, p. 143.

[17] Kröner, E., 1961, "Zur Plastichen Verformung des Vielkristalls," Acta Metall., **9**, p. 155.

[18] Budianski, B., and Wu, T. T., 1962, "Theoretical Prediction of Plastic Strains

of Polycrystals,'' *Proc. 4th U.S. Nat. Cong. Appl. Mech.*, ASME, New York, p. 1175.
[19] Hill, R., 1965, ''Continuum Micro-mechanical Elastoplastic Polycrystals,'' J. Mech. Phys. Solids, **13**, p. 89.
[20] Hutchinson, J. W., 1970, ''Elastic-Plastic Behaviour of Polycrystalline Metals and Composites,'' Proc. R. Soc. London, Ser. A, **A319**, p. 247.
[21] Berveiller, M., and Zaoui, A., 1979, ''An Extension of the Self-Consistent Scheme to Plasticity Flowing Polycrystals,'' J. Mech. Phys. Solids, **26**, p. 325.
[22] Weng, G. J., 1982, ''A Unified Self-Consistent Theory for the Plastic-Creep Deformation of Metals,'' ASME J. Appl. Mech., **49**, p. 728.
[23] Iwakuma, T., and Nemat-Nasser, S., 1984, ''Finite Elastic Deformation of Polycrystalline Metals,'' Proc. R. Soc. London, Ser. A, **A394**, p. 87.
[24] Nemat-Nasser, S., and Obata, M., 1986, ''Rate-Dependent, Finite Elasto-Plastic Deformation of Polycrystals,'' Proc. R. Soc. London, Ser. A, **A407**, p. 343.
[25] Lipinski, P., Krier, J., and Berveiller, M., 1990, ''Elastoplasticité des Métaux en Grandes Déformations: Comportement Global et Evolution de la Structure Interne,'' Rev. Phys. Appl., **25**, p. 361.
[26] Lipinski, P., Naddari, A., and Berveiller, M., 1992, ''Recent Results Concerning the Modeling of Polycrystalline Plasticity at Large Strains,'' Int. J. Solids Struct., **92**, p. 1873.
[27] Rougier, Y., Stola, C., and Zaoui, A., 1994, ''Self-Consistent Modelling of Elastic-Viscoplastic Polycrystals,'' C. R. Acad. Sci. Paris, **319**, p. 145.
[28] Molinari, A., Ahzi, S., and Kouddane, R., 1997, ''On the Self-Consistent Modeling of Elasto-Plastic Behavior of Polycrystals,'' Mech. Mater., **26**, p. 43.
[29] Schmitt, C., Lipinski, P., and Berveiller, M., 1997, ''Micromechanical Modelling of the Elastoplastic Behavior of Polycrystals Containing Precipitates—Application to Hypo- and Hyper-eutectoid Steels,'' Int. J. Plast., **13**, p. 183.
[30] Abdul-Latif, A., Dingli, J. Ph., and Saanouni, K., 1998, ''Modeling of Complex Cyclic Inelasticity in Heterogeneous Polycrystalline Microstructure,'' J. Mech. Mater., **30**, p. 287.
[31] Schmid, E., 1924, *Proc. Int. Congr. Appl. Mech*, Delft, p. 342.
[32] Abdul-Latif, A., and Saanouni, K., 1994, ''Damaged Anelastic Behavior of FCC Poly-Crystalline Metals With Micromechanical Approach,'' Int. J. Damage Mech., **3**, p. 237.
[33] Abdul-Latif, A., and Saanouni, K., 1996, ''Micromechanical Modeling of Low Cycle Fatigue Under Complex Loadings—Part II Applications,'' Int. J. Plast., **12**, p. 1123.
[34] Abdul-Latif, A., and Saanouni, K., 1997, ''Effect of Some Parameters on the Plastic Fatigue Behavior With Micromechanical Approach,'' Int. J. Damage Mech., **6**, p. 433.
[35] Saanouni, K., and Abdul-Latif, A., 1996, ''Micromechanical Modeling of Low Cycle Fatigue Under Complex Loadings—Part I. Theoretical Formulation,'' Int. J. Plast., **12**, p. 1111.
[36] Abdul-Latif, A., 1999, ''Unilateral Effect in Plastic Fatigue with Micromechanical Approach,'' Int. J. Damage Mech., **8**, p. 316.
[37] Abdul-Latif, A., Ferney, V., and Saanouni, K., 1999, ''Fatigue Damage of Waspaloy Under Complex Loading,'' ASME J. Eng. Mater. Technol., **121**, p. 278.
[38] Cailletaud, G., 1992, ''A Micromechanical Approach to Inelastic Behaviour of Metals,'' Int. J. Plast., **8**, p. 55.
[39] Dingli, J. P., 1997, ''Modélisation du Comportement Anélastique des Matériaux Polycristallins Hétérogènes,'' *Mémoire de DEA*, Université de Technologie de Compiègne.
[40] François, D., Pineau, A., and Zaoui, A., 1993, *Comportement Mécanique des Matériaux*, Hermens, Paris.
[41] Kocks, U. F., and Brown, T. J., 1966, ''Latent Hardening in Aluminum,'' Acta Metall., **14**, pp. 87–98.
[42] Jackson, P. J., and Basinski, Z. S., 1967, ''Latent Hardening and the Flow Stress in Copper Single Crystal,'' Can. J. Phys., **45**, p. 421.
[43] Franciosi, P., 1978, ''Plasticité à froid des monocristaux C.F.C.: Etude du Durcissement Latent,'' Thèse d'état, Univ. of Paris XIII.
[44] Dingli, J. P., Abdul-Latif, A., and Saanouni, K., 2000, ''Predictions of the Complex Cyclic Behavior of Polycrystals Using a New Self-Consistent Modeling,'' Int. J. Plast., **16**, p. 411.

R. D. Blevins
Goodrich Aerostructures, Inc.,
Mail Stop 107P,
850 Lagoon Drive,
Chula Vista, CA 91910

Non-Gaussian Narrow-Band Random Fatigue

Fatigue is produced by the sum of randomly phased sine waves. Fatigue damage during this non-Gaussian, finite random process is a function of the noise bandwidth, the fatigue slope, mean stress, and the rms stress. Methods are developed for predicting the cycles to failure. Comparison is made with data. [DOI: 10.1115/1.1428332]

1 Introduction

Many processes that cause fatigue are not Gaussian distributed. For example, fatigue damage accumulated by normally operating machinery is the result of a finite number of operational cycles in quasi-random environments. Manufactures guarantees exclude most, if not all, extreme events. The as-designed operation of machinery is a random process with finite limits.

Palmgren and Miner proposed that fatigue damage is only dependent on the peaks in a time history and that fatigue failure occurs when the sum of the fatigue damage reaches unity (Fuchs and Stevens [1], pp. 190–191). By applying their hypothesis, Crandall and Mark ([2], p. 117) obtained an integral expression for the expected number of cycles to fatigue failure, N_F, for a stationary narrow band random process.

$$N_F=\left[\int_0^\infty \frac{p_{S_p}(s)}{N_f(s,S_m)}\,ds\right]^{-1} \tag{1}$$

$p_{S_p}(s)$ is the probability density of peaks in cyclic stress and $N_f(s,S_m)$ is the expected number of cycles to failure of a material during sinusoidal cycling at stress amplitude s. The mean stress is S_m.

Equation (1) has been applied to stationary Gaussian random processes by Wirsching [3], Sobczyk and Spencer, Jr. [4], Bishop and Sherrat [5], Lin [6], Crandall and Mark [2], Powell [7], and Miles [8]. This paper develops methods to predict fatigue damage produced by non-Gaussian stationary narrow-band random processes consisting of the sum of finite amplitude randomly phased sine waves. This is finite random narrow-band pink noise. The single sine wave and the Gaussian process are the limit cases. The results allow the prediction of the fatigue as a function of the number of sine waves (bandwidth, or equivalently, randomness) as a parameter.

Application of Eq. (1) requires (1) a fatigue model, or data, for the number of cycles to failure with constant amplitude cycling and, (2) the probability density of peaks during random cycling. Section 2 presents models for fatigue of materials including mean stress. The probability distribution of the peaks in the finite random noise time history is developed in Section 3. The Sections 4 determines fatigue life in closed form and by numerical integration. Applications and comparisons with data are given in Sections 5 and 6.

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, January 13, 2000; final revision, November 13, 2000. Associate Editor: A. Needleman. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Fatigue Models

Two widely used fatigue models for constant amplitude cycling are (1) the exponential fatigue law and (2) the equivalent stress model. The exponential fatigue law postulates that the stress cycle amplitude above the mean stress is a straight line when plotted on a log-log scale against cycles to failure: $N_f S_a^b = C$ (Basquin [9], Crandall and Mark [2] p. 113, and Bannantine, [10]). $1/b$ is the slope of the fatigue curve when plotted as log (stress amplitude) versus log (cycles to failure). The dimensionless parameter b is a positive real number. Further, Goodman and Soderberg postulated that stress amplitude for fatigue failure is proportional to the difference between the material yield stress and the mean stress (Fuchs and Stephens [1], p. 72). The exponential fatigue model with mean stress effect thus is

$$S_a=(1-S_m/S_y)C^{1/b}N_f^{-1/b}. \tag{2}$$

S_a is the oscillating stress amplitude and S_m is the mean stress. S_y is the stress that causes yield or failure. For fully reversed cycling $S_m=0$. See Fig. 1. Solving Eq. (2) for the number of cycles to fatigue failure (N_f) gives the form required for insertion in Eq. (1).

$$N_f=C(1-S_m/S_y)^b S_a^{-b} \tag{3}$$

In the limit as $b=0$, failure occurs at the Cth cycle regardless of the stress amplitude. In the limit $b=\infty$, the failure stress is independent of the number of cycles.

The equivalent stress fatigue model (MIL-HDBK-5G, 1994, Section 9.3.4.9) gives a relationship between cycles to failure and cyclic stress in terms of four parameters A_1, A_2, A_3, A_4: $\log_{10}N = A_1 + A_2 \log_{10}[(1-R)^{A_3}S_a - A_4]$. The oscillating stress amplitude is

$$S_a=(A_4+10^{-A_1/A_2}N_f^{1/A_2})/(1-R)^{A_3}. \tag{4}$$

The stress ratio R is the ratio of the minimum stress to the maximum stress during one cycle.

$$R=(S_m-S_a)/(S_m+S_a) \tag{5}$$

Equations (4) and (5) can be solved for cycles to failure.

$$N_f=10^{A_1}[2^{A_3}S_a^{A_3}(S_m+S_a)^{1-A_3}-A_4]^{A_2}, \quad S_a>S_{\lim} \tag{6}$$

This is the form required for insertion in Eq. (1).

Equation (6) has an endurance limit stress below which no fatigue damage occurs. The oscillating stress amplitude ($S_{\lim}$) at the endurance limit is found by setting $N_f=\infty$ in Eq. (6) which is equivalent to solving $2^{A_3}S_a^{A_3}(S_m+S_a)^{1-A_3}=A_4$ for the mean or oscillating stress at the endurance limit with the conditions $S_a>0$ and $S_a+S_m>0$. For zero mean stress the endurance limit is $S_{\lim}=A_4/2^{A_3}$. Both the equivalent stress model (Eqs. (4), (6)) and the exponential model (Eqs. (2), (3)) can be solved for mean stress as a function of peak stress and number of cycles to failure.

$$S_m = \begin{cases} S_y(1 - S_a C^{-1/b} N^{1/b}), & \text{exponential model} \quad (7a) \\ (2S_a)^{-A_3/(1-A_3)}[10^{-A_1/A_2} N_f^{1/A_2} + A_4]^{1/(1-A_3)} - S_a, & \\ & \text{equivalent stress} \quad (7b) \end{cases}$$

Tensile mean stress reduces fatigue life; compressive mean stress increase fatigue life. The exponential and equivalent stress fatigue models are identical if there is no endurance limit or mean stress $S_m = A_4 = 0$ so that $R = -1$, $b = -A_2$, and $C = 10^{A_1} 2^{A_2 A_3}$.

Fatigue models are semi-empirical fits to constant-amplitude fatigue data. For aluminum Al 2024-T3 with a notch factor $K_t = 1.5$, $A_1 = 7.5$, $A_2 = -2.13$, $A_3 = 0.66$, $A_4 = 23.7$ ksi (MIL-HDBK-5G [11], p. 3-113) in units of ksi (1 ksi = 1000 lb/sq in. = 6.895 MPa). A_4 has units of stress. With stress in mega Pascals, these are $A_1 = 7.5 + 2.13 \log_{10}(6.895) = 9.286$ and $A_4 = 23.7 * 6.895 = 163.4$ MPa. The units of A_1 are such that $10^{-A_1/A_2}$ has units of stress. A_2 and A_3 are dimensionless.

If fatigue life is known at two points, $N_{f1}, S_{a1}; N_{f2}, S_{a2}$, then the fitted parameters b and C in the exponential fatigue law are $b = -\log_{10}[N_{f2}/N_{f1}]/\log_{10}[S_{a2}/S_{a1}]$, $C = N_{f1} S_{a1}^b$. $b \approx 5$ to 20 for high cycle fatigue of most metals (Manson [12] and Bannantine [10]) and $C \approx S_y^b$ where S_y is the yield or ultimate stress as can be seen by setting $N_f = 1$ and $S_m = 0$ in Eq. (3). For aluminum alloy Al 2024-T3 with Kt = 1.5, fitting the equivalent stress model at $N_f = 5 \times 10^3$ and $N_f = 10^6$ gives $b = 4.911$ and $C = 1.545$E 12 (ksi)$^{4.91}$ [2.030 E16 (MPa)$^{4.911}$]. These curves are shown in Fig. 2.

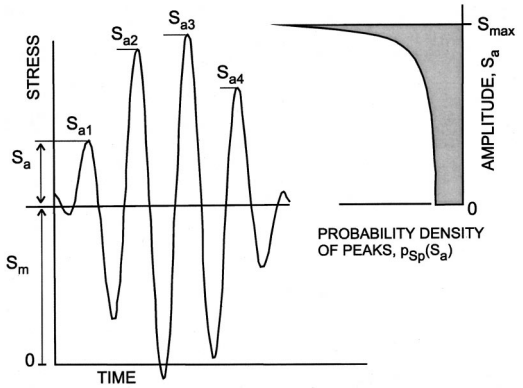


Fig. 1 Time history of a narrow-band random process with mean stress (S_m) and the associated probability density of the peaks. Process shown is the sum of two sine waves.

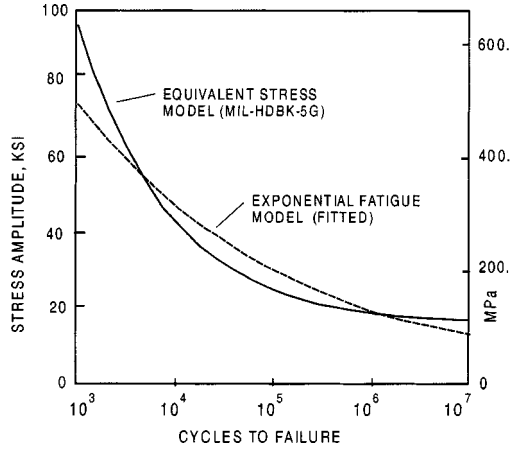


Fig. 2 Fatigue curves for 2024-T3 Al $K_t = 1.5$ with no mean stress and sinusoidal cycling

3 Probability Density of Peaks in Finite Narrow Band Random Noise

The spectral components of a narrow band process are assumed to fall within narrow frequency band, typically one-third-octave band or less wide. Each time the mean value (axis) is crossed with a positive slope, a single peak is generated before the axis is recrossed. Positive peaks only occur above the mean value. The peaks occur at nearly constant rate ([6]) as shown in Fig. 1 for a narrow-band process consisting of the sum of two equal amplitude sine waves.

The elemental narrow-band process is a constant amplitude sinusoidal stress time history.

$$s(t) = a \cos(\omega t + \phi), \quad 0 \leq t \leq T, \quad a \geq 0, \quad 0 \leq \phi < 2\pi. \quad (8)$$

The probability that the amplitude (a) of stress cycle peak falls in the range between s and $s + ds$ is zero unless that range includes the amplitude a. Thus the probability density of the amplitude of a constant amplitude sine wave is a Dirac delta function.

$$p_{S_p}(s) = \delta[s(t) - a] \quad (9)$$

Substituting this expression into Eq. (1) reproduces the fatigue law for constant amplitude sinusoidal cycling.

Consider finite narrow-band noise that is the sum of M equal amplitude ($a_m = a, m = 1, M$), randomly phased sine waves whose frequencies fall in a relatively narrow band $\omega_m \approx \omega$, $m = 1, M$.

$$s(t) = \sum_{m=1}^{M} a_m \cos(\omega_m t + \phi_m), \quad 0 \leq t \leq T,$$
$$a_m \geq 0, \quad 0 \leq \phi_m < 2\pi. \quad (10)$$

The amplitudes, a_m, are positive constants. The circular frequency of each sine wave, ω_m, is a positive, nonzero integer multiple of $2\pi/T$ where T is the sampling time interval. The number of sine waves in the series, M, is a measure of the bandwidth or randomness of the noise.

The phases ϕ_m in Eq. (10) are independent random variables that are uniformly distributed over the range $0 \leq \phi_m < 2\pi$.

$$p_{\phi_m}(\phi) = \begin{cases} 1/(2\pi), & \text{if } 0 \leq \phi < 2\pi \\ 0, & \text{if } 0 > \phi \text{ or } \phi > 2\pi \end{cases} \quad (11)$$

The maximum value of the oscillating component $s(t)$, Eq. (10), is the sum of the amplitudes of the sine waves (recall $a_m \geq 0$).

$$S_{\max} = \sum_{m=1}^{M} a_m \quad (12a)$$

$$= Ma, \quad \text{for } a_m = a, \quad m = 1,2,..M. \quad (12b)$$

The root-mean-square (rms) of the sum of mutually independent sine waves over the ensemble of random independent sine waves is the square root of the sum of the mean squares of the terms.

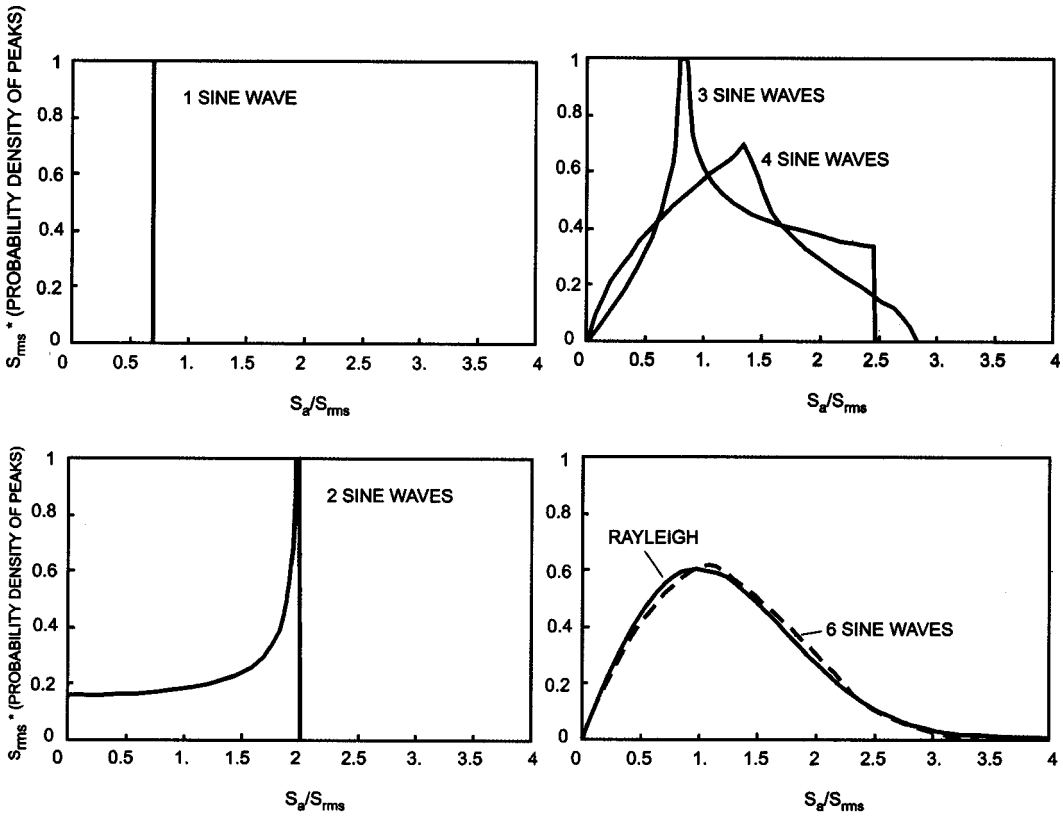


Fig. 3 Probability density of amplitude for sum of randomly phased sine waves (Table 2)

Table 1 Fourier coefficients B_i of M-equal-amplitude sine probability distribution. Case 4 of Table 2.

M	2	3	4	5	6	8
$\frac{S_{max}}{S_{rms}}$	2	2.45	2.83	3.16	3.46	4
B_1	1.03613	1.47952	1.54854	1.56185	1.53894	1.46782
B_2	−0.25911	0.10399	0.33872	0.62573	0.84299	1.14488
B_3	0.50665	0.13080	−0.2551E−1	−0.2748E−1	0.9322E−1	0.30870
B_4	−0.22002	−0.34356	−0.6865E−1	0.1071E−1	−0.1879E−2	0.2405E−1
B_5	0.36795	−0.4612E−1	0.11990	0.02888E−1	−0.4861E−2	−0.6822E−4
B_6	−0.19353	−0.7622E−1	−0.5270E−1	−0.3711E−1	−0.1235E−1	−0.4310E−3
B_7	0.30192	0.19239	−0.1186E−1	−0.4610E−1	0.8122E−2	−0.8269E−3
B_8	−0.17112	0.2985E−1	−0.3011E−1	−0.4037E−2	0.2378E−1	−0.1955E−2
B_9	0.26396	0.5330E−1	0.4651E−1	−0.3387E−2	0.4599E−2	−0.5061E−3
B_{10}	−0.15644	−0.13365	0.2257E−1	−0.8987E−2	−0.5762E−3	0.4020E−2
B_{11}	0.23347	−0.2219E−1	−0.7055E−2	0.1019E−1	−0.1168E−2	0.2896E−2
B_{12}	−0.14555	−0.4076E−1	−0.1767E−1	0.1525E−1	−0.2870E−2	0.3765E−3
B_{13}	0.21531	0.10286	0.2601E−1	0.1600E−2	0.1294E−2	−0.3401E−4
B_{14}	−0.13510	0.1801E−1	0.1312E−1	0.1589E−2	0.5600E−2	−0.5775E−4
B_{15}	0.19753	0.3304E−1	−0.4767E−2	0.4262E−2	0.1160E−2	−0.1051E−3

(a) **Computed with 20 terms in each sum.** $B_i(M)=2\pi^2 i\sum_{k=0}^{\infty} g_{ik}[J_0(\pi/M\sqrt{i^2+k^2})]^M$

$$g_{ik}=\begin{cases}1/8, & i=k=0;\\ 1/4, & i>0,k=0;\\ (1/2)[(-1)^k-1]/(k\pi)^2, & i=0;k>0,\\ [(-1)^k-1]/(k\pi)^2, & i>0,k>0\end{cases}$$

Table 2 Probability densities for peaks in finite narrow band random noise with *M* equal amplitude spectral components, Eq. (10) ([13]). *a*=amplitude of spectral component. *M*=number of spectral components.

1 **One Sine Wave, *M*=1.** $a=2^{1/2}S_{\rm rms}$. $p_{S_p}(s)=\delta[s(t)-a]$

$\int_{a-\epsilon}^{a+\epsilon}p_{S_p}(s)ds=1$ for $\epsilon>0$.

2 **Two Sine Waves, *M*=2.** $a=S_{\rm rms}$,

$$p_{S_p}(s)=\begin{cases}\dfrac{2}{\pi\sqrt{(2a)^2-s^2}}, & 0\leqslant s\leqslant 2a\\ 0, & s>2a \text{ or } s<0.\end{cases}$$

$\int_{2a-\epsilon}^{2a}p_{S_p}(d)ds\approx(2/\pi)(\epsilon/a)^{1/2}$ for $\epsilon/a\ll 1$.

3 **Three Sine waves, *M*=3**. $a=(2/3)^{1/2}S_{\rm rms}$. Envelope distribution in terms of the complete elliptic integral *K*, discussed in the Appendix.

$$p_{S_p}(s)=\begin{cases}\dfrac{4s}{\pi^2}\dfrac{1}{\sqrt{[4a^2-(s-a)^2](s+a)^2}}K\left(\sqrt{\dfrac{16sa^3}{[4a^2-(s-a)^2](s+a)^2}}\right), & 0\leqslant s\leqslant a,\\ \dfrac{4s}{\pi^2}\dfrac{1}{\sqrt{16sa^3}}K\left(\sqrt{\dfrac{[4a^2-(s-a)^2](s+a)^2}{16sa^3}}\right), & a\leqslant s\leqslant 3a,\\ 0, & s>3a \text{ or } s<0.\end{cases}$$

$\int_{a-\epsilon}^{a+\epsilon}p_{S_p}(s)ds\approx(2/\pi^2)(\epsilon/a)(3+\ln(16(a/\epsilon)^3))$ for $\epsilon/a\ll 1$.

4 ***M* Sine Waves, *M*=2, 3, 4, 5... .** $a=(2/M)^{1/2}S_{\rm rms}$.

$$p_{S_p}(s)=\begin{cases}\dfrac{1}{Ma}\displaystyle\sum_{i=0}^{\infty}B_i(M)\sin((i\pi s)/(Ma)), & 0\leqslant s\leqslant Ma,\\ 0, & s>Ma \text{ or } s<0.\end{cases}$$

where Table 1 gives values of B_i. See Blevins [20] for nonequal amplitude case. An alternate expression is a Cramer power series valid for $s<Ma$,

$$p_{S_p}(s)=\frac{s}{S_{\rm rms}^2}e^{-s^2/2S_{\rm rms}^2}\left(1-\frac{1}{2M}\left(1-\frac{s^2}{S_{\rm rms}^2}+\frac{s^4}{8S_{\rm rms}^4}\right)+\frac{1}{12M^2}\left(1-\frac{6s^2}{S_{\rm rms}^2}+\frac{15s^4}{4S_{\rm rms}^4}-\frac{7s^6}{12S_{\rm rms}^6}+\frac{3s^8}{128S_{\rm rms}^8}\right)\right.$$
$$\left.-\frac{11}{8M^3}\left(1-\frac{2s^2}{S_{\rm rms}^2}+\frac{3s^4}{4S_{\rm rms}^4}-\frac{s^6}{12S_{\rm rms}^6}+\frac{s^8}{384S_{\rm rms}^8}\right)\right)+O(1/M^4).$$

Both series converges most rapidly for $M>2$.

5 **Infinite Sine Waves, *M*=∞**, the Rayleigh distribution ([2]),

$$p_{S_p}(s)=\frac{s}{S_{\rm rms}^2}e^{-s^2/(2S_{\rm rms}^2)},\quad 0\leqslant s<\infty.$$

$$S_{\rm rms}^2=\frac{1}{T}\int_0^{2\pi}\left[\sum_{m=1}^{M}a_m\cos(\omega_m t/T+\phi_m)\right]^2dt=\frac{1}{2}\sum_{m=1}^{M}a_m^2 \tag{13a}$$

$$=\tfrac{1}{2}Ma^2,\quad \text{for } a_m=a,\quad m=1,2,..M \tag{13b}$$

Note the rms and the maximum defined by Eqs. (12) and (13) do not include the mean (constant) stress. The maximum oscillating stress is $S_{\rm max}=Ma=(2M)^{1/2}S_{\rm rms}$ is all *M* amplitudes are equal.

The maximum-to-rms ratio, which is a measure of randomness, increases with the number of component sine waves in Eq. (10).

$$\frac{S_{\rm max}}{S_{\rm rms}}=2^{1/2}\sum_{m=1}^{M}a_m\Big/\left(\sum_{m=1}^{M}a_m^2\right)^{1/2}, \tag{14a}$$

$$=(2M)^{1/2},\quad \text{for } a_1=a_2=a_m=a,\quad m=1,2,..M. \tag{14b}$$

Since *M* is finite the probability of *S* exceeding the sum of the amplitudes (Eq. (12)) is zero. For example for $M=4$ process with equal amplitudes, there is no chance the ratio $S/S_{\rm rms}$ will exceed 2.828. See Fig. 3. Table 2 provides a summary the probability density of peaks in finite narrow-band noise with equal amplitude spectral components. There are simple expressions for $M=1$, 2, 3, and ∞. The $M=1$, 2, and 3 probability densities have singular points at $S/S_{\rm rms}=2^{1/2}$, 2 and 3, respectively, but like all probability density functions, the area under the probability density curve is unity.

The probability density of peaks for the process of (Eq. (10)) for arbitrary integer *M* is given in Case 4 of Table 2 by two equivalent series: (1) a Fourier series over the range of nonzero probabilities, and (2) a Cramer power series which is adapted from Blevins [13] but with corrections and additional terms. As *M* approaches infinity, the central limit theorem implies that the process becomes Gaussian and the distribution of peaks approaches the Rayleigh distribution ([2]). This can be seen in Table 2, Case 4b, in Cramer series by letting *M* approach infinity and in Case 3 of Table 3 by setting $\gamma(1+b/2,\infty)=\Gamma(1+b/2)$.

Table 3 Solutions for cycles to failure with exponential fatigue, Eq. (15)

Sine Waves, M	Cycles to Failure, N_F
1	$C\left(1-\frac{S_m}{S_y}\right)^b S_{\rm rms}^{-b}2^{-b/2}$, or equivalently, $C\left(1-\frac{S_m}{S_y}\right)^b S_a^{-b}$
2 (a)	$C\left(1-\frac{S_m}{S_y}\right)^b S_{\rm rms}^{-b}\pi^{1/2}2^{-b}\Gamma(1+b/2)/\Gamma(1/2+b/2)$
3,..M (a,b)	$C\left(1-\frac{S_m}{S_y}\right)^b S_{\rm rms}^{-b}2^{-b/2}\left[\gamma\left(\frac{2+b}{2},M\right)\times\left(1-\frac{1}{2M}+\frac{1}{12M^2}-\frac{11}{8M^3}\right)\right]^{-1}$
∞ (a,c)	$C\left(1-\frac{S_m}{S_y}\right)^b S_{\rm rms}^{-b}2^{-b/2}/\Gamma(1+b/2)$

[a] See Appendix for definition and evaluation of $\gamma(a,x)$ and $\Gamma(a)$.
[b] Cramer series, Case 4(b) of Table 2 used. Increasing number of terms in series apparently does not substantially increase accuracy.
[c] Crandall and Mark [2].

4 Cycles to Failure

The expected number of cycles to failure of a narrow band random process is found by substituting the peak probability density of the peaks (Table 2) and the sinusoidal cycling fatigue failure model (Eqs. (3) or (6)) into Eq. (1) and integrating.

Substituting Eq. (3) into Eq. (1) gives the following equation for cycles to failure for the exponential fatigue law:

$$N_F=C\left(1-\frac{S_m}{S_y}\right)^b\left[\int_0^{S_{\max}} s^b p_{S_p}(s)\,ds\right]^{-1}. \tag{15}$$

The integration is made over the range of nonzero probabilities. Table 3 lists exact solutions to Eq. (15) using the exact probability densities of amplitude given in Table 2 (Gradshteyn, Ryzhik, and Jeffrey [14]). The random cycles to failure is proportional to $S_{\rm rms}^{-b}$. The Cramer power series for probability is used for the general case. These solutions are in terms of two special mathematical functions, the gamma function $\Gamma(x)$ and the partial gamma function $\gamma(x,a)$, that are discussed in the Appendix.

Equation (1) is numerically integrated to obtain solutions for other fatigue laws or probability densities. Using the Fourier series probability density representation (Table 2, Case 4b) and the exponential fatigue law and Simpson's's numerical integration, this is

$$N_F\approx C\left(1-\frac{S_m}{S_y}\right)^b S_{\rm rms}^{-b}\left[\frac{(2M)^{b/2}}{K}\times\sum_{k=1}^{K}\sum_{i=1}^{I}\left(\frac{k}{K}\right)^b B_i(M)\sin\left(\frac{i\pi k}{K}\right)\right]^{-1}. \tag{16}$$

The stress range has been divided into K intervals with $ds=\Delta s=S_{\max}/K$.

With the equivalent stress fatigue model, numerical integration is carried out over the stress amplitude range between the endurance stress amplitude ($S_{\lim}$—see Eq. (6) with $N_f=\infty$) and the maximum stress $S_{\max}$ (Eq. (12)) by dividing this range is divided into K stress intervals of size $\Delta s=(S_{\max}-S_{\lim})/K$. Simpson's rule of numerical integration is applied. With narrow-band finite random noise and the equivalent stress model (Eq. (6)) the result is

$$N_F=\left[\frac{1}{Ma}\int_{S_{\lim}}^{Ma}\sum_{i=0}^{\infty}\left(B_i(M)\sin\left(\frac{i\pi s}{Ma}\right)\Big/\,10^{A_1}[2^{A_3}s^{A_3}(S_m+s)^{1-A_3}-A_4]^{A_2}\right)ds\right]^{-1} \tag{17a}$$

$$\approx\left[\frac{Ma-S_{\lim}}{10^{A_1}KMa}\sum_{k=0}^{K-1}\sum_{i=0}^{I}\frac{B_i(M)\sin(i\pi(S_{\lim}+k\Delta s)/(Ma))}{[2^{A_3}(S_{\lim}+k\Delta s)^{A_3}(S_m+S_{\lim}+k\Delta s)^{1-A_3}-A_4]^{A_2}}\right]^{-1}. \tag{17b}$$

Good numerical accuracy requires $K=20$ or more integration intervals and I approximately 10 to 15 for $M=3,4,5\ldots$. Note $a=(2S_{\rm rms}/M)^{1/2}$.

The exact probability densities for the $M=2$ and $M=3$ cases (Cases 2 and 3 of Table 2) give better accuracy than their series approximation (Case 4 of Table 2). The numerical integration of Eq. (1) across singularities in these distributions can be accomplished by creating small stress interval ϵ wide about the singularities at $s=2a$ and $s=a$ in the $M=2$ and $M=3$ distributions, respectively, and integrating the probability densities across the singularities while holding $N_f=$constant, as indicated in Table 2. In general, numerical evaluation using Fourier series representation of probability density, Eqs. (16) and (17), gives better accuracy than using the general Cramer series solution, Case 3 of Table 3.

5 Randomness and Fatigue Life

We can define a sinusoidal process which will generate the same average fatigue damage per cycle as a narrow band random process by setting the number of cycles to failure of a random process (Eq. (1)) equal to that of a sinusoidal process (Eq. (3) or Eq. (6)) and solving for the sine amplitude. For the exponential fatigue law and equal-amplitude random sinusoidal terms in the process, the equally damaging sinusoidal amplitude is

$$S_a|\text{equal}-\text{damage}=\left[\int_0^{\infty}s^b p_{S_a}(s)\,ds\right]^{1/b}, \tag{18}$$

The right-hand side is only a function of the slope b of the fatigue curve and the bandwidth M. The ratio of the equally damaging sinusoidal amplitude to the rms stress for $M=1$ (sinusoidal time history), $M=2$, $M=5$, $M=20$, and $M=\infty$, Eq. (36) are given Table 4. Numerical integration of Eq. (18) was used for $M=3$ in this table. For $M=1$ or a very steeply dropping fatigue curve, $b<3$, the equivalent amplitude is $2^{1/2}$ times the rms; it is larger for more components M and flatter fatigue curves.

Following Crandall and Mark [2], the damage per vibration cycle is defined as the rms stress times the probability density of the stress amplitude, divided by the number of cycles to failure at that stress amplitude. The most damaging stress with the Rayleigh distribution and the exponential fatigue law is $(1+b)^{1/2}S_{\rm rms}$. Stress amplitudes up to twice this value contribute significantly to

Table 4 Ratio of equally damaging sinusoidal amplitude stress to random RMS stress as function of random bandwidth (*M*) and fatigue slope (*b*)

	$S_a/S_{\text{rms-random}}$						
	M						
b	1	2	3	5	10	20	∞
2	1.4142	1.4142	1.4185	1.4143	1.4142	1.4142	1.4142
3	1.4142	1.5030	1.5214	1.5349	1.5451	1.5500	1.5550
4	1.4142	1.5651	1.6073	1.6383	1.6605	1.6712	1.6818
6	1.4142	1.6475	1.7380	1.8078	1.8583	1.8823	1.9064
8	1.4142	1.7007	1.8308	1.9419	2.0248	2.0642	2.1039
10	1.4142	1.7384	1.8999	2.0504	2.1689	2.2251	2.2826
15	1.4142	1.7981	2.0148	2.2477	2.4624	2.5637	2.6730
20	1.4142	1.8337	2.0862	2.3805	2.6929	2.8378	3.0096
30	1.4142	1.8751	2.1719	2.5491	3.3039	3.2039	3.5842
50	1.4142	1.9144	2.2561	2.7240	3.4565	4.1482	4.5116

the overall damage accumulation as shown in Fig. 4. So for the flatter fatigue curves or broader probability densities, the most damaging stress is pushed towards the rare, high amplitude, stress cycles.

As noted by Wolfe [15], the random vibration amplitudes produced by standard electromechanical shakers are not Rayleigh distributed beyond three to four times the rms. Figure 5 shows the results of fatigue testing a 321 austenitic steel tube with a fitting on a shaker (Pozefsky [16]). Two sets of test were made: (1) with sinusoidal input, and (2) using narrow band random noise as input to the shaker. The sinusoidal data was fitted with an exponential fatigue law giving $S_a = 151 * N_f^{-0.131}$, ksi. dividing the stress by $2^{1/2}$ gives the sinusoidal result in terms of rms $S_{\text{rms-sine}} = 107 * N_f^{-0.131}$, ksi rms. Fitting the random test results with the same exponent gave $S_{\text{rms-random}} = 79.9 * N_F^{-0.131}$, ksi rms. The magnitude corresponds to a sine-peak-to random rms value of 151.3/79=1.91. Interpolating in Table 4 this gives $M \approx 5$, rather than $M = \infty$ which would give the Rayleigh result.

A second comparison of fatigue data taken with both random and sinusoidal cycling is given for annealed titanium Ti-6Al-4V. Schneider's [17] fit to his random fatigue data is $N_F = 10^{14.6} * S_{\text{rms}}^{-5.99}$ whereas the sinusoidal data of MIL-HDBK-5G ([11] p. 5-68) gives $N_f = 10^{19.18} S_a^{-7.55}$. This implies $b = 7.55$. By equating cycles to failure, the ratio of the sinusoidal peak stress to the random rms stress is found as follows:

Cycles to Failure	10^4	10^5	10^6	10^7
S_a/S_{rms}	1.74	1.88	2.04	2.21

Comparing this table with Table 4 using $b \approx 8$ suggests that a finite random noise with $M = 4$ to 15 is an appropriate model for this random process. It also suggests that the Rayleigh distribution is conservative for computation of random fatigue and that "3-sigma" criteria ([3]), $S_{\text{equiv}}/S_{\text{rms}} = 3$, is overly conservative for all but the flattest fatigue curves.

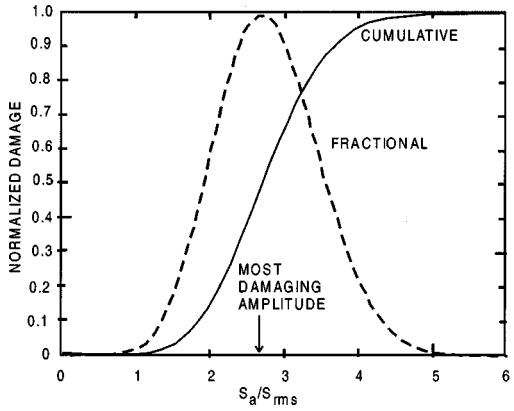


Fig. 4 Fractional (Eq. (21)) and cumulative damage with Rayleigh distribution. Curves have been normalized to a maximum of unity. Material is Al 7075-T6, $K_t = 2$, $b = 6.15$.

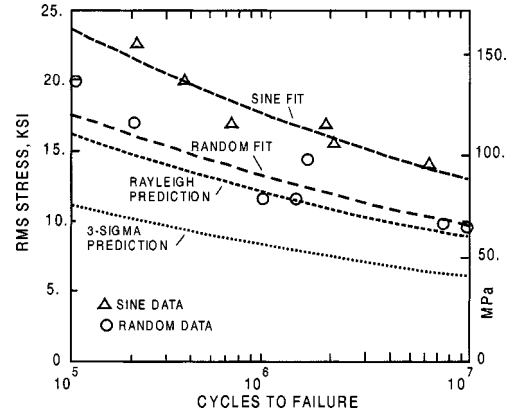


Fig. 5 Sinusoidal and random fatigue data for austenitic stainless steel 321 in comparison with exponential fatigue law fits and predictions. Random fit is identical to $M = 5$ prediction.

6 Example

Application is made for the aluminum alloy Al 2024-T3 sheet with a notch $K_t = 1.5$ at room temperature ([11], p. 3-113—also

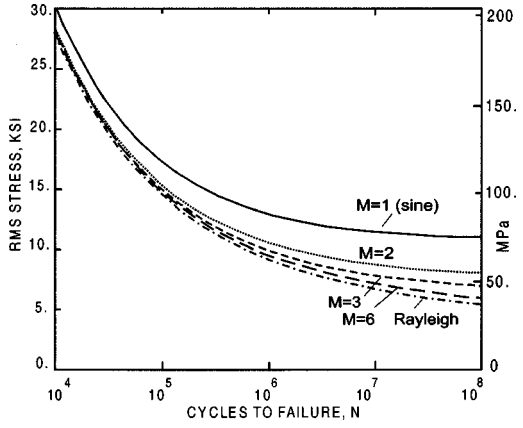


Fig. 6 Effect of increasing randomness of fatigue of Al 2024-T3 with $K_t = 1.5$ and no mean stress. *N* is number of randomly phased sine waves in the process.

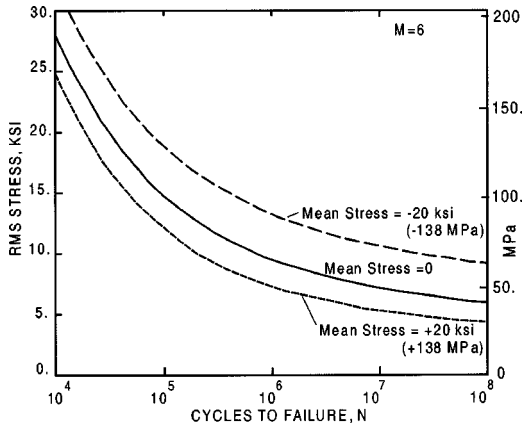


Fig. 7 Effect of mean stress on $M=6$ finite random noise induced fatigue of Al 2024-T3 with $K_t=1.5$. Note RMS stress refers to oscillating component only.

see Section 2). The finite random fatigue curves for zero mean stress are shown in Fig. 6 for sinusoidal oscillation ($M=1$, $S_{\max}/S_{\rm rms}=2^{1/2}$), random oscillations with two spectral components ($M=1$, $S_{\max}/S_{\rm rms}=2^{1/2}$), six spectral components ($M=6$, $S_{\max}/S_{\rm rms}=3.46$), and the Rayleigh distribution which is achieved in the limit of infinite spectral components ($S_{\max}/S_{\rm rms}=\infty$). Fatigue life decreases as random bandwidth of the process (M) increases. Figure 7 shows that a tensile mean stress decreases random fatigue life.

7 Conclusions

Analysis has been made to determine the fatigue life of materials for stationary finite narrow-band random noise consisting of the sum of a finite number of randomly phased sine waves. The following are the conclusions of the analysis.

1 Fatigue damage during this non-Gaussian, finite, pink noise is a function of the noise bandwidth, that is, the number of component sine waves in the sum, the fatigue slope, mean stress, and the rms stress.

2 A method of computing cycle to failure for a finite narrow band noise consisting of the sum of randomly phased sine waves was developed.

3 The fatigue damage for a given rms stress increases with the random band width of the process. The results fall between two limit processes, the single sine wave, and the Rayleigh distribution.

4 Comparison with data suggest that extreme amplitudes of normally operating machinery are not Rayleigh distributed and that finite random pink noise can provide a more representative and less conservative model than the Rayleigh distribution or the 3-sigma criterion.

Nomenclature

$A_1,..A_4$ = parameters in equivalent amplitude stress model, Eq. (4)
a = amplitude of sinusoidal stress oscillation
a_m = amplitude of mth sinusoidal stress oscillation
B_i = coefficient in expansion for probability density of amplitude, Table 2, Case 4
b = inverse of slope of exponential fatigue model on log-log plot
C = exponential fatigue model parameter, Eq. (3)
$J_o(x)$ = Bessel function of first kind and zero order (see Appendix)
I, i, K, k = counting integers
$K(x)$ = complete elliptic integral of first kind (see Appendix)
K_t = fatigue stress notch factor
$\ln(x)$ = natural logarithm
M = integer number of sinusoidal components in process
m = integer index of sinusoidal component
N_F = expected number of cycles to failure for random stress cycle
N_f = expected number of cycles to failure for sinusoidal stress cycle
$p_\phi(\phi)$ = probability density of phase
$p_{sp}(s)$ = probability density of stress amplitude s
R = ratio of maximum to minimum stress over one cycle
S, s = stress
S_a = amplitude of cyclic stress
$S_{\rm lim}$ = stress amplitude below which no fatigue damage occurs
S_m = mean stress
$S_{\max}$ = maximum value of oscillating stress
$S_{\rm rms}$ = root mean square stress
S_y = yield stress
t = time
$\delta(x)$ = Dirac delta function
$\Gamma(x)$ = gamma function (see Appendix)
$\gamma(a,x)$ = partial gamma function (see Appendix)
ϕ = phase angle
ω = circular frequency

Appendix

Evaluation of Special Functions. The gamma function $\Gamma(x)$ and the incomplete gamma function $\gamma(a,x)$ are defined by the following expressions (Abramowitz and Stegun [18], pp. 255, 260):

$$\Gamma(a)=\int_0^\infty e^{-t}t^{a-1}dt$$

$$\gamma(a,x)=\int_0^x e^{-t}t^{a-1}dt=\Gamma(a)-\Gamma(a,x)$$

$$\text{where}\quad \Gamma(a,x)=\int_x^\infty e^{-t}t^{a-1}dt.$$

There are relatively simple expressions available for evaluation of the gamma function (Abramowitz and Stegun [18], p. 255).

$$\Gamma(x)=\begin{cases}\approx(2\pi)^{1/2}x^{x-1/2}e^{-x}(1+1/(12x)+\ \ldots), & x>1\\ (n-1)!, \quad \text{integer } x=n=1,2,3\ \ldots\end{cases}.$$

The incomplete gamma function is related to the area of the upper tail of the statistical chi squared distribution Q (Abramowitz and Stegun [18], pp. 940–941).

$$\gamma(a,x)=\Gamma(a)*(1-Q(2x|2a))$$

Tables of the chi-squared distribution are available in statistical and mathematical literature ([17]), for example.

These functions are available in software. For example in Mathematica [19], $\Gamma(a)=$Gamma$[a]$ and $\gamma(a,x)=$Gamma$[a,0,x]$. In EXCEL 2000 ([20]) spreadsheet they are expressed in terms of related functions, $\Gamma(a)=$EXP(GAMMALN(a)) and $\gamma(a,x)$ =EXP(GAMMALN(a))*(1−CHIDIST(2*x,2*a)).

The Bessel Function of zero order $J_o(x)$ has a series forms. The complete elliptic integral of first kind $K(x)$ has approximate expression (Abramowitz and Stegun [18], pp. 360, 591).

$$J_o(x)=\sum_{k=0}^{\infty}\frac{(-\frac{1}{4}x^2)^k}{k!\Gamma(k+1)}$$

$$K(k)=\int_0^1 [(1-u^2)(1-k^2u^2)]^{-1/2}du, \quad 0<k<1;$$

$$\approx(\alpha_0+\alpha_1k_1^2+\alpha_2k_1^4)+(\beta_0+\beta_1k_1^2+\beta_2k_1^4)\ln(1/k_1^2).$$

$k_1=(1-k^2)^{1/2}$ and $\alpha_0=1.38629$, $\alpha_1=0.11197$, $\alpha_2=0.07252$, $\beta_0=0.5$, $\beta_1=0.12134$, $\beta_2=0.02887$. These functions are tabulated and available in software. In Mathematica they are $J_o(x)$ = Bessel $J[0,x]$, $K(x)$ = Elliptic $K[x^2]$. In EXCEL, $J_o(x)$ = BESSEL $J[x,0]$.

References

[1] Fuchs, H. O., and Stephens, R. I., 1980, *Metal Fatigue in Engineering*, John Wiley and Sons, New York.

[2] Crandall, S. H., and Mark, C. H., 1963, *Random Vibrations in Mechanical Systems*, Academic Press, San Diego, CA.

[3] Wirsching, P. H., Paez, T. L., and Ortiz, K., 1995, *Random Vibrations, Theory and Practice*, Wiley-Interscience, NY, pp. 162–166.

[4] Sobczyk, K., and Spencer, Jr., B. F., 1992, *Random Fatigue From Data to Theory*, Academic Press, San Diego, CA.

[5] Bishop, N. W. M., and Sherrat, F., 1989, "Fatigue Life Prediction for Power Spectral Density Data, Part 1: Traditional Approaches; Part II: Recent Developments," Envir. Engr., **2**, pp. 11–29.

[6] Lin, P. K., 1976, *Probabilistic Theory of Structural Dynamics*, Krieger, Melbourne, FL (reprint of 1967 edition with corrections).

[7] Powell, A., 1958, "On the Fatigue Failure of Structure due to Vibrations Excited by Random Pressure Fields," J. Acoust. Soc. Am., **30**, pp. 1130–1135.

[8] Miles, J., 1954, "On Structural Fatigue Under Random Loading," J. Aeronaut. Sci., **21**, pp. 753–762.

[9] Basquin, O. H., 1910, "The Exponential Law of Endurance Tests," *Proceedings Thirteench Annual Meeting*, Vol. X, ASTM, Philadelphia, PA, pp. 625–630.

[10] Bannantine, J. A., Comer, J. J., and Handrock, J. L., 1990, *Fundamental of Metal Fatigue Analysis*, Prentice-Hall, Englewood Cliffs, NJ, pp. 59, 64.

[11] MIL-HDBK-5G, 1994, "*Metallic Materials and Elements for Aerospace Vehicle Structures*," US Department of Defense.

[12] Manson, S. S., 1966, *Thermal Stress and Low-Cycle Fatigue*, McGraw-Hill, New York, p. 159.

[13] Blevins, R. D., 1997, "Probability Density of Finite Fourier Series with Random Phases," J. Sound Vib., **208**, pp. 617–652.

[14] Gradshteyn, I. S., Ryzhik, I. M., and Jeffrey, A., 1994, *Table of Integrals, Series, and Products*, 5th Ed., Academic Press, San Diego, CA, Articles 3.251, 3.952, 8.338, 8.384.

[15] Wolfe, H. F., Camden, M. P., and Simmons, L. W., 1996, "Six Sigma Effects on the Random Response of a Cantilevered Beam with Random Excitation," *Proceedings of the 67th Shock and Vibrations Symposium*, Vol. I, Monterey, CA, pp. 499–508.

[16] Pozefsky, P., Blevins, R. D., and Laganelli, A., 1989, "Thermo-Vibro Acoustic Loads and Fatigue of Hypersonic Flight Vehicle Structure," AFWAL-TR-3014, Flight Dynamics Laboratory, Wright-Patterson Air Force Base, OH.

[17] Schneider, C. W., 1974, "Acoustic Fatigue of Aircraft Structures at Elevated Temperatures," AFFDL-TR-73-155, Part 1, pp. 38, 41.

[18] Abramowitz, M., and Stegun, I. A., 1964, *Handbook of Mathematical Functions*, National Bureau of Standards, Washington DC (reprinted by Dover).

[19] Mathematica, 1997, Ver 3.0, Wolfram Research, Champaign, IL.

[20] Microsoft Corporation, 1999, EXCEL 2000.

Open-Loop Nonlinear Vibration Control of Shallow Arches via Perturbation Approach

W. Lacarbonara[1]
Assistant Professor,
Dipartimento di Ingegneria Strutturale e Geotecnica,
University of Rome La Sapienza,
via Eudossiana, 18
Rome 00184, Italy

C.-M. Chin
Senior Research Engineer,
VSAS Center General Motors Corporation,
MC 480-305-200
6440 E. 12 Mile Rd.,
Warren, MI 48090-9000

R. R. Soper
Senior Engineer,
Westvaco Covington Research Laboratory,
752 N. Mill Road,
Covington, VA 24426

An open-loop nonlinear control strategy applied to a hinged-hinged shallow arch, subjected to a longitudinal end-displacement with frequency twice the frequency of the second mode (principal parametric resonance), is developed. The control action—a transverse point force at the midspan—is typical of many single-input control systems; the control authority onto part of the system dynamics is high whereas the control authority onto some other part of the system dynamics is zero within the linear regime. However, although the action of the controller is orthogonal, in a linear sense, to the externally excited first antisymmetric mode, beneficial effects are exerted through nonlinear actuator action due to the system structural nonlinearities. The employed mechanism generating the effective nonlinear controller action is a one-half subharmonic resonance (control frequency being twice the frequency of the excited mode). The appropriate form of the control signal and associated phase is suggested by the dynamics at reduced orders, determined by a multiple-scales perturbation analysis directly applied to the integral-partial-differential equations of motion and boundary conditions. For optimal control phase and gain—the latter obtained via a combined analytical and numerical approach with minimization of a suitable cost functional—the parametric resonance is cancelled and the response of the system is reduced by orders of magnitude near resonance. The robustness of the proposed control methodology with respect to phase and frequency variations is also demonstrated. [DOI: 10.1115/1.1459069]

Introduction

Shallow arches are common structural members widely used either in civil engineering (e.g., bridges) or in mechanical and aerospace engineering as subelements of more complex structures. More recently, they are being investigated also as components of nanostructures. External resonant excitations may be sources of undesirable flexural vibrations which may be either catastrophic (due to coupling with torsional modes as, e.g., in the collapse of the Tacoma Narrow Bridge) or may significantly reduce the service life due to fatigue. An insidious dynamic instability in these systems is the parametric resonance which can be excited by longitudinal end-displacements or loads above a threshold level ([1]) and can cause violent and complex vibrations ([2]).

To cope with these large-amplitude vibrations, an open-loop control strategy is developed for a hinged-hinged shallow arch excited by a longitudinal end-displacement which is parametrically resonant with the first antisymmetric mode. The control input is a transverse force at the midspan (Fig. 1).

The task of mitigating the effects of resonant disturbances such as parametric excitations has been tackled in a number of different ways ranging from direct disturbance rejection via classical control theory techniques to the use of vibration absorbers attached to the main system as dedicated substructures. For example, a number of works have addressed both theoretically and experimentally the problem of controlling transverse oscillations in distributed-parameter systems by parametric-type control actions ([3,4]) or by coupling autoparametrically the system to an electronic circuit thereby exploiting the saturation phenomenon due to a 2:1 internal resonance mechanism ([5]). Yabuno et al. ([6]) showed that a parametric resonance in a cantilever beam can be suppressed by attaching a pendulum absorber to the beam tip. Using a more theoretical framework, Maschke et al. ([7]) are currently developing a port-controlled Hamiltonian formulation for the dynamics of nonlinear distributed-parameter systems to represent the energy flows through the boundaries of these systems aimed at extending some control schemes proposed for nonlinear finite-dimensional systems ([8]).

When excitations and actuations enter a system at the same point and in the same way, direct cancellation of the disturbance, resulting in no net energy transfer to the system, is possible. In systems such as the one under consideration, wherein the actuation and excitation are noncollocated, direct cancellation of the system disturbance is not possible. Furthermore, the linear modal projection of the control force onto the antisymmetric modes is zero entailing that the system is linearly uncontrollable with the given control input. However, due to the system structural nonlinearities, the nonlinear controller action may be exploited to cancel the external principal parametric resonance. Therefore, the same actuator, optimally collocated to control symmetric modes, may be still employed for controlling antisymmetric modes. One of the objectives of this paper is, in fact, to show how an intelligent exploitation of nonlinear phenomena can greatly expand on the capabilities to control a distributed-parameter system.

A direct perturbation expansion of the system dynamics facilitates understanding of the mechanism by which the full nonlinear actuator input may be used to suppress the resonant part of the excitation and further reduce the residual steady-state oscillations. The key mechanism here used is a subharmonic resonance of order one-half arising from the quadratic nonlinearities (due to initial curvature).

This work is not the first to examine the effects of the interaction of parametric resonances with subharmonic resonances of order one-half although for single-degree-of-freedom systems only ([9]). In a previous work by the same authors ([10]), similar concepts were employed to address noncollocated disturbances

[1]To whom all correspondence should be addressed.

Contributed by the Applied Mechanics Division of The American Society of Mechanical Engineers for publication in the ASME Journal of Applied Mechanics. Manuscript received by the ASME Applied Mechanics Division, Oct. 18, 2000; final revision, Oct. 25, 2001. Associate Editor: N. C. Perkins. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

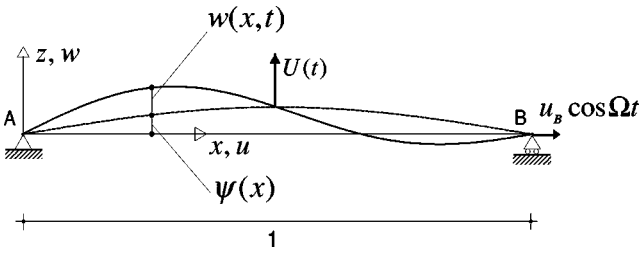


Fig. 1 Shallow arch geometry with the disturbance and the control input

via nonlinear actuator action in a pendulum-type crane architecture. It is worth pointing out that the use of a perturbation technique, namely the method of multiple scales, is here aimed at designing the type of control inputs rather than simply at obtaining closed-form approximate responses of the system to external disturbances.

The control strategy is referred to as ''open loop'' because neither the system states nor a measured output are employed in direct feedback. Nevertheless, the approach tacitly assumes direct availability of the disturbance levels and relative phases. In practice, these values might be directly measured or estimated through use of an augmented state observer although this may introduce significant complexity. Also, determination of the resonant-excitation character could be extracted from base-excitation measurements by means of phase-locked-loop electronics ([11]). However, the control methodology is shown to be robust with respect to limited phase and frequency variations.

Equations of Motion and Problem Formulation

Nonlinear vibrations of shallow elastic arches around the initial configuration $\hat{\psi}$ are governed, in dimensional form, by the following integral-partial-differential equation ([12]):

$$
\begin{aligned}
&m\frac{\partial^2 \hat{w}}{\partial \hat{t}^2}+EI\frac{\partial^4 \hat{w}}{\partial \hat{x}^4}-\frac{EA}{l}\frac{d^2\hat{\psi}}{d\hat{x}^2}\int_0^l \frac{\partial \hat{w}}{\partial \hat{x}}\frac{d\hat{\psi}}{d\hat{x}}d\hat{x}-\frac{EA}{l}\hat{u}(l,\hat{t})\frac{\partial^2 \hat{w}}{\partial \hat{x}^2}\\
&\quad-\frac{EA}{l}\frac{\partial^2 \hat{w}}{\partial \hat{x}^2}\int_0^l \frac{\partial \hat{w}}{\partial \hat{x}}\frac{d\hat{\psi}}{d\hat{x}}d\hat{x}-\frac{EA}{2l}\frac{d^2\hat{\psi}}{d\hat{x}^2}\int_0^l\left(\frac{\partial \hat{w}}{\partial \hat{x}}\right)^2 d\hat{x}\\
&\quad-\frac{EA}{2l}\frac{\partial^2 \hat{w}}{\partial \hat{x}^2}\int_0^l\left(\frac{\partial \hat{w}}{\partial \hat{x}}\right)^2 d\hat{x}\\
&\quad=-\hat{c}\frac{\partial \hat{w}}{\partial \hat{t}}+\hat{U}_c(\hat{t})\,\delta\left(\hat{x}-\frac{l}{2}\right)+\frac{EA}{l}\hat{u}(l,\hat{t})\frac{d^2\hat{\psi}}{d\hat{x}^2} \qquad (1)
\end{aligned}
$$

where m is the mass per unit length; l is the span of the arch; A and I denote the area and moment of inertia of the cross section, respectively; $\hat{c}$ is the coefficient of linear viscous damping; E is Young's modulus; $\hat{u}(l,\hat{t})$ and $\hat{U}_c(\hat{t})$ represent the prescribed end-displacement (external boundary disturbance) and the control force at the midspan, respectively; and δ denotes the Dirac delta function (see Fig. 1 for the definitions of $\hat{\psi}$, $\hat{x}$, and $\hat{w}$). Using the following nondimensional variables and parameters: $x=\hat{x}/l$, $\psi=\hat{\psi}/r$, $w=\hat{w}/r$, $t=\hat{t}\sqrt{EI/ml^4}$ (r is the radius of gyration of the cross section), Eq. (1) is transformed into its simplest nondimensional form as

$$
\begin{aligned}
&\frac{\partial^2 w}{\partial t^2}+\frac{\partial^4 w}{\partial x^4}-\frac{d^2\psi}{dx^2}\int_0^1 \frac{\partial w}{\partial x}\frac{d\psi}{dx}dx\\
&\quad-\frac{\partial^2 w}{\partial x^2}\int_0^1 \frac{\partial w}{\partial x}\frac{d\psi}{dx}dx-\frac{1}{2}\frac{d^2\psi}{dx^2}\int_0^1\left(\frac{\partial w}{\partial x}\right)^2 dx\\
&\quad-\frac{1}{2}\frac{\partial^2 w}{\partial x^2}\int_0^1\left(\frac{\partial w}{\partial x}\right)^2 dx\\
&\quad=-\epsilon^{\nu_2}c\frac{\partial w}{\partial t}+\epsilon^{\nu_3}U_c(t)\,\delta\left(x-\frac{1}{2}\right)\\
&\quad+\epsilon^{\nu_1}u(1,t)\left(\frac{\partial^2 w}{\partial x^2}+\frac{d^2\psi}{dx^2}\right) \qquad (2)
\end{aligned}
$$

where $\epsilon^{\nu_2}c=\hat{c}l^2/\sqrt{mEI}$, $\epsilon^{\nu_3}U_c=\hat{U}_c l^4/(rEI)$, and $\epsilon^{\nu_1}u(1,t)=\hat{u}(l,\hat{t})l/r^2$, with ϵ denoting a small nondimensional number used as a bookkeeping device.

For hinged-hinged arches, the boundary conditions are

$$
w=0 \quad \text{and} \quad \frac{\partial^2 w}{\partial x^2}=0 \quad \text{at} \quad x=0 \quad \text{and} \quad x=1. \qquad (3)
$$

The linear unforced undamped problem is obtained from Eq. (2) by dropping the damping term, the disturbance, the control force, and the nonlinear terms; that is,

$$
\frac{\partial^2 w}{\partial t^2}+\mathcal{L}w=\frac{\partial^2 w}{\partial t^2}+\frac{\partial^4 w}{\partial x^4}-\frac{d^2\psi}{dx^2}\int_0^1 \frac{\partial w}{\partial x}\frac{d\psi}{dx}dx=0 \qquad (4)
$$

with boundary conditions (3). In Eq. (4), $\mathcal{L}$ denotes the linear stiffness operator.

Because the linear unforced undamped problem is self-adjoint, the eigenfunctions $\phi_m(x)$ are mutually orthogonal and they are normalized as follows: $\int_0^1 \phi_m\phi_n dx=\langle \phi_m\phi_n\rangle=\delta_{mn}$, $\langle \phi_m\mathcal{L}\phi_n\rangle=\omega_n^2\delta_{mn}$ where δ_{mn} is the Kronecker delta. For a simply-supported shallow arch with initial shape $\psi(x)=b\sin\pi x$, the eigenmodes are readily obtained in the form of the trigonometric series

$$
\phi_n(x)=\sqrt{2}\,\sin n\pi x, \quad n=1,2,\ldots \qquad (5)
$$

and the associated natural frequencies are given by $\omega_1=\pi^2\sqrt{1+b^2/2}$ and $\omega_n=n^2\pi^2$, $n=2,3,\ldots$. A few bimodal two-to-one and one-to-one internal resonances may be possible in hinged-hinged shallow arches.

In the next section, we develop an open-loop control strategy to reduce the nonlinear resonant vibrations arising from the boundary-excited principal parametric resonance of the second mode when this mode is away from the mentioned internal resonances.

Open-Loop Control Of The Principal Parametric Resonance Of The Second Mode

In this section, we construct the response of the system to a principal parametric resonance of the second mode when no internal resonances engage this mode with any other mode and the system is subjected to a control force introduced to suppress the principal parametric resonance and further minimize the overall steady-state vibrations. The boundary disturbance is sinusoidal; namely, $u(1,t)=u_B\cos\Omega t$ with $\Omega\approx 2\omega_2$.

The ordering of the excitation and damping demands that $\nu_1=\nu_3=1$ and $\nu_2=2$. This ordering promotes the excitation to first order to activate the principal parametric resonance. Further, the control input is introduced also at first order so as to balance and possibly inhibit the external principal parametric resonance. On the other hand, the damping force is demoted to third order where

there appear the secondary-resonant terms produced by the control force and the resonant effects of the structural nonlinearities yielding the frequency correction.

We assume the control signal as a pure tone; that is, $U_c(t) = U_c \exp(i(\Omega_c t+\psi_c))+cc$ where cc denotes the complex conjugate of the preceding term. The objective is to design a control term at first order that is capable of producing resonant terms at second order counteracting the effects of the principal parametric resonance which are proportional to $\bar{A}\exp(i\omega_2 t)$. Here, A indicates the complex-valued amplitude of the arch response at the natural frequency of the second mode (ω_2) and the overbar denotes its complex conjugate. Hence, to create nonlinear controller terms at second-order proportional to $\bar{A}\exp(i\omega_2 t)$, the control frequency needs to satisfy the relation $\pm\Omega_c \pm \omega_2 = \omega_2$. Consequently, we choose $\Omega_c \approx 2\omega_2$. In addition, it is required that the resulting nonlinear controller action have nonzero projection onto the mode to be controlled (i.e., the first antisymmetric mode). The designed control input is thus expected to produce resonant terms via a one-half subharmonic resonance.

The method of multiple scales is employed to determine a third–order uniform expansion of the solutions of (2) and (3). As mentioned, the resonant dynamics arising from the disturbance and control input will appear at second order. However, to capture the nonlinear frequency correction, we seek a third-order expansion. Consequently, one needs to use the method of reconstitution ([13]). To obtain consistent reconstituted modulation equations, the equations of motion need to be cast as a system of first-order equations in time (i.e., state-space formulation), first, and, then, they can be treated with the method of multiple scales. Therefore, we first rewrite Eq. (2) as a system of two first-order equations in time by adding the equation $\dot{w}=v$ and putting $\dot{v}$ instead of $\ddot{w}$ in the equation of motion. Moreover, we directly attack the equations of motion and boundary conditions instead of treating finite-degree-of-freedom discretized versions. In fact, it has been extensively shown that treatment of a discretized set of distributed-parameter systems with quadratic and cubic nonlinearities may lead to erroneous quantitative and, in some cases, qualitative results ([14,15]). Thus, we overcome the problem of order reduction and the associated problems as spillover effects which are critical when designing control laws for such systems.

We seek a third-order uniform expansion in the form

$$\begin{aligned} w(x,t) &= \sum_{k=1}^{3} \epsilon^k w_k(x,T_0,T_1,T_2)+ \\ v(x,t) &= \sum_{k=1}^{3} \epsilon^k v_k(x,T_0,T_1,T_2)+ \end{aligned} \tag{6}$$

where $T_k=\epsilon^k t$ are the time scales. Then, the first derivative with respect to time is defined as $\partial/\partial t = D_0+\epsilon D_1+\epsilon^2 D_2+$ where $D_n=\partial/\partial T_n$. To express the nearness of the principal parametric resonance, we introduce the detuning parameter σ such that $\Omega = 2\omega_2+\epsilon\sigma$. We assume at this stage that the excitation and control signals are phase-locked and one-to-one; that is, $\Omega_c=\Omega$ (this assumption will be later relaxed).

Substituting (6) into the system of first-order (in time) equations of motion and boundary conditions (3), using the independence of the time scales, and equating coefficients of like powers of ϵ yields

Order ϵ:

$$D_0 w_1 - v_1 = 0 \tag{7}$$

$$D_0 v_1 + \mathcal{L} w_1 = \frac{1}{2} U_c e^{i(\Omega T_0+\psi_c)} \delta\left(x-\frac{1}{2}\right) + \frac{1}{2}\frac{d^2\psi}{dx^2} u_B e^{i\Omega T_0} + cc \tag{8}$$

Order ϵ^2:

$$D_0 w_2 - v_2 = -D_1 w_1 \tag{9}$$

$$D_0 v_2 + \mathcal{L} w_2 = -D_1 v_1 + \frac{\partial^2 w_1}{\partial x^2}\int_0^1 \frac{\partial w_1}{\partial x}\frac{d\psi}{dx}dx + \frac{1}{2}\frac{d^2\psi}{dx^2}\int_0^1 \left(\frac{\partial w_1}{\partial x}\right)^2 dx + \frac{1}{2}\frac{\partial^2 w_1}{\partial x^2}(u_B e^{i\Omega T_0}+cc) \tag{10}$$

Order ϵ^3:

$$D_0 w_3 - v_3 = -D_2 w_1 - D_1 w_2 \tag{11}$$

$$\begin{aligned} D_0 v_3 + \mathcal{L} w_3 = &-D_2 v_1 - D_1 v_2 + \frac{\partial^2 w_1}{\partial x^2}\int_0^1 \frac{\partial w_2}{\partial x}\frac{d\psi}{dx}dx \\ &+ \frac{\partial^2 w_2}{\partial x^2}\int_0^1 \frac{\partial w_1}{\partial x}\frac{d\psi}{dx}dx + \frac{d^2\psi}{dx^2}\int_0^1 \frac{\partial w_1}{\partial x}\frac{\partial w_2}{\partial x}dx \\ &+ \frac{1}{2}\frac{\partial^2 w_1}{\partial x^2}\int_0^1 \left(\frac{\partial w_1}{\partial x}\right)^2 dx - 2\mu v_1 \\ &+ \frac{1}{2}\frac{\partial^2 w_2}{\partial x^2}(u_B e^{i\Omega T_0}+cc) \end{aligned} \tag{12}$$

The boundary conditions at all orders are given by (3).

Because the second mode is directly excited by the principal parametric resonance of the disturbance and indirectly by the subharmonic resonance of the control input; moreover, because there are no internal resonances involving this mode, we assume the solution at order ϵ as

$$\begin{aligned} w_1 = &A(T_1,T_2)e^{i\omega_2 T_0}\phi_2(x)+\mathcal{U}_c(x)e^{i(\Omega T_0+\psi_c)} \\ &+\mathcal{B}(x)e^{i\Omega T_0}+cc \end{aligned} \tag{13}$$

where the functions $\mathcal{U}_c(x)$ and $\mathcal{B}(x)$ are solutions of the following boundary value problems:

$$\mathcal{L}\mathcal{U}_c - 4\omega_2^2\mathcal{U}_c = \frac{1}{2}U_c\delta\left(x-\frac{1}{2}\right) \tag{14}$$

$$\mathcal{L}\mathcal{B} - 4\omega_2^2\mathcal{B} = \frac{1}{2}u_B\frac{d^2\psi}{dx^2}. \tag{15}$$

The function $\mathcal{U}_c$ can be expressed as an infinite series of the eigenfunctions in the form

$$\begin{aligned} \mathcal{U}_c(x) &= \frac{1}{2}U_c\sum_{k=0}^{\infty}\frac{\phi_{2k+1}(1/2)}{\omega_{2k+1}^2-4\omega_2^2}\phi_{2k+1}(x) \\ &= \frac{U_c}{\pi^4}\left[\frac{2}{b^2-126}\sin\ \pi x+\sum_{k=1}^{\infty}\frac{\sin((2k+1)(\pi/2))}{(2k+1)^4-64}\right. \\ &\quad \left.\times\sin(2k+1)\pi x\right]. \end{aligned} \tag{16}$$

On the other hand, the function $\mathcal{B}$ can be readily obtained as

$$\mathcal{B}(x) = -\frac{u_B}{\pi^2(b^2-126)}b\ \sin\ \pi x. \tag{17}$$

Substituting (13) and $v_1=D_0 w_1$ into the second-order problem, Eqs. (9) and (10), we obtain

$$D_0 w_2 - v_2 = -(D_1 A)e^{i\omega_2 T_0}\phi_2(x)+cc \tag{18}$$

$$D_0 v_2 + \mathcal{L} w_2 = -i\omega_2 (D_1 A) e^{i\omega_2 T_0} \phi_2(x) + \overline{A} e^{i(\omega_2 T_0 + \sigma T_1)} \phi_2'' \left(\langle \mathcal{B}' \psi' \rangle + \frac{1}{2} u_B + e^{i\psi_c} \langle \mathcal{U}_c' \psi' \rangle \right)$$
$$+ A e^{i(3\omega_2 T_0 + \sigma T_1)} \phi_2'' \left(\langle \mathcal{B}' \psi' \rangle + \frac{1}{2} u_B + e^{i\psi_c} \langle \mathcal{U}_c' \psi' \rangle \right) + e^{i(4\omega_2 T_0 + 2\sigma T_1 + 2\psi_c)} \left(\frac{1}{2} \psi'' \langle \mathcal{U}_c' \mathcal{U}_c' \rangle + \mathcal{U}_c'' \langle \mathcal{U}_c' \psi' \rangle \right)$$
$$+ e^{i(4\omega_2 T_0 + 2\sigma T_1 + \psi_c)} \left(\mathcal{U}_c'' \langle \mathcal{B}' \psi' \rangle + \mathcal{B}'' \langle \mathcal{U}_c' \psi' \rangle + \psi'' \langle \mathcal{U}_c' \mathcal{B}' \rangle + \frac{1}{2} u_B \mathcal{U}_c'' \right)$$
$$+ e^{i(4\omega_2 T_0 + 2\sigma T_1)} \left(\frac{1}{2} \psi'' \langle \mathcal{B}' \mathcal{B}' \rangle + \mathcal{B}'' \langle \mathcal{B}' \psi' \rangle + \frac{1}{2} u_B \mathcal{B}'' \right)$$
$$+ \frac{1}{2} A^2 e^{2i\omega_2 T_0} \psi'' \langle \phi_2' \phi_2' \rangle + \frac{1}{2} A \overline{A} \psi'' \langle \phi_2' \phi_2' \rangle + e^{i\psi_c} (\mathcal{U}_c'' \langle \mathcal{B}' \psi' \rangle + \mathcal{B}'' \langle \mathcal{U}_c' \psi' \rangle + \psi'' \langle \mathcal{U}_c' \mathcal{B}' \rangle) + \mathcal{U}_c'' \langle \mathcal{U}_c' \psi' \rangle$$
$$+ \frac{1}{2} \psi'' \langle \mathcal{U}_c' \mathcal{U}_c' \rangle + \mathcal{B}'' \langle \mathcal{B}' \psi' \rangle + \frac{1}{2} \psi'' \langle \mathcal{B}' \mathcal{B}' \rangle + \frac{1}{2} u_B \mathcal{B}'' + cc \tag{19}$$

where the prime indicates differentiation with respect to x. Because the associated homogeneous problem admits nontrivial solutions, the resulting inhomogeneous problem, Eqs. (18), (19), and (3), possesses solutions only if solvability conditions are satisfied. The solvability conditions of Eqs. (18), (19), and (3) demand that the right-hand side of Eqs. (18) and (19) be orthogonal to every solution of the associated adjoint homogeneous problem. The transposes of the solutions of the adjoint homogeneous problem are $(i\omega_k, 1)\phi_k(x)\exp(-i\omega_k T_0)$. Hence, imposing that the right-hand side of Eqs. (18) and (19) be orthogonal to these adjoints, we obtain the following solvability condition:

$$2i\omega_2 D_1 A = \overline{A} e^{i\sigma T_1} (K_1 + e^{i\psi_c} K_2) \tag{20}$$

where the *external parametric excitation coefficient* and the *subharmonic control gain* are given, respectively, by

$$K_1 = \langle \phi_2'' \phi_2 \rangle \left(\langle \mathcal{B}' \psi' \rangle + \frac{1}{2} u_B \right) = \frac{252\pi^2}{b^2 - 126} u_B \tag{21}$$

and

$$K_2 = \langle \phi_2'' \phi_2 \rangle \langle \mathcal{U}_c' \psi' \rangle = -\frac{4b}{b^2 - 126} U_c . \tag{22}$$

Substituting (20) into (18) and (19), we seek the solutions of the resulting equations along with (3) in the form

$$w_2 = [\chi_1(x) + e^{i\psi_c} \chi_2(x)] \overline{A} e^{i(\omega_2 T_0 + \sigma T_1)} + \chi_3(x) A^2 e^{2i\omega_2 T_0}$$
$$+ \chi_4(x) A \overline{A} + [\chi_5(x) + e^{i\psi_c} \chi_6(x)] A e^{i(3\omega_2 T_0 + \sigma T_1)}$$
$$+ [\chi_7(x) + e^{i\psi_c} \chi_8(x) + e^{2i\psi_c} \chi_9(x)] e^{i(4\omega_2 T_0 + 2\sigma T_1)} + \chi_{10}(x)$$
$$+ e^{i\psi_c} \chi_{11}(x) + cc \tag{23}$$

$$v_2 = [\eta_1(x) + e^{i\psi_c} \eta_2(x)] \overline{A} e^{i(\omega_2 T_0 + \sigma T_1)} + \eta_3(x) A^2 e^{2i\omega_2 T_0}$$
$$+ \eta_4(x) A \overline{A} + [\eta_5(x) + e^{i\psi_c} \eta_6(x)] A e^{i(3\omega_2 T_0 + \sigma T_1)}$$
$$+ [\eta_7(x) + e^{i\psi_c} \eta_8(x) + e^{2i\psi_c} \eta_9(x)] e^{i(4\omega_2 T_0 + 2\sigma T_1)} + \eta_{10}(x)$$
$$+ e^{i\psi_c} \eta_{11}(x) + cc. \tag{24}$$

The second-order displacement and velocity fields depend on the shape functions χ_j and η_j. The functions χ_j are solutions of a number of boundary value problems, Eqs. (40)–(49) with boundary conditions (50), reported in the Appendix. The functions governing the system response at first and second-order are shown in Fig. 2 for the values $b = 16.5$, $u_B = 1$, and $U_c = 28.936$. Using the solutions at second order, Eqs. (23) and (24), and substituting them into the third-order problem, Eqs. (11) and (12), we impose the solvability condition governing at this order the dependence of A on the scale T_2; that is,

$$2i\omega_2 (D_2 A + \mu A) = \alpha A^2 \overline{A} + \sigma_s A \tag{25}$$

where the *effective nonlinearity coefficient* is given by

$$\alpha = \langle \phi_2 \phi_2'' \rangle \left[2\langle \psi' \chi_4' \rangle + \langle \psi' \chi_3' \rangle + \frac{3}{2} \langle \phi_2' \phi_2' \rangle \right]$$
$$= 8\pi^4 \left[b^2 \left(\frac{2}{b^2 + 2} + \frac{1}{b^2 - 126} \right) - 3 \right]. \tag{26}$$

On the other hand, the coefficient of the linear frequency shift is given by

$$\sigma_s = \frac{16 U_c^2}{(b^2 + 2)} \left[\frac{b^2}{\pi^4 (b^2 - 126)^2} - \frac{\pi^2}{2} S_1^{(\infty)} \right]$$
$$+ \frac{500\pi^2 b^2 u_B^2 - 16b(b^2 - 1) U_c u_B}{\pi^2 (b^2 + 2)(b^2 - 126)^2} + \frac{(63\pi^2 u_B - U_c b)^2}{4\pi^4 (b^2 - 126)^2} \tag{27}$$

where use of the phase condition $\psi_c = 2n\pi$ was made (this condition will be discussed in the next section).

Employing the method of reconstitution by substituting (20) and (25) into $\dot{A} = \epsilon D_1 A + \epsilon^2 D_2 A + \ldots$, and setting $\epsilon = 1$, we obtain the normal form governing the modulation of the amplitude and phase as

$$2i\omega_2 (\dot{A} + \mu A) = (K_1 + e^{i\psi_c} K_2) e^{i\sigma t} \overline{A} + \sigma_s A + \alpha A^2 \overline{A}. \tag{28}$$

The deflection of the arch, to second order, is given by the sum of (13) and (23) with the complex-valued amplitude A being governed by the reconstituted modulation Eq. (28).

Control Input Design. Our objective is the reduction of the overall parametrically excited vibrations. Therefore, to optimize the nonlinear control action, we choose the following cost functional:

$$J = \max_{0 \leq t \leq T_c} \left\{ \int_0^1 w(x,t;a_m)^2 dx \right\} \tag{29}$$

where T_c is the period of oscillation and a_m is the maximum stable value of the real part of the complex-valued amplitude A at the disturbance frequency detuning σ. This cost is calculated based on steady-state rather than transient behavior. Hence, it reflects the cost predicted over long periods of operation where transient effects are negligible. The constraint for the optimization problem is given by the reconstituted modulation Eq. (28). The domain where minimization of J is sought is spanned by U_c and ψ_c. It is clear that if we seek conditions for minimizing a_m, then

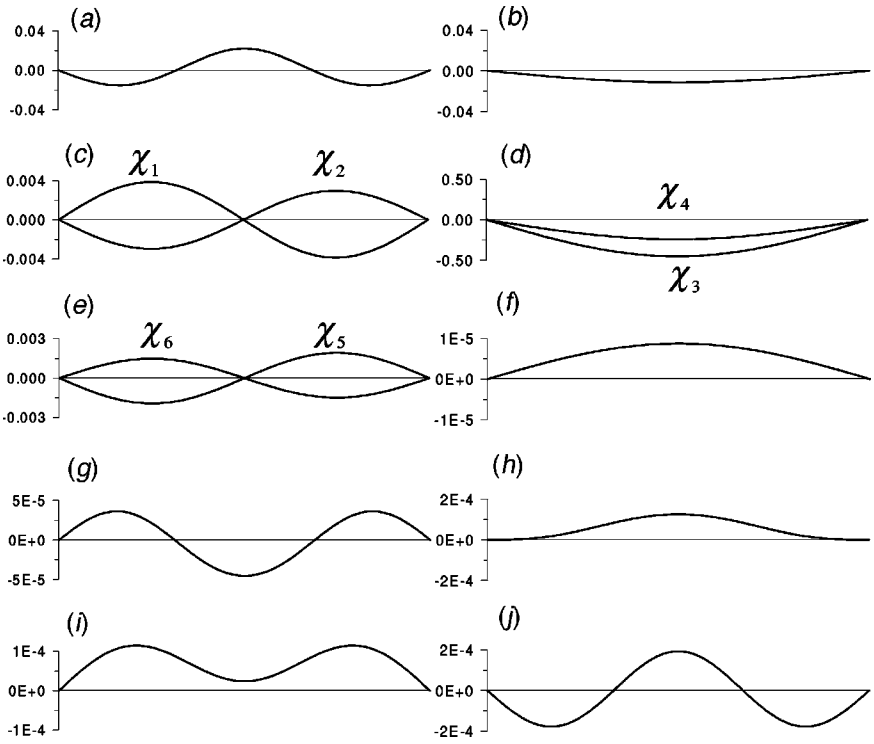


Fig. 2 First and second-order shape functions: (*a*) $\mathcal{U}_c$; (*b*) $\mathcal{B}$; (*c*) χ_1 and χ_2; (*d*) χ_3 and χ_4; (*e*) χ_5 and χ_6; (*f*) χ_7; (*g*) χ_8; (*h*) χ_9; (*i*) χ_{10}; and (*j*) χ_{11} when $b=16.5$, $u_B=1.0$, and $U_c=28.936$

we can more efficiently minimize J within the design parameter subset where the minimum of a_m is attained. The form of the effective parametric excitation coefficient (sum of the external parametric and subharmonic control coefficients) suggests that we have the ability to affect this coefficient through the control gain and phase. Then, the objective is to lower this coefficient below the activation threshold of the parametric resonance. In this way, we prevent the principal parametric resonance from being activated and the only stable solution becomes the trivial solution.

It is clear that the magnitude of the overall effective parametric resonance coefficient—$K_1+K_2\exp(i\psi_c)$—is minimum with respect to changes in the control phase when $K_2\exp(i\psi_c)$ lies in a complex-vector direction opposite to K_1. Inspecting Eqs. (21) and (22), we note that K_1 and K_2 have always opposite sign; hence, the control phase condition is easily obtained as $\psi_c=2n\pi$, $n=0,1,\ldots$. Using this condition, we can transform the complex-valued modulation Eq. (28) into the real-valued modulation equations for the amplitude and phase. To this end, we assume the following polar transformation $A=1/2a\exp(i\beta)\exp(-i\sigma_s/2t)$ and obtain the phase $\gamma=(\sigma+\sigma_s)t-2\beta$ in order to render the modulation equations autonomous. Substituting the polar transformation into (28) and using the control phase condition (i.e., $\psi_c=2n\pi$) yields

$$\dot{a}=-\mu a+\frac{1}{2}\frac{K_e}{\omega_2}a\sin\gamma \tag{30}$$

$$a\dot{\gamma}=a\left(\sigma+\frac{\sigma_s}{\omega_2}\right)+\frac{\alpha}{4\omega_2}a^3+\frac{K_e}{\omega_2}a\cos\gamma, \tag{31}$$

where $K_e=K_1+K_2$.

The frequency-response equation for the steady-state amplitude is given by

$$\sigma=-\left(\frac{\sigma_s}{\omega_2}+\frac{\alpha}{4\omega_2}a^2\right)\pm\sqrt{\frac{K_e^2}{\omega_2^2}-4\mu^2} \tag{32}$$

whereas the phase is given by $\tan\gamma=\mp\mu(K_e^2/4\omega_2^2-\mu^2)^{-(1/2)}$.

We gain more insight into the effects of the control scheme by preliminarily investigating the system uncontrolled dynamics excited by the principal parametric resonance. In this case, the frequency-response equation can be obtained from (32) by putting $U_c=0$ into σ_s and K_e (hence, $K_e=K_1$). The amplitudes of the nontrivial uncontrolled responses are given by

$$a=2\left[\frac{\alpha}{\omega_2}\left(-\sigma-\frac{\sigma_{s,uc}}{\omega_2}\pm\sqrt{\frac{K_1^2}{\omega_2^2}-4\mu^2}\right)\right]^{1/2}. \tag{33}$$

Seeking conditions for the existence of real solutions for a, we conclude that there are three regions in the plane of disturbance amplitude and detuning as shown in Fig. 3 (where $b=16.5$ and $\mu=0.05$). Namely, (i) in region I ($u_B<u_{B,cr}$ $\forall$ σ or $\sigma>-\sigma_{s,uc}/\omega_2+(K_1^2/\omega_2^2-4\mu^2)^{1/2}$), there are no solutions; (ii) in region II $(-\sigma_{s,uc}/\omega_2-(K_1^2/\omega_2^2-4\mu^2)^{1/2}<\sigma<-\sigma_{s,uc}/\omega_2+(K_1^2/\omega_2^2-4\mu^2)^{1/2})$, there is only one real solution; (iii) in region III ($\sigma<-\sigma_{s,uc}/\omega_2-(K_1^2/\omega_2^2-4\mu^2)^{1/2}$), there are two real solutions. We note that subregion I′ (i.e., $u_B<u_{B,cr}$), embedded in region I, is such that the nonactivation of the parametric resonance therein is independent of the frequency detuning. The activation threshold is obtained from (33) imposing vanishing of the argument of the inner square root which yields $u_{B,cr}=2\mu|b^2-126|/63$.

Computation of the effective nonlinearity coefficient, when $b=16.5$, leads to $\alpha=660.008$. The linear frequency shift is $\sigma_{s,uc}=6.9599\times10^{-2}$. In Fig. 4, we show the frequency-response curve of the uncontrolled case when $u_B=1$ and $\mu=0.05$. Clearly, for the considered sag level, the second mode of the arch is of the softening type. Point S_1 (S_2) corresponds to a supercritical (subcritical) pitchfork bifurcation.

In the presence of control input, the activation threshold for the parametric resonance is obtained from (32) letting $K_e^2=4\mu^2\omega_2^2$.

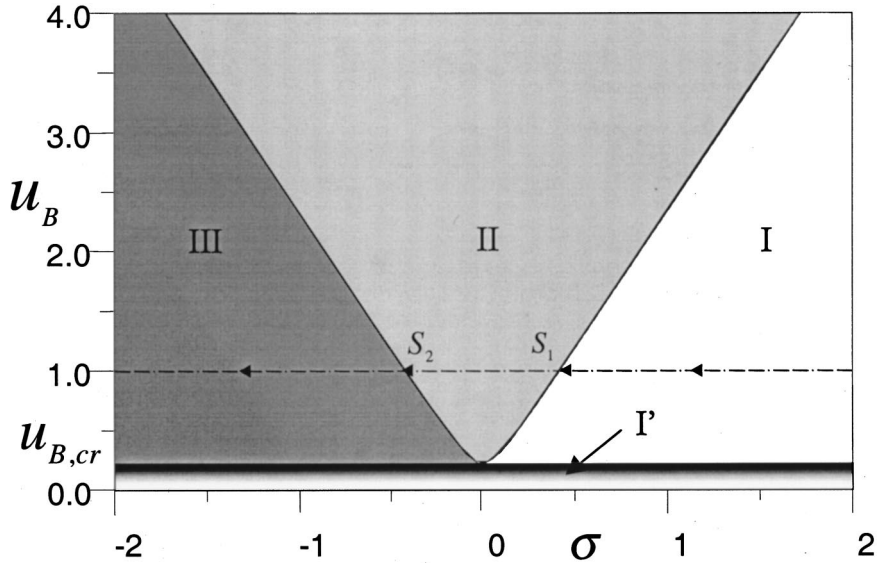


Fig. 3 Regions of activation/nonactivation of the principal parametric resonance in the plane of the disturbance frequency detuning and gain when $b=16.5$ and $\mu=0.05$

Substituting the expressions (21) and (22) for K_1 and K_2 into K_e for fixed disturbance level and damping, we determine the range of control gain where the parametric resonance is not activated as $U_c^{(1)}<U_c<U_c^{(2)}$ where $U_c^{(1,2)}=63\pi^2/b(u_B\mp u_{B,cr})$. Within this range (lightly shaded region in Fig. 5), there is a locus of control gains where the effective parametric coefficient becomes zero; that is, $U_c^{(3)}=63\pi^2 u_B/b$. It is worth emphasizing that requiring the control gain to be within the cancellation region in Fig. 5 entails enforcing the system to be in subregion I′ (Fig. 3) where the parametric resonance is annihilated for any frequency detuning.

It is interesting to investigate the sensitivity of the developed control strategy to variations in the relative phase and frequency detuning. We first investigate the effect of the phase variations with perfect frequency tuning (i.e., $\Omega=\Omega_c$). To this end, the analysis is simplified by neglecting third-order contributions which are not expected to influence significantly the resonance activation threshold. The region of resonance cancellation is obtained as

$$K_1^2+K_2^2+2K_1K_2\cos\psi_c\leqslant 4\mu^2\omega_2^2. \tag{34}$$

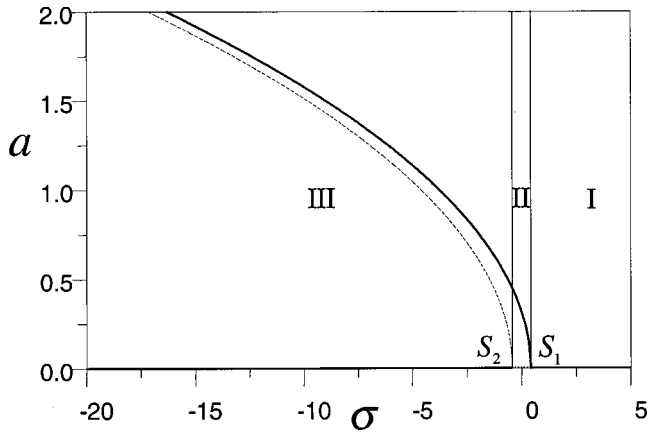


Fig. 4 Frequency-response curve of the uncontrolled arch when $b=16.5$, $\mu=0.05$, and $u_B=1$. Solid (dashed) line indicates stable (unstable) solutions.

Substituting the expressions for K_1 and K_2 into (34), we obtain the cancellation regions in the plane of disturbance and control gains for different phase angles ψ_c. Evidently, on account of (34), the regions are independent of the sign of ψ_c. Further, it can be shown that the cancellation regions are physically meaningful when $\psi_c\in(-\pi/2,\pi/2)$. Figure 5 shows clearly that $\psi_c=2n\pi$ is the optimal phase angle because the associated region is nonterminating indicating that, in principle, cancellation is achievable for any disturbance level. A variation of ψ_c by 6 deg renders the region terminating thereby entailing that past a disturbance threshold, cancellation cannot be achieved. However, this occurs for rather high values of disturbances. On the other hand, for low-to-medium excitation amplitudes, small changes in phase angle do not practically vary the cancellation region. Nevertheless, Fig. 5 also shows that significant increases in phase angle variations, as expected, degrade the control performance until it becomes ineffective when $\psi_c=\pi/2$.

Results And Discussion

Within the outlined regions of resonance cancellation (Fig. 5), the optimal gains are determined via the presented optimization scheme. Therefore, for the given nondimensional sag $b=16.5$, damping coefficient $\mu=0.05$, and disturbance level $u_B=1$, we computed the cost when U_c varies in the range $U_c^{(1)}=28.936$ to $U_c^{(2)}=46.432$ where the parametric resonance is cancelled and determine, accordingly, the optimal control gain as the value where the cost is minimum. Clearly, in this case, the only stable solution for the amplitude is the trivial solution. Accordingly, the solution for the displacement field at steady state can be expressed as

$$w(x,t)=2\ \cos\ \Omega t[\mathcal{B}(x)+\mathcal{U}_c(x)]+2\{\cos(2\Omega t)[\chi_7(x)+\chi_8(x)+\chi_9(x)]+\chi_{10}(x)+\chi_{11}(x)\}. \tag{35}$$

Thus, using (35), we computed $w(x,t)$ when U_c is equal to $U_c^{(1)}$, $U_c^{(2)}$, and $U_c^{(3)}$. Here, we note that the convergence of the first and second-order shape functions expressed as infinite series is very fast. In fact, numerical tests conducted with MATHEMATICA, showed that, for the given sag level, three/four terms are sufficient for convergence. Using five terms, we ob-

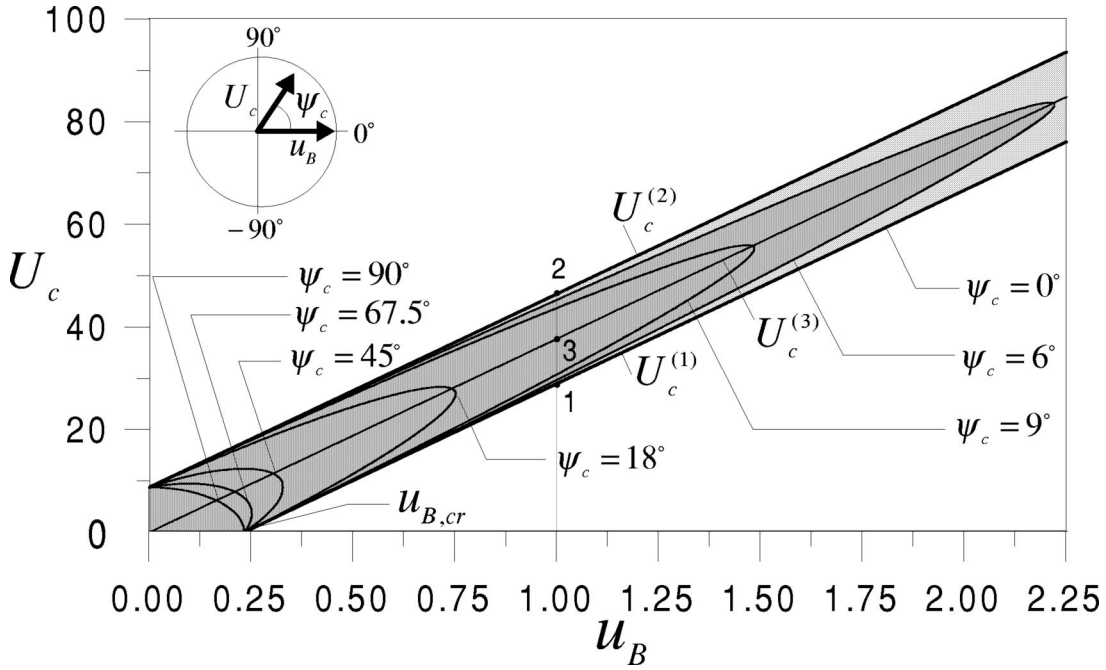


Fig. 5 Regions (shaded) of nonactivation of the principal parametric resonance in the plane of the disturbance and control gains for different control phase angles when $b=16.5$ and $\mu=0.05$

tained $S_1^{(5)}=1.6341\times10^{-5}$ and (i) $K_1=17.006$, $K_2=-13.058$, and $\sigma_s=-7.6441\times10^{-3}$ when $U_c=28.936$; (ii) $K_1=-K_2=17.006$, and $\sigma_s=-1.926\times10^{-2}$ when $U_c=63\pi^2u_B/b=37.684$; (iii) $K_1=17.006$, $K_2=-20.954$, and $\sigma_s=-2.543\times10^{-2}$ when $U_c=46.432$. Computation of the cost according to Eq. (29) shows that it is minimum when $U_c=U_c^{(1)}=28.936$. Hence, we assume this control gain as the optimal control gain.

It is interesting to compare the optimally controlled case with the uncontrolled case when the frequency detuning is $\sigma=-10$ and $u_B=1$, $b=16.5$, and $\mu=0.05$. First, we compare the steady-state responses. Then, we compare the transient responses in the presence of initial conditions within the basin of attraction of the uncontrolled parametrically resonant response.

The steady-state deflection for the controlled case is given by (35). On the other hand, the steady-state uncontrolled displacement field, to second order, is expressed as

$$w(x,t)=a\cos\left[\frac{1}{2}(\Omega t-\gamma)\right]\phi_2(x)+2\cos\Omega t\mathcal{B}(x)+a\cos\left[\frac{1}{2}(\Omega t+\gamma)\right]\chi_1(x)+2\cos(2\Omega t)\chi_7(x)+\frac{1}{2}a^2[\cos(\Omega t-\gamma)\chi_3(x)+\chi_4(x)]+a\chi_5(x)\cos\left[\frac{1}{2}(3\Omega t-\gamma)\right]+2\chi_{10}(x) \quad (36)$$

where χ_{10} is given by (61) inserting $U_c=0$ in it.

We computed the stable steady-state response amplitude and phase when $\sigma=-10$ and $u_B=1$ and found $a=1.579$ and $\gamma=-13.42$ deg. The period of oscillation is $T=0.182$ and the cost is $J=2.890$. In Fig. 6, we show the uncontrolled steady-state deflections contrasted with the optimally controlled deflections (magnified ten times in the scale of the uncontrolled deflections)

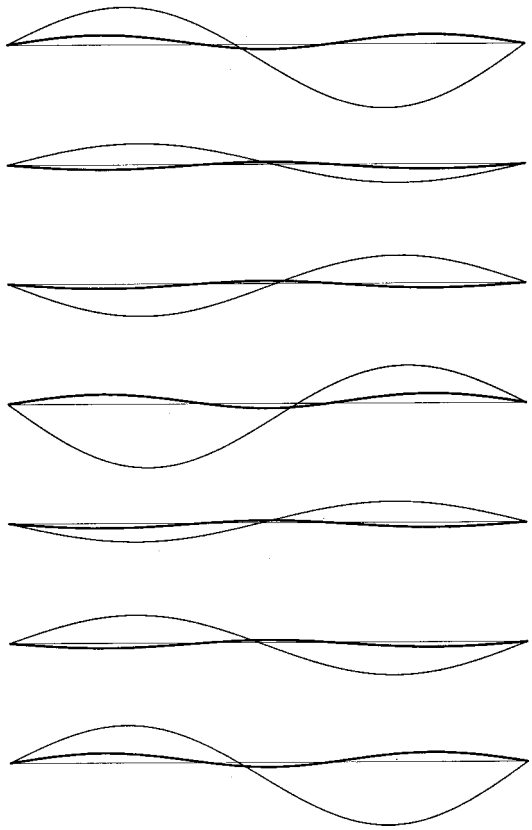


Fig. 6 Uncontrolled (thin line) and optimally controlled (thick line) dynamic deflections at seven discrete times equally spaced within a period of oscillation when $b=16.5$, $\mu=0.05$, $u_B=1$, $\sigma=-10$, and $U_c=28.936$

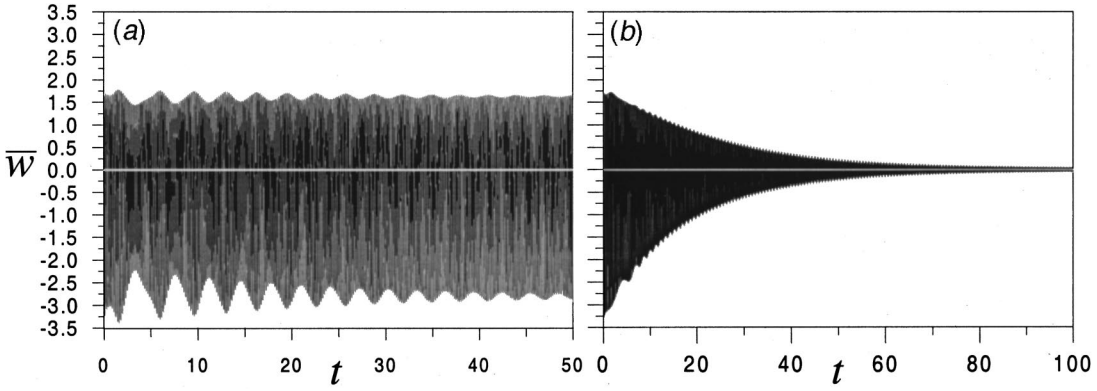


Fig. 7 Time histories of the uncontrolled and optimally controlled deflections ($\psi_c=0$ and $U_c=28.936$) at $x=1/4$ when $u_B=1$, $\sigma=-10$, $b=16.5$, $\mu=0.05$, $p(0)=1.5$, and $q(0)=-0.95$

at seven discrete times equally spaced in a period of the uncontrolled oscillations (twice the period of the controlled oscillations). The result is that the overall optimally controlled deflection at steady state is two orders of magnitude smaller than that of the uncontrolled case and the relative decrease of the cost is 97.5%.

Finally, to contrast the behavior of the uncontrolled response with the optimally controlled response in the transient regime, we cast the modulation equations and the displacement field in Cartesian form by using the transformation $A=1/2/(p-iq)\exp(i\sigma/2t)$. Using this transformation, the modulation equations become

$$\dot{p}=-\mu p+\frac{1}{2}\left(\frac{K_e}{\omega_2}-\sigma-\frac{\sigma_s}{\omega_2}\right)q-\frac{1}{8}\frac{\alpha}{\omega_2}(p^2+q^2)q \qquad (37)$$

$$\dot{q}=-\mu q+\frac{1}{2}\left(\frac{K_e}{\omega_2}+\sigma+\frac{\sigma_s}{\omega_2}\right)p+\frac{1}{8}\frac{\alpha}{\omega_2}(p^2+q^2)p. \qquad (38)$$

The displacement field, to second order, is also expressed in terms of Cartesian components as

$$\begin{aligned} w(x,t)=&\left(p\cos\frac{\Omega}{2}t+q\sin\frac{\Omega}{2}t\right)\phi_2(x)+2\cos\Omega t(\mathcal{B}(x)+\mathcal{U}_c(x))\\ &+\left(p\cos\frac{\Omega}{2}t-q\sin\frac{\Omega}{2}t\right)(\chi_1(x)+\chi_2(x))\\ &+\left(p\cos\frac{3\Omega}{2}t+q\sin\frac{3\Omega}{2}t\right)(\chi_5(x)+\chi_6(x))\\ &+2\cos(2\Omega t)(\chi_7(x)+\chi_8(x)+\chi_9(x))\\ &+\frac{1}{2}[(p^2-q^2)\cos\Omega t+2pq\sin\Omega t]\chi_3(x)\\ &+\frac{1}{2}(p^2+q^2)\chi_4(x)+2(\chi_{10}(x)+\chi_{11}(x)). \end{aligned} \qquad (39)$$

We integrated Eqs. (37) and (38) using a variable stepsize fifth-order Runge-Kutta scheme when $p(0)=1.5$ and $q(0)=-0.95$ for the optimally controlled and uncontrolled cases. Using Eq. (39), we computed the time histories of the deflections at $x=1/4$ in whose neighborhood the maximum vibration amplitude is attained and show the results in Fig. 7. We note that, after the transient dies out, the response in the controlled case is two orders of magnitude smaller than that in the uncontrolled case. Evidently, as expected, the decay rate of the transients in the controlled case is not affected by the control strategy.

Last, to show the robustness of the control strategy with respect to variations in frequency detuning, we performed some numeri cal tests. To this end, we introduced the control frequency detuning as $\Omega_c=2\omega_2+\epsilon\sigma_c$ and carried out the perturbation expansion up to second order. We integrated the obtained nonautonomous modulation equations in Cartesian form using the same parameters chosen thus far with different values of frequency detuning defined as $\Delta\sigma=\sigma-\sigma_c$. We found that resonance cancellation is achieved for a wide range of positive and negative frequency detunings (i.e., excitation frequency lower or greater than the control frequency) with differences with respect to the perfectly tuned case in the transient regime only. A small range of positive detunings (approximately 9.7 to 10) was found where cancellation can-

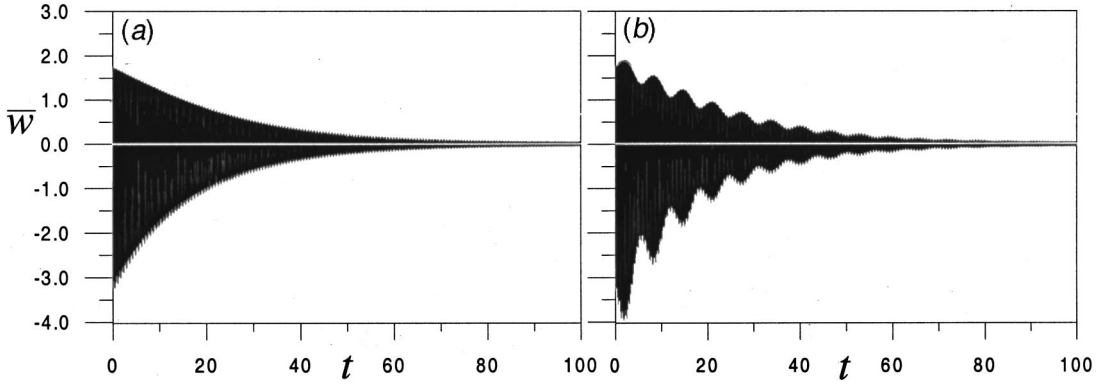


Fig. 8 Time histories of the controlled deflections (second-order solution) at $x=1/4$ in the detuned case when (*a*) $\Delta\sigma=\sigma-\sigma_c=-9.55$ and (*b*) $\Delta\sigma=8.95$ and $\psi_c=0$, $U_c=28.936$, $\sigma=-10$, $b=16.5$, $\mu=0.05$, $u_B=1$, $p(0)=1.5$, and $q(0)=-0.95$

not be achieved resulting in unbounded growth of the response in the investigated second-order solution. However, past this range, cancellation is recovered again. Figure 8 shows the response of the system in two detuned cases when (a) $\Delta\sigma=-9.55$ with a frequency variation of about 13.8% and (b) $\Delta\sigma=8.95$ with a frequency variation of about 13%. While in the negatively detuned case, no appreciable differences are noted, in the positively detuned case (prior the onset of instability at $\Delta\sigma \approx 9.7$), the presence of a large-amplitude excursion in the initial phase of the motion and modulation in the response are observed although the response decays with the same rate as in the perfectly tuned case.

Conclusions

In this work we have demonstrated the effectiveness of a nonlinear control strategy to cancel the parametrically forced oscillations of a shallow arch and further minimize the residual nonresonant oscillations. In particular, the control strategy inhibits activation of the principal parametric resonance using a point actuator at the midspan of the arch. The control algorithm is open loop because state parameters are not used for feedback. The difficulty arises because the disturbance input and actuator are not physically collocated and the control action is linearly orthogonal to the excited mode (first antisymmetric mode). The key idea is to rely on part of the nonlinear controller action which is not orthogonal to the mode and further has resonant effects onto it.

We show that a perturbation analysis can be used to develop intuition as to the proper form of the control input. In this paper, a one-half subharmonically resonant control law is used for sup pressing the parametric resonance by enforcing the effective parametric resonance coefficient to be below the activation threshold. This condition yields the control input in phase with the disturbance in the optimal case. On the other hand, the optimal gain value that minimizes the resulting steady-state vibrations, according to a chosen cost functional, is also determined.

The technique is shown to reduce the overall vibrations by orders of magnitude with respect to the uncontrolled case. The robustness of the proposed control methodology with respect to phase and frequency variations has also been demonstrated. Application of perturbation techniques to problems of this type seems to provide a unified approach that may be applicable to a broad class of weakly nonlinear distributed-parameter systems with either noncollocated inputs or linearly uncontrollable modes.

Acknowledgments

This work was partially supported by a M.U.R.S.T. COFIN99-2000 grant. Interesting comments by one of the reviewers are gratefully acknowledged.

Appendix

Boundary Value Problems And Solutions.

$$\mathcal{L}\chi_j-\omega_2^2\chi_j=0,\quad j=1,2 \tag{40}$$

$$\mathcal{L}\chi_3-4\omega_2^2\chi_3=\frac{1}{2}\psi''\langle\phi_2'\phi_2'\rangle \tag{41}$$

$$\mathcal{L}\chi_4=\frac{1}{2}\psi''\langle\phi_2'\phi_2'\rangle \tag{42}$$

$$\mathcal{L}\chi_5-9\omega_2^2\chi_5=\phi_2''\left[\langle\mathcal{B}'\psi'\rangle+\frac{1}{2}u_B\right] \tag{43}$$

$$\mathcal{L}\chi_6-9\omega_2^2\chi_6=\phi_2''\langle\mathcal{U}_c'\psi'\rangle \tag{44}$$

$$\mathcal{L}\chi_7-16\omega_2^2\chi_7=\mathcal{B}''\langle\mathcal{B}'\psi'\rangle+\frac{1}{2}\psi''\langle\mathcal{B}'\mathcal{B}'\rangle+\frac{1}{2}u_B\mathcal{B}'' \tag{45}$$

$$\mathcal{L}\chi_8-16\omega_2^2\chi_8=\mathcal{U}_c''\langle\mathcal{B}'\psi'\rangle+\mathcal{B}''\langle\mathcal{U}_c'\psi'\rangle+\psi''\langle\mathcal{U}_c'\mathcal{B}'\rangle+\frac{1}{2}u_B\mathcal{U}_c'' \tag{46}$$

$$\mathcal{L}\chi_9-16\omega_2^2\chi_9=\mathcal{U}_c''\langle\mathcal{U}_c'\psi'\rangle+\frac{1}{2}\psi''\langle\mathcal{U}_c'\mathcal{U}_c'\rangle \tag{47}$$

$$\mathcal{L}\chi_{10}=\mathcal{B}''\langle\mathcal{B}'\psi'\rangle+\frac{1}{2}\psi''\langle\mathcal{B}'\mathcal{B}'\rangle+\frac{1}{2}u_B\mathcal{B}''+\mathcal{U}_c''\langle\mathcal{U}_c'\psi'\rangle+\frac{1}{2}\psi''\langle\mathcal{U}_c'\mathcal{U}_c'\rangle \tag{48}$$

$$\mathcal{L}\chi_{11}=\mathcal{B}''\langle\mathcal{U}_c'\psi'\rangle+\mathcal{U}_c''\langle\mathcal{B}'\psi'\rangle+\psi''\langle\mathcal{U}_c'\mathcal{B}'\rangle \tag{49}$$

subject to the boundary conditions

$$\chi_i=0 \text{ and } \chi_i''=0 \text{ at } x=0 \text{ and } x=1. \tag{50}$$

The functions η_j, in turn, are given by

$$\eta_j=i\omega_2\chi_j-i\frac{K_j}{2\omega_2}\phi_2,\quad j=1,2 \tag{51}$$

$$\eta_3=2i\omega_2\chi_3,\quad \eta_j=3i\omega_2\chi_j,\quad j=5,6 \tag{52}$$

$$\eta_j=4i\omega_2\chi_j,\quad j=7,8,9,\quad \eta_{10}=\eta_{11}=0 \tag{53}$$

Solving (40) and (50) and using (51), we obtain

$$\chi_j=d_j\phi_2(x),\quad \eta_j=i\left[\omega_2 d_j-\frac{K_j}{2\omega_2}\right]\phi_2(x),\quad j=1,2. \tag{54}$$

To remove the indeterminacy associated with the arbitrary constants d_j, we impose that the functions χ_1 and η_1 and χ_2 and η_2 are orthogonal to the adjoint $(i\omega_2,1)\phi_2(x)$. The result is

$$\chi_j=\frac{K_j}{4\omega_2^2}\phi_2(x),\quad \eta_j=-i\omega_2\chi_j,\quad j=1,2. \tag{55}$$

Solving the remaining boundary value problems (41)–(49) with boundary conditions (50), we obtain the following shape functions:

$$\chi_3=-\frac{4b}{b^2-126}\sin\pi x,\quad \chi_4=-\frac{4b}{b^2+2}\sin\pi x \tag{56}$$

$$\chi_5=-\frac{63u_B}{32\pi^2(b^2-126)}\phi_2,\quad \chi_6=\frac{bU_c}{32\pi^4(b^2-126)}\phi_2, \tag{57}$$

$$\chi_7=-\frac{u_B^2b(b^2+252)}{2\pi^4(b^2-126)^2(b^2-510)}\sin\pi x \tag{58}$$

$$\chi_8=\frac{U_cu_B}{\pi^6(b^2-126)}\left\{\frac{4(b^2+63)}{(b^2-126)(b^2-510)}\sin\pi x+63\sum_{k=1}^{\infty}\left[\frac{s_{2k+1}(2k+1)^2}{[(2k+1)^4-256][(2k+1)^4-64]}\right]\times\sin(2k+1)\pi x\right\} \tag{59}$$

$$\chi_9=-\frac{2bU_c^2}{\pi^4(b^2-510)}\left[\frac{2}{\pi^4(b^2-126)^2}+\frac{\pi^2}{2}S_1^{(\infty)}\right]\sin\pi x-\frac{bU_c^2}{\pi^8(b^2-126)}\sum_{k=1}^{\infty}\left\{\frac{s_{2k+1}(2k+1)^2}{[(2k+1)^4-256][(2k+1)^4-64]}\right\}\times\sin(2k+1)\pi x \tag{60}$$

$$\chi_{10}=-\left[\frac{2bU_c^2}{\pi^4(b^2+2)}\left(\frac{2}{\pi^4(b^2-126)^2}+\frac{\pi^2}{2}S_1^{(\infty)}\right)-\frac{bu_B^2(b^2+252)}{2\pi^4(b^2+2)(b^2-126)^2}\right]\sin\ \pi x-\frac{bU_c^2}{\pi^8(b^2-126)}\sum_{k=1}^{\infty}\left[\frac{s_{2k+1}}{(2k+1)^2[(2k+1)^4-64]}\right]\times\sin(2k+1)\pi x \tag{61}$$

$$\chi_{11}=\frac{b^2U_cu_B}{\pi^6(b^2-126)}\left\{\frac{6}{(b^2-126)(b^2+2)}\sin\ \pi x+\frac{1}{2}\sum_{k=1}^{\infty}\left[\frac{s_{2k+1}}{(2k+1)^2[(2k+1)^4-64]}\right]\sin(2k+1)\pi x\right\} \tag{62}$$

where

$$S_1^{(\infty)}=\frac{1}{2\pi^6}\left[\frac{4}{(b^2-126)^2}+\sum_{k=1}^{\infty}\frac{s_{2k+1}^2(2k+1)^2}{[(2k+1)^4-64]^2}\right] \tag{63}$$

and $s_{2k+1}=\sin[(2k+1)\pi/2]$.

References

[1] Nayfeh, A. H., and Mook, D. T., 1979, *Nonlinear Oscillations,* Wiley-Interscience, New York.

[2] Chin, C. M., Nayfeh, A. H., and Lacarbonara, W., 1997, ''Two-to-One Internal Resonances in Parametrically Excited Buckled Beams,'' *Proceedings of the 38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials*, Kissimmee, FL, Apr. 7–10, AIAA Paper No. 97-1081.

[3] Fujino, Y., Warnitchai, P., and Pacheco, B. M., 1993, ''Active Stiffness Control of Cable Vibration,'' ASME J. Appl. Mech. **60**, pp. 948–953.

[4] Gattulli, V., Pasca, M., and Vestroni, F., 1997, ''Nonlinear Oscillations of a Nonresonant Cable Under In-Plane Excitation With a Longitudinal Control,'' Nonlinear Dyn. **14**, pp. 139–156.

[5] Oueini, S. S., Nayfeh, A. H., and Pratt, J. R., 1998, ''A Nonlinear Vibration Absorber for Flexible Structures,'' Nonlinear Dyn. **15**, pp. 259–282.

[6] Yabuno, H., Kawazoe, J., and Aoshima, N., 1999, ''Suppression of Parametric Resonance of a Cantiliver Beam by a Pendulum-Type Vibration Absorber,'' *Proceedings of the 17th Biennial ASME Conference on Mechanical Vibration and Noise*, Las Vegas, NV, Sept. 12–15, ASME, New York, Paper No. DETC99/VIB-8072.

[7] Maschke, B. M. J., and van der Schaft, A. J., 2000, ''Port Controlled Hamiltonian Representation of Distributed Parameter Systems,'' *Proceedings of the IFAC Workshop Lagrangian and Hamiltonian Methods for Nonlinear Control*, N. E. Leonard and R. Ortega, eds., Princeton University, Princeton, NJ, March 16–18, Elsevier Science, Oxford, UK.

[8] Ortega, R., van der Schaft, A. J., and Maschke, B. M. J., 1999, ''Stabilization of Port Controlled Hamiltonian Systems,'' *Stability and Stabilization of Nonlinear Systems*, Vol. 246, D. Aeyels, F. Lamnabhi-Lagarrigue, and A. J. van der Schaft, eds., Springer-Verlag, New York, pp. 239–260.

[9] Nayfeh, A. H., 1984, ''Interaction of Fundamental Parametric Resonances with Subharmonic Resonances of Order One-Half,'' *Proceedings of the 25th Structures, Structural Dynamics and Materials Conference*, Palm Springs, CA, May 14–16, AIAA, Washington, DC.

[10] Soper, R. R., Lacarbonara, W., Chin, C. M., Nayfeh, A. H., and Mook, D. T., 2001, ''Open-Loop Resonance-Cancellation Control for a Base-Excited Pendulum,'' J. Vib. Control (in press).

[11] Algrain, M., Hardt, S., and Ehlers, D., 1997, ''A Phase-Lock-Loop-Based Control System for Suppressing Periodic Vibration in Smart Structural Systems,'' Smart Mater. Struct. **6**, pp. 10–22.

[12] Mettler, E., 1962, *Dynamic Buckling* in *Handbook of Engineering Mechanics*, W. Flugge, ed., McGraw-Hill, New York.

[13] Nayfeh, A. H., 2000, *Nonlinear Interactions*, Wiley-Interscience, New York.

[14] Lacarbonara, W., Nayfeh, A. H., and Kreider, W., 1998, ''Experimental Validation of Reduction Methods for Nonlinear Vibrations of Distributed-Parameter Systems: Analysis of a Buckled Beam,'' Nonlinear Dyn. **17**, pp. 95–117.

[15] Lacarbonara, W., 1999, ''Direct Treatment and Discretizations of Non-Linear Spatially Continuous Systems,'' J. Sound Vib. **221**, pp. 849–866.

F. E. Udwadia
Professor of Aerospace and Mechanical Engineering, Mathematics, and Civil Engineering,
University of Southern California,
Los Angeles, CA 90089-1453

R. E. Kalaba
Professor of Biomedical Engineering, Economics, and Electrical Engineering,
University of Southern California,
Los Angeles, CA 90089

What is the General Form of the Explicit Equations of Motion for Constrained Mechanical Systems?

This paper presents the general form of the explicit equations of motion for mechanical systems. The systems may have holonomic and/or nonholonomic constraints, and the constraint forces may or may not satisfy D'Alembert's principle at each instant of time. The explicit equations lead to new fundamental principles of analytical mechanics.
[DOI: 10.1115/1.1459071]

Introduction

Since its inception more than 200 years ago, analytical mechanics has been continually drawn to the determination of the equations of motion for constrained mechanical systems. Following the fundamental work of Lagrange [1] who bequeathed to us the so-called Lagrange multipliers in the process of determining these equations, numerous scientists and mathematicians have attempted this central problem of analytical dynamics. A comprehensive reference list would run into several hundreds; hence we shall provide here, by way of a thumbnail historical review of the subject, only some of the significant milestones and discoveries. In 1829, Gauss [2] introduced a general principle for handling constrained motion, which is commonly referred to today as Gauss's Principle; Gibbs [3] and Appell [4] independently obtained the so-called Gibbs-Appell equations of motion using the concept of (felicitously chosen) quasi-coordinates; Poincare [5], using group theoretic methods, generalized Lagrange's equations to include general quasi-coordinates; and Dirac [6], in a series of papers provided an algorithm to give the Lagrange multipliers for constrained, singular Hamiltonian systems. Udwadia and Kalaba [7] gave the explicit equations of motion for constrained mechanical systems using generalized inverses of matrices, a concept that was independently discovered by Moore [8] and Penrose [9]. The use of this powerful concept, which was further developed from the late 1950s to the 1980s, allows the generalized-inverse equations (Udwadia and Kalaba [7]) to go beyond, in a sense, those provided earlier; for, they are valid for sets of constraints that could be nonlinear in the generalized velocities, and that could be functionally dependent. Thus the problem of obtaining the equations of motion for constrained mechanical systems has a history that is indeed as long as that of analytical dynamics itself.

Yet, all these efforts have been solely targeted towards obtaining the equations of motion for holonomically and nonholonomically constrained systems that *all* obey D'Alembert's principle of virtual work at each instant of time. This principle, though introduced by D'Alembert, was precisely stated for the first time by Lagrange. The principle in effect makes *an assumption* about the nature of the forces of constraint that act on a mechanical system. It assumes that at each instant of time, *t*, during the motion of the mechanical system, the constraint forces do *no* work under virtual displacements.

This seemingly sweeping assumption is indeed a tribute to the genius of Lagrange, because: (1) it gives exactly the right amount of additional information regarding the nature of the constraint forces in a general constrained mechanical system so that the equations of motion are *uniquely* determined, and are thus in conformity with practical observation; (2) in the mathematical modeling of a mechanical system, it obviates the need for the mechanician to investigate each specific mechanical system at hand and to determine the nature of the constraint forces prevalent; and, (3) it yields equations of motion for constrained systems that seem to work well (or at least sufficiently well) in numerous practical situations.

However, there are many mechanical systems that are commonplace in Nature where D'Alembert's principle is not valid, such as when sliding friction becomes important. Such situations have so far been considered to lie beyond the compass of the Lagrangian formulation of mechanics. As stated by Goldstein [10], ''This [total work done by forces of constraint under virtual displacements equal to zero] is no longer true if sliding friction is present, and we must exclude such systems from our [Lagrangian] formulation.'' And Pars [11] (p. 14) in his treatise on analytical dynamics writes, ''There are in fact systems for which the principle enunciated [D'Alembert's Principle] . . . does not hold. But such systems will not be considered in this book.''

Constraint forces that *do* work under virtual displacements are called nonideal constraint forces, and such constraints themselves are often referred to as being nonideal. While it is possible, at times, to handle problems with holonomic, nonideal constraints (like sliding friction) by using a Newtonian approach, to date we do not have a general formulation for obtaining the equations of motion for systems where we have nonholonomic, nonideal constraints, i.e., nonholonomic constraints where the constraint forces do work under virtual displacements. The aim of this paper is to include such systems within the Lagrangian formulation of mechanics, and further to develop the general form of the explicit equations of motion for constrained systems that may or may not obey D'Alembert's principle at each instant of time. The approach we follow here is based on linear algebra, and it is different from that of Refs. [12], [13], and [14]. It leads us to the general structure of the equation of motion for constrained systems, and culminates in the statement of two fundamental principles of analytical dynamics.

Formulation of the Problem of Constrained Motion

Consider an ''unconstrained'' mechanical system described by the Lagrange equations

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received and accepted by the ASME Applied Mechanics Division, April 18, 2000. Associate Editor: L. T. Wheeler. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

$$M(q,t)\ddot{q}=Q(q,\dot{q},t), \quad q(0)=q_0, \quad \dot{q}(0)=\dot{q}_0 \tag{1}$$

where $q(t)$ is the n-vector (i.e., n by 1 vector) of generalized coordinates, M is an n by n symmetric, positive-definite matrix, Q is the "known" n-vector of impressed (also, called "given") forces, and the dots refer to differentiation with respect to time. By unconstrained, we mean that the components of the n-vector $\dot{q}_0$ can be arbitrarily specified. By "known," we mean that the n-vector Q is a known function of its arguments. The acceleration, a, of the unconstrained system at any time t is then given by the relation $a(q,\dot{q},t)=M^{-1}(q,t)Q(q,\dot{q},t)$.

We next subject the system to a set of $m=h+s$ consistent, equality constraints of the form

$$\varphi(q,t)=0 \tag{2}$$

and

$$\psi(q,\dot{q},t)=0, \tag{3}$$

where φ is an h-vector and ψ an s-vector. Furthermore, we shall assume that the initial conditions q_0 and $\dot{q}_0$ satisfy these constraint equations at time $t=0$, i.e., $\varphi(q_0,0)=0$, $\dot{\varphi}(q_0,\dot{q}_0,0)=0$, and $\psi(q_0,\dot{q}_0,0)=0$.

Assuming that Eqs. (2) and (3) are sufficiently smooth,[1] we differentiate Eq. (2) twice with respect to time, and Eq. (3) once with respect to time, to obtain an equation of the form

$$A(q,\dot{q},t)\ddot{q}=b(q,\dot{q},t), \tag{4}$$

where the matrix A is m by n, and b is the m-vector that results from carrying out the differentiations. We place no restrictions on the rank of the matrix A.

This set of constraint equations includes, among others, the usual holonomic, nonholonomic, scleronomic, rheonomic, catastatic, and acatastatic varieties of constraints; combinations of such constraints may also be permitted in Eq. (4). Furthermore, the functions in (3) could be nonlinear in $\dot{q}$, and the m constraint equations need not be independent of one another.

It is important to note that Eq. (4), together with the initial conditions, is equivalent to Eqs. (2) and (3).

The equation of motion of the constrained mechanical system can then be expressed as

$$M(q,t)\ddot{q}=Q(q,\dot{q},t)+Q^c(q,\dot{q},t), \quad q(0)=q_0, \quad \dot{q}(0)=\dot{q}_0 \tag{5}$$

where the additional "constraint force" n-vector, $Q^c(q,\dot{q},t)$, arises by virtue of the constraints that are imposed on the unconstrained system, which we have described by Eq. (1). Since the n-vector Q is known, our aim is to determine a *general* explicit form for Q^c at any time t.

We shall see below that in any constrained mechanical system, the total constraint force n-vector, Q^c, at each instant of time t, can be thought of as made up of two components: $Q^c=Q_i^c+Q_{ni}^c$. The first component corresponds to the force of constraint, Q_i^c, *that would act were all the constraints ideal at that instant of time*; the second component, Q_{ni}^c, arises because of the nonideal nature of the constraints. This latter component is *situation specific* and needs to be specified by the mechanician entrusted with modeling the mechanical system. However, we shall show that this component too must always occur in the explicit equation of motion *in a specific form*.

In what follows, for brevity, we shall suppress the arguments of the various quantities, unless necessary for purposes of clarification.

[1]We assume throughout this paper that the presence of constraints does not change the rank of the matrix M. This is almost always true in mechanical systems.

The General Form of the Explicit Equation of Motion for any Constrained Mechanical Systems

We begin by stating our general result in the following three-part statement.

(1) The general "explicit" equation of motion at time t for *any* constrained mechanical system, *whether or not the constraint forces satisfy D'Alembert's Principle at that time t,* is given by

$$\begin{aligned} M\ddot{q}&=Q+Q^c=Q+Q_i^c+Q_{ni}^c \\ &=Q+M^{1/2}B^+(b-AM^{-1}Q)+M^{1/2}(I-B^+B)z \end{aligned} \tag{6}$$

where the matrix $B=AM^{-1/2}$, B^+ is the generalized inverse[2] of the matrix B, and $z(q(t),\dot{q}(t),t)$ is some suitable n-vector. (When z is C^1, Eq. (6) yields a unique solution.) The matrix A is defined in relation (4), as is the m-vector b. The n-vector Q is the impressed force. By "explicit" we mean here that the acceleration n-vector, $\ddot{q}$, on the left-hand side of Eq. (6) is explicitly expressed in terms of quantities that are functions of q, $\dot{q}$, and t on the right-hand side.

Alternately stated, the total constraint force n-vector, Q^c, at any instant of time t is made up of the sum of two components Q_i^c and Q_{ni}^c that can be explicitly written as

$$Q_i^c=M^{1/2}B^+(b-AM^{-1}Q), \tag{7}$$

and,

$$Q_{ni}^c=M^{1/2}(I-B^+B)z. \tag{8}$$

(2) To mathematically model a *given* constrained mechanical system adequately, the mechanician must *specify* the vector $z(q,\dot{q},t)$ in the third member on the right-hand side of Eq. (6) at each instant of time. This may be done by inspection of the specific system at hand, by analogy with other systems that the mechanician may have dealt with in the past, by experimentation with the specific system or similar systems, or otherwise.

(3) However, no matter how the mechanician comes up with the prescription of the n-vector z for adequately modeling a *given* constrained mechanical system under consideration, specification of this n-vector at each time t uniquely determines Q_{ni}^c, and hence the acceleration n-vector, $\ddot{q}(t)$, of the constrained system. Such a prescription of $z(t)$ is *equivalent* to prescribing the work done by all the constraint forces under virtual displacements at that time t, in the following sense.

(a) When the vector $z(t)$ is prescribed, it can always be expressed as

$$z(t)=M^{-1/2}(q,t)C(q,\dot{q},t) \tag{9}$$

since, M is a positive definite matrix. The total work done, $W:=v^TQ^c$, by all the forces of constraint under (nonzero) virtual displacements v at time t, is then given by

$$W(t):=v(t)^TQ^c=v(t)^TC(q,\dot{q},t). \tag{10}$$

(b) When, for a given specific constrained mechanical system, the work done, W, at time t by the forces of constraint under virtual displacements v is prescribed through specification of the n-vector $C(q,\dot{q},t)$ such that

$$W(t)=v(t)^TC(q,\dot{q},t), \tag{11}$$

this determines the equation of motion of the constrained system *uniquely* at time t. This equation of motion is obtained by setting $z(t)=M^{-1/2}(q,t)C(q,\dot{q},t)$, in Eq. (6). The work done, $W(t)$, may be positive, zero, or negative, at the instant of time t. □

[2]Some of the basic properties of the Moore-Penrose generalized inverse that are used throughout this paper may be found in Chapter 2 of Ref. [15].

We note from Eq. (9) above, that prescribing z to be the zero n-vector at any time t, is equivalent to specifying $C=0$ at that specific time t, and then by (10), the constraint forces do *no* work under virtual displacements and therefore they satisfy D'Alembert's principle at that instant of time t. In what follows we shall also show that when the constraints do no work under virtual displacements at time t, because of Eq. (10), the n-vector C must belong to the range space of A^T; the third member on the right in Eq. (6) then becomes zero at that time. Further, if throughout the motion of the constrained system the work done by the constraint forces under virtual displacements is zero, then the third member on the right-hand side in Eq. (6) disappears for all time. The equation of motion (6) then becomes

$$M\ddot{q}=Q+Q^c=Q+Q_i^c=Q+M^{1/2}B^+(b-AM^{-1}Q), \quad (12)$$

which is identical to that obtained by Udwadia and Kalaba [7] for systems that obey D'Alembert's principle. Equation (12) is equivalent to the Gibbs-Appell equations (see Ref. [15]). We then see that the component Q_i^c in Eq. (7) therefore gives the constraint force at time t that *would be generated were all the constraints ideal at that time*. And Q_{ni}^c explicitly gives the contribution to the total constraint force, Q^c, made by the nonideal nature of the constraints.

Were the acceleration, $a=M^{-1}Q$, of the unconstrained system at time t to be inserted into the equation of constraint (4), this equation would not, in general, be satisfied at that time. The extent to which the constraint (Eq. (4)) would not be satisfied by this acceleration, a, of the unconstrained system at time t would then be given by

$$e=b-Aa=b-AM^{-1}Q. \quad (13)$$

The force of constraint can now be rewritten as

$$Q^c=Q_i^c+Q_{ni}^c=M^{1/2}B^+e+M^{1/2}(I-B^+B)z. \quad (14)$$

Also, the effect of this constraint force in altering the acceleration of the unconstrained system can be explicitly determined. For, the deviation, $\Delta\ddot{q}$, at time t of the acceleration of the constrained system from that of the unconstrained system becomes, by Eq. (6),

$$\Delta\ddot{q}=\ddot{q}-a=M^{-1/2}B^+e+M^{-1/2}(I-B^+B)z. \quad (15)$$

Equations (14) and (15) lead us to a new fundamental principle of Lagrangian mechanics which we now state in two equivalent forms.

1 A constrained mechanical system evolves in such a way that, at each instant of time, the deviation, $\Delta\ddot{q}$, of its acceleration from what it would have been at that instant had there been no constraints on it, is given by a sum of two components: the first component is proportional to the extent, e, to which the unconstrained acceleration does not satisfy the constraints at that instant of time, the matrix of proportionality being the matrix $M^{-1/2}B^+$; the second is proportional to an n-vector z that needs, in general, to be specified at each instant of time, the matrix of proportionality being $M^{-1/2}(I-B^+B)$, where $B=AM^{-1/2}$. The specification of z at any time, t, is dependent on the nature of the forces of constraint that are generated. Its specification for a given system at hand is tantamount to the specification of the total work done under virtual displacements by all the forces of constraint at that time. Such a specification of the work done at each instant of time uniquely determines the equation of motion of the constrained system.

2 At each instant of time t, the force of constraint acting on a constrained mechanical system is made up of two components: the first component is proportional to the extent, e, to which the unconstrained acceleration of the system does not satisfy the constraints at that instant of time, and the matrix of proportionality is $M^{1/2}B^+$; the second is proportional to an n-vector z that, in general, needs specification at each instant of time, the matrix of proportionality being $M^{1/2}(I-B^+B)$, where $B=AM^{-1/2}$. This vector z is specific to a given mechanical system and needs to be prescribed by the mechanician who is modeling the system. Whether or not the constraints are ideal, the first component is always present and constitutes the constraint force at the instant of time t that would have been generated were all the constraints ideal at that time. The second component depends on the nature of the constraint forces generated in the specific mechanical system that is being modeled; it prevails only when the total work done by the constraint forces under virtual displacements differs from zero.

Proof of the General Form of the Equations of Motion for Constrained Systems

We begin by considering the ''scaled accelerations'' defined by the relations

$$\ddot{q}_s=M^{1/2}\ddot{q}; \quad (16)$$

$$a_s=M^{-1/2}Q=M^{1/2}a; \quad (17)$$

and,

$$\ddot{q}_s^c=M^{-1/2}Q^c=M^{1/2}\ddot{q}^c. \quad (18)$$

By Eq. (5), we then have

$$\ddot{q}_s=a_s+\ddot{q}_s^c. \quad (19)$$

Furthermore, Eq. (4) can be expressed as

$$B\ddot{q}_s=b, \quad (20)$$

where

$$B=AM^{-1/2}. \quad (21)$$

Consider the matrices $T=B^+B$ and $N=(I-B^+B)$, where the matrix B^+ is the Moore-Penrose (MP) inverse of the matrix B. The matrix T is an orthogonal projection operator since $(B^+B)^T=B^+B$, and $T^2=(B^+B)(B^+B)=B^+B=T$. Also, N is an orthogonal projection operator since $(I-B^+B)^T=I-(B^+B)^T=I-B^+B$, and $N^2=N$. Since $R^n=\mathcal{R}(B^T)\oplus\mathcal{N}(B)$, any n-vector w has a unique orthogonal decomposition $w=B^+Bw+(I-B^+B)w$; and so also our n-vector $\ddot{q}_s$. This yields the identity

$$\ddot{q}_s=B^+B\ddot{q}_s+(I-B^+B)\ddot{q}_s. \quad (22)$$

Using relation (20) in the first member on the right, and relation (19) in the second member, we obtain

$$\ddot{q}_s=a_s+B^+(b-Ba_s)+(I-B^+B)\ddot{q}_s^c. \quad (23)$$

Comparison of Eq. (19) with Eq. (23) then yields

$$B^+B\ddot{q}_s^c=B^+(b-Ba_s) \quad (24)$$

which can be solved for $\ddot{q}_s^c$ to yield

$$\ddot{q}_s^c=B^+BB^+(b-Ba_s)+\{I-(B^+B)^+(B^+B)\}z$$
$$=B^+(b-Ba_s)+(I-B^+B)z \quad (25)$$

for some n-vector z.

Equation (18), then gives

$$Q^c=M^{1/2}B^+(b-Aa)+M^{1/2}(I-B^+B)z \quad (26)$$

and the general equation of motion of the constrained system, by Eq. (5), becomes

$$M\ddot{q}=Q+Q^c=Q+M^{1/2}B^+(b-Aa)+M^{1/2}(I-B^+B)z \quad (27)$$

where z is some n-vector. q.e.d.

To obtain the unique equation of motion for a *specific* mechanical system, the mechanician needs to prescribe the vector $z(q(t),\dot{q}(t),t)$ at each instant of time. Specification of the vector

$z(t)$ yields explicitly and uniquely the component Q^c_{ni} of the constraint force, Q^c, at each instant of time t. In fact, given an n-vector z at a specific time t, we can form the n-vector $C = M^{1/2}z$ at time t. The vector C can *now be interpreted* as providing the work done, $W = v^T C$, by the constraint force n-vector Q^c under virtual displacements v at time t.

We now show that Q^c_{ni} can also be uniquely determined at each instant of time t by specifying the work done by the constraint force n-vector, Q^c, under virtual displacements at that time. *Proof*: A virtual displacement is any nonzero n-vector v such that $Av = 0$ (see Ref. [15]). Using Eq. (21) this relation can also be written as $Av = (AM^{-1/2})M^{1/2}v = B(M^{1/2}v) = B\mu = 0$, where we have denoted the n-vector $M^{1/2}v$ by μ. Thus a virtual displacement can also be considered as any (nonzero) n-vector μ such that $B\mu = 0$. Using Eq. (27), the work done by the force of constraint under all virtual displacements v is then given by

$$\begin{aligned} W &:= v^T Q^c = v^T(Q^c_i + Q^c_{ni}) \\ &= v^T M^{1/2} B^+ (b - Aa) + v^T M^{1/2}(I - B^+ B)z \\ &= \mu^T B^+ (b - Aa) + \mu^T (I - B^+ B)z. \end{aligned} \tag{28}$$

The first member in the last expression on the right of equation (28) is zero since $B\mu = 0$ implies $\mu^T B^+ = 0$. *Hence the component Q^c_i of the total force of constraint, Q^c, does no work under virtual displacements*. Equation (28) then becomes

$$W := v^T Q^c = v^T Q^c_{ni} = \mu^T z = v^T (M^{1/2} z). \tag{29}$$

Let $W(t)$ to be prescribed at time t by the mechanician through a specification of the n-vector $C(q,\dot{q},t)$ so that $W := v^T Q^c = v^T C$. Then by Eq. (29), we have

$$v^T (M^{1/2} z) = v^T C. \tag{30}$$

Since v is such that $Av = 0$, this requires that

$$z = M^{-1/2}(C + A^T w) = M^{-1/2} C + B^T w \tag{31}$$

where w is any arbitrary m-vector. Using this expression for z in Eq. (27) we obtain the *unique* equation of motion of the constrained system to be

$$\begin{aligned} M\ddot{q} = Q + Q^c = Q + Q^c_i + Q^c_{ni} &= Q + M^{1/2} B^+ (b - Aa) \\ &+ M^{1/2}(I - B^+ B) M^{-1/2} C, \end{aligned} \tag{32}$$

since $(I - B^+ B)B^T = \{B(I - B^+ B)\}^T = 0$.

We now see that Eq. (6) is identical to Eq. (32) with $z = M^{-1/2}C$! The component of z in the range space of B^T—the second member on the right in Eq. (31)—does not affect Q^c_{ni}, and therefore the equation of motion of the constrained system.

Though the n-vector $C(t)$ specifies the work done, $W := v^T Q^c = v^T Q^c_{ni} = v^T C$, by the constraint force under all virtual displacements v at time t, Eq. (32) states that, in general, $Q^c_{ni} \neq C$. At instants of time t when $W = (v^T M^{1/2})(M^{-1/2} C(q,\dot{q},t)) = 0$, $M^{-1/2}C$ belongs to the range space of B^T, and hence by Eq. (32), $Q^c_{ni} = 0$ since $(I - B^+ B)B^T = 0$. If further, $W = 0$ for *all* time, then the force of constraint satisfies D'Alembert's principle, and $Q^c_{ni}(t) \equiv 0$; the equation of motion for the constrained system then reduces to that given in (12). At instants of time t when $M^{-1/2}C$ belongs to the null space of B, $Q^c_{ni} = C$. In general, the n-vector $M^{-1/2}C$ can have components in both the null space of B and the range space of B^T. We note that at each instant of time, it is only the component of $M^{-1/2}C$ in the null space of B that contributes to Q^c_{ni}, and hence to the equation of motion of the constrained system.

Conclusions

The equations of motion for constrained systems obtained to date have all been based upon D'Alembert's principle of virtual work. So far, no general equations of motion have been discovered within the Lagrangian formalism in situations where this central principle of analytical dynamics is not applicable.

This paper provides the general explicit form of the equation of motion for *any* holonomically and/or nonholonomically constrained mechanical system. The equation is

$$M\ddot{q} = Q + M^{1/2} B^+ (b - AM^{-1} Q) + M^{1/2}(I - B^+ B)z. \tag{33}$$

The n-vector Q is the given force, the m by n matrix A and the m-vector b are defined in Eq. (4), $B = AM^{-1/2}$, and B^+ is the generalized inverse of B. *The equation applies to all constrained mechanical systems whether or not they satisfy D'Alembert's principle*. The second member on the right in Eq. (33) explicitly gives the force of constraint, Q^c_i that *would have been* generated at time t were all the constraint forces ideal, and thus satisfy D'Alembert's principle. The third member on the right in Eq. (33) explicitly gives the contribution, Q^c_{ni}, to the total force of constraint because of the presence of nonideal constraints.

To obtain the equation of motion for a given, specific, mechanical system, the mechanician needs to provide the n-vector $z(q,\dot{q},t)$ suitably at each instant of time, thereby uniquely specifying the third member on the right in Eq. (33). The provision of this vector $z(t)$ depends on the judgement and discernment of the mechanician and may be determined by experiment, experience, intuition, inspection, or otherwise. However, no matter how this vector is arrived at, the total work done, $W(t) := v^T(t) Q^c(t)$, by the force of constraint under virtual displacements $v(t)$ at any instant of time t is always given by $v^T(t)C(t)$, where the n-vector $C(t) = M^{1/2}(q,t)z(q,\dot{q},t)$. This work, $W(t)$, may, in general, be positive, zero, or negative.

We show that to model a *given* constrained mechanical system adequately one needs, in general, to provide *more than* just the equations of constraint (Eqs. (2) and (3)), be they holonomic or nonholonomic. While at each instant of time the component Q^c_i of the total constraint force n-vector, Q^c, is determined solely from the kinematical description of the constraints (Eqs. (2) and (3)), to determine the component Q^c_{ni} one *always* needs to rely on the mechanician's discernment and judgement. However, as shown in the equation above, this component (see also Eq. (8)) *must* appear in a *specific form* in the explicit equation of motion of the constrained system. When the mechanical system satisfies D'Alembert's principle at every instant of time, $Q^c_{ni}(t) \equiv 0$, and the third member on the right in (33) becomes zero. Then our general equation yields the known equation of motion ([15]) for constrained systems that satisfy D'Alembert's principle.

It is perhaps noteworthy that though the equations of motion of even very simple mechanical systems are often highly nonlinear, the general form of the equation of motion obtained here relies on techniques from linear algebra. The fundamental principles of analytical dynamics obtained in this paper may have been impossible to state in such a simple form without the concept of the generalized inverse of a matrix, a concept first invented by Penrose [9].

The equation of motion obtained in this paper appears to be the simplest and most general so far discovered for mechanical systems within the framework of classical mechanics.

References

[1] Lagrange, J. L., 1787, *Mecanique Analytique*, Mme Ve Courcier, Paris.
[2] Gauss, C. F., 1829, "Uber Ein Neues Allgemeines Grundgesetz der Mechanik," J. Reine Agnew. Math., **4**, pp. 232–235.
[3] Gibbs, J. W., 1879, "On the Fundamental Formulas of Dynamics," Am. J. Math., **2**, pp. 49–64.
[4] Appell, P., 1899, "Sur une forme generale des equations de la dynamique," C. R. Acad. Sci., Paris, **129**, pp. 459–460.
[5] Poincare, H., 1901, "Sur une forme nouvelle des equations de la mecanique," C. R. Acad. Sci., Paris, **132**, pp. 369–371.
[6] Dirac, P. A. M., 1964, *Lectures in Quantum Mechanics*, Yeshiva University, New York.

[7] Udwadia, F. E., and Kalaba, R. E., 1992, ''A New Perspective on Constrained Motion,'' Proc. R. Soc. London, Ser. A, **439**, pp. 407–410.
[8] Moore, E. H., 1920, ''On the Reciprocal of the General Algebraic Matrix,'' Bull. Am. Math. Soc., **26**, pp. 394–395.
[9] Penrose, R., 1955, ''A Generalized Inverse of Matrices,'' Proc. Cambridge Philos. Soc., **51**, pp. 406–413.
[10] Goldstein, H., 1981, *Classical Mechanics*, Addison-Wesley, Reading, MA.
[11] Pars, L. A., 1979, *A Treatise on Analytical Dynamics*, Oxbow Press, Woodridge, CT.
[12] Udwadia, F. E., and Kalaba, R. E., 2000, ''Non-ideal Constraints and Lagrangian Dynamics,'' Journal of Aerospace Engineering, **13**, Jan., pp. 17–22.
[13] Udwadia, F. E., and Kalaba, R. E., 2001, ''Explicit Equations of Motion for Systems With Non-ideal Constraints,'' ASME J. Appl. Mech., **68**, pp. 462–467.
[14] Udwadia, F. E., 2000, ''Fundamental Principles and Lagrangian Dynamics: Mechanical Systems With Non-ideal, Holonomic, and Nonholonomic Constraints,'' J. Math. Anal. Appl., **251**, pp. 341–355.
[15] Udwadia, F. E., and Kalaba, R. E., 1996, *Analytical Dynamics: A New Approach*, Cambridge University Press, England.

Exact Analysis of Dynamic Sliding Indentation at any Constant Speed on an Orthotropic or Transversely Isotropic Half-Space

L. M. Brock
Fellow ASME
Mechanical Engineering,
University of Kentucky
Lexington, KY 40506
e-mail: brock@engr.uky.edu

A plane-strain study of steady sliding by a smooth rigid indentor at any constant speed on a class of orthotropic or transversely isotropic half-spaces is performed. Exact solutions for the full displacement fields are constructed, and applied to the case of the generic parabolic indentor. The closed-form results obtained confirm previous observations that physically acceptable solutions arise for sliding speeds below the Rayleigh speed, for a single critical transonic speed, and for all supersonic speeds. Continuity of contact zone traction is lost for the latter two cases. Calculations for five representative materials indicate that contact zone width achieves minimum values at high, but not critical, subsonic sliding speeds. A key feature of the analysis is the factorization that gives, despite anisotropy, solution expressions that are rather simple in form. In particular, a compact function of the Rayleigh-type emerges that leads to a simple exact formula for the Rayleigh speed itself. [DOI: 10.1115/1.1464874]

Introduction

Studies of rapid sliding contact by rigid indentors allow insight into the operation of mechanisms, and has been modeled, e.g. [1–4], as a dynamic process involving linear isotropic elastic solids. More recently, Georgiadis and Barber [5] and Brock and Georgiadis [6] have shown for, respectively, smooth and frictional contact, that physically acceptable solutions may not exist for sliding speeds in various portions of the sub, tran, and supersonic ranges. Similar behavior is seen for sliding on hyperelastic isotropic solids under prestress ([7]). That study treats, in the manner of [8,9], sliding contact as the superposition of infinitesimal deformations upon the possibly finite deformations due to prestress. Thus, the anisotropy induced by the prestress influences solution behavior.

The present article complements efforts such as [1–7] by treating initially unstressed solids that are either orthotropic or transversely isotropic in their rest states. A plane-strain analysis of steady sliding is performed, the rigid indentor is smooth, and the sliding speed can be any constant value. In keeping with standard definitions ([6,10]), the subsonic range refers to speeds below the shear wave speed, the transonic range denotes speeds between the shear and dilatational wave speeds, and the supersonic range refers to speeds that exceed the dilatational wave speed. An unbounded half-space is treated. The results will, therefore, be most applicable when the sliding duration is brief enough and the indented body large enough to neglect the effects of other boundaries.

Exact solutions are obtained, and key steps in the analysis are the factorization of certain functions in the integral transform space that give, despite anisotropy, simple expressions for field quantities. One factorization, in particular, produces a function of the Rayleigh type, and allows a compact exact formula for the Rayleigh speed along the material surface. The analysis begins in the next section with the basic equations for the material.

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, April 11, 2001; final revision, November 6, 2001. Associate Editor: A. K. Mal. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

Basic Equations

Consider a half-space, defined in terms of Cartesian coordinates (x, y, z) as the region $y>0$. The half-space material is of a class of linear homogeneous anisotropic solids whose nontrivial governing equations in plane strain in the absence of body forces are

$$c_{11}u_{x,xx}+c_{44}u_{x,yy}+(c_{13}+c_{44})u_{y,xy}=\rho\ddot{u}_x \tag{1a}$$

$$c_{44}u_{y,xx}+c_{33}u_{y,yy}+(c_{13}+c_{44})u_{x,xy}=\rho\ddot{u}_y \tag{1b}$$

and are augmented by the associated stress-strain formulas

$$\sigma_x=c_{11}u_{x,x}+c_{13}u_{y,y} \tag{2a}$$

$$\sigma_y=c_{33}u_{y,y}+c_{13}u_{x,x} \tag{2b}$$

$$\sigma_{xy}=c_{44}(u_{x,y}+u_{y,x}). \tag{2c}$$

These equations hold for both orthotropic and transversely isotropic materials, where the x and y-axes are axes of material symmetry. The (u_x, u_y) are the (x, y)-components of displacement, while $\dot{(\ )}$ and $(\ )_{,s}$ denote differentiation by time and a variable s, respectively. The constants $(c_{11}, c_{13}, c_{33}, c_{44})$ are a subset of the elasticities $c_{ik}(i,k=1,2,\ldots,6)$ that appear in the generalized Hooke's law ([11]), and ρ is the mass density. Equation (1) is a special case of a more general form that involves four constants that can be linearly related to various subsets of c_{ik} ([12]). Discussions of relations between crystal structure and the elasticities can be found in [13,14]. In the present case, the isotropic limit can be extracted by setting $c_{11}=c_{33}=\lambda+2\mu$, $c_{13}=\lambda$, $c_{44}=\mu$, where (λ,μ) are the Lame' constants.

The half-space is at rest when a smooth rigid indentor of infinite length and invariant profile in the z-direction is pressed into the half-space surface $y=0$ with compressive force (per unit of length) N. The indentor is simultaneously translated in the positive x-direction with a constant speed v. Eventually, a dynamic plane-strain state of steady sliding is reached, as depicted schematically in Fig. 1. As indicated there, it is convenient to translate the coordinates (x, y, z) with the indentor, so that the boundary conditions along $y=0$ take the form

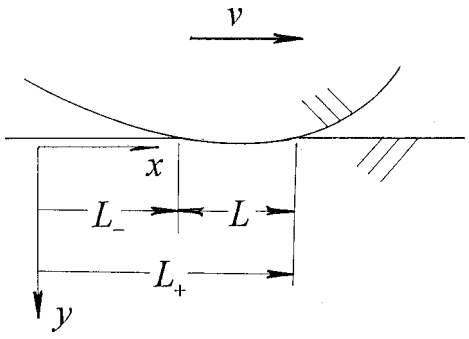


Fig. 1 Schematic of sliding rigid indentation

$$\sigma_{xy}=0; \tag{3a}$$

$$\sigma_y=0 (x\notin L), \tag{3b}$$

$$u_y=V(x) (x\in L). \tag{3c}$$

In Fig. 1, $x=L_\pm$ locates the leading and trailing edges of the indentor contact zone. The symbol L is used to represent both the contact zone itself, and its width $L=L_+-L_-$. The function $V(x)$ is the normal motion imposed by the indentor in the contact zone. The constants $L_\pm$ are a priori unknown.

Because the process is one of steady sliding in plane strain, field quantities depend only on (x, y), and $(\,)$ in the inertial frame can be replaced by the operator $-v(\,)_{,x}$. For convenience, therefore, the dimensionless quantities

$$\alpha=\frac{c_{33}}{c_{44}},\quad \beta=\frac{c_{11}}{c_{44}},\quad \gamma=1+\alpha\beta-m^2,\quad m=1+\frac{c_{13}}{c_{44}},\quad c=\frac{v}{v_r} \tag{4}$$

are introduced, where $(\alpha, \beta, \gamma, m)$ follow from [15], and the speed

$$v_r=\sqrt{\frac{c_{44}}{\rho}} \tag{5}$$

is, in the isotropic limit, the classical rotational wave speed ([10]). Equations (1) and (2) become

$$(\beta-c^2)u_{x,xx}+u_{x,yy}+mu_{y,xy}=0 \tag{6a}$$

$$u_{y,xx}+(\alpha-c^2)u_{y,yy}+mu_{x,xy}=0 \tag{6b}$$

and

$$\frac{1}{c_{44}}\sigma_x=\beta u_{x,x}+(m-1)u_{y,y} \tag{7a}$$

$$\frac{1}{c_{44}}\sigma_y=(m-1)u_{x,x}+\alpha u_{y,y} \tag{7b}$$

$$\frac{1}{c_{44}}\sigma_{xy}=u_{x,y}+u_{y,x}. \tag{7c}$$

For purposes of illustration, we consider, after [15,16], the constraints

$$2\sqrt{\alpha\beta}\leqslant\gamma\leqslant 1+\alpha\beta \quad (1<\beta<\alpha) \tag{8a}$$

$$\alpha+\beta\leqslant\gamma\leqslant 1+\alpha\beta \quad (1<\alpha<\beta) \tag{8b}$$

$$2\alpha\leqslant\gamma\leqslant 1+\alpha^2 \quad (1<\beta=\alpha). \tag{8c}$$

As will be seen, condition (8) guarantees that the rotational and dilatational wave speeds associated with the x-axis of material symmetry arise as distinct, real-valued branch points in the complex c-plane. The class of materials governed by (8) includes beryl, cobalt, ice, magnesium, and titanium, as well as the isotropic limit. Because displacements in steady motion can be determined only to within an arbitrary rigid-body motion, boundary condition (3c) can be replaced by

$$u_{y,x}=\frac{dV(x)}{dx} \quad (x\in L) \tag{9}$$

along $y=0$. In addition, displacements and their gradients should vanish when $\sqrt{x^2+y^2}\to\infty$, $y>0$ and be nonsingular and continuous almost everywhere. This latter requirement, as well as the assumption, implicit in Fig. 1 and (3a–c) and (9), that multiple contact zones ([4]) do not in fact arise, can be justified in part by requiring that $(V, dV/dx, d^2V/dx^2)$ be finite and continuous for $x\in L$. Then, the resultant of σ_y on the contact zone should be the specified compressive force $-N$, and the contact zone parameters $L_\pm$ must be determined as part of the solution. Finally, two unilateral constraints ([5]) must be satisfied: (a) contact zone normal stress is nontensile, and (b) indentor and half-space surfaces do not interpenetrate.

Solution Candidates

Following a standard ([17]) procedure for two-dimensional mixed boundary value problems in classical elasticity, the solution to this problem is obtained by first considering Eqs. (6)–(8) and the boundedness/continuity conditions, but with (3a,b) and (9) replaced by the unmixed conditions

$$\sigma_{xy}=0, \quad \sigma_y=\sigma(x) \tag{10}$$

for $y=0$, where $\sigma(x)$ represents the contact zone normal traction; it must vanish identically for $x\in L$, and should be continuous at $x=L_\pm$. The solution to this simpler problem will provide candidates for the sliding contact solution if a $\sigma(x)$ can be found such that (9) is also satisfied.

Solutions to (6)–(8) and (10) are found by integral transform techniques, and expressions for $(u_{x,x}, u_{y,x})$ for all $y>0$ are given in the Appendix. For the subsonic sliding case $(0<v<v_r)$ use of (*A*13*b*) and (*A*18) in (9) gives the equation

$$\frac{A(b+a)}{\pi c_{44}R}\fint_L\frac{\sigma(t)dt}{t-x}=\frac{dV(x)}{dx} \quad (x\in L). \tag{11}$$

Here $\fint$ denotes Cauchy principal value integration, (*A*, *a*, *B*, *b*, *R*) are defined by (*A*4) and (*A*7) and, as in the Appendix, the symbol L affixed to an integral signifies integration over the real interval (L_-, L_+). The relative simplicity of the left-hand side, when compared with (*A*13*b*) itself, follows from the observation made in the Appendix, that setting $y=0$ allows an explicit cancellation of terms proportional to *b*-*a*.

Equation (11) is a standard Cauchy integral equation ([17]) whose solution is

$$\sigma(x)=\frac{-c_{44}R}{\pi A(b+a)}\sqrt{\frac{x-L_-}{L_+-x}}\fint_L\frac{dV(t)}{dt}\sqrt{\frac{L_+-t}{t-L_-}}\frac{dt}{t-x} \quad (x\in L). \tag{12}$$

The result is appropriately bounded at L_-, but boundedness as $x\to L_+$ occurs only when

$$\int_L\frac{dV(t)}{dt}\frac{dt}{\sqrt{t-L_-}\sqrt{L_+-t}}=0. \tag{13}$$

This requirement serves as one equation for determining $L_\pm$. The other arises from the requirement that $-N$ be the resultant of σ_y on the contact zone:

$$\int_L\sigma(x)dx=-N. \tag{14}$$

Substitution of (12) into (14) and the use of standard tables ([18]) gives the more explicit form

$$\frac{c_{44}R}{A(b+a)}\int_L \frac{dV(t)}{dt}\sqrt{\frac{L_+-t}{t-L_-}}\,dt=N. \tag{15}$$

For sliding in the transonic ($v_r<v<\sqrt{\beta}v_r$) speed range, the result (*A*14*b*) and (9) give the Cauchy singular integral equation

$$\frac{A}{c_{44}C}\left[(c^2Ab'-aB'T)\sigma(x)+(c^2aA+b'B'T)\frac{1}{\pi}\int\!\!\!\!\!\!\diagup_L \frac{\sigma(t)dt}{t-x}\right] = \frac{dV(x)}{dx}\quad (x\in L) \tag{16}$$

where (*A*4), (*A*5), and (*A*15*b*) hold and, again, the cancellation of terms that occurs when $y=0$ has been exploited. Following [17], the solution to (16) is

$$\sigma(x)=\frac{c_{44}\sqrt{K}}{A\sqrt{a^2-b^2}}I\left(\frac{dV}{dx};x\right)\quad (x\in L) \tag{17}$$

where the operator I and eigenvalue υ are defined as

$$I(g;x)=g(x)\cos\pi\upsilon+\frac{1}{\pi}\left(\frac{L_+-x}{x-L_-}\right)^{\upsilon}\sin\pi\upsilon\int\!\!\!\!\!\!\diagup_L g(t)\times\left(\frac{t-L_-}{L_+-t}\right)^{\upsilon}\frac{dt}{t-x}\quad (x\in L) \tag{18a}$$

$$\upsilon=\frac{1}{\pi}\tan^{-1}\frac{b'}{a}\frac{\alpha a^2T-c^2A^2}{\alpha b^2T-c^2A^2}-\frac{1}{2}\quad\left(-\frac{1}{2}<\upsilon<0\right). \tag{18b}$$

The counterparts to the auxiliary conditions (13) and (15) are

$$\int_L \frac{dV(t)}{dt}\left(\frac{t-L_-}{L_+-t}\right)^{\upsilon}\frac{dt}{t-L_+}=0 \tag{19a}$$

$$\frac{c_{44}\sqrt{K}}{A\sqrt{a^2-b^2}}\int_L \frac{dV(t)}{dt}\left(\frac{t-L_-}{L_+-t}\right)^{\upsilon}dt=N. \tag{19b}$$

Finally, for sliding speeds in the supersonic ($v>\sqrt{\beta}v_r$) range, (9) and (A16b) lead directly to the results

$$\sigma(x)=\frac{c_{44}R'}{A'(b'+a')}\frac{dV(x)}{dx}\quad (x\in L) \tag{20a}$$

$$\frac{c_{44}R'}{A'(b'+a')}[V(L_+)-V(L_-)]=N \tag{20b}$$

where (A',a',b',R') are defined by (A15b) and (A17) and, in this case, boundedness at $x=L_\pm$ is controlled more directly by the form of $V(x)$. The translational part of any rigid-body motion cancels out in (20b).

Equations (12), (17), and (20*a*), along with their auxiliary conditions, constitute in light of (A13), (A14), and (A16), respectively, the solution candidates for the sliding indentation problem. The actual solutions are those that satisfy in addition the unilateral constraints (*a*) and (*b*).

Problem Solution: Subsonic Case

To illustrate the solution identification process for subsonic ($0<v<v_r$) sliding speeds, the generic parabolic indentor characterized by the function

$$V(x)=V_0+V_1x+\frac{1}{2}V_2x^2 \tag{21}$$

is considered, where the V_k are real constants. Substitution of (21) into (12), (13), and (15) gives, upon the use of standard tables ([18]),

$$\sigma(x)=-\frac{c_{44}V_2R}{A(a+b)}\sqrt{L_+-x}\sqrt{x-L_-}\quad (x\in L) \tag{22}$$

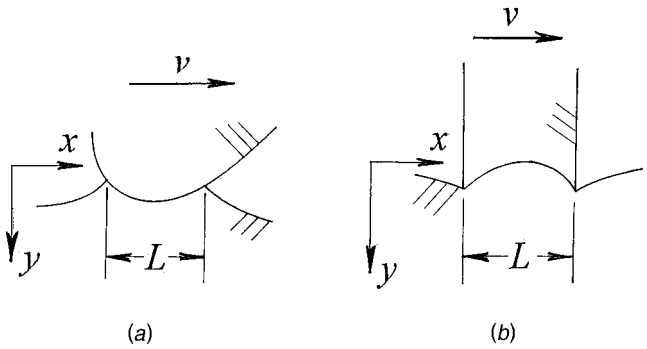


Fig. 2 (*a*) Surface deformation schematic for subcritical sliding; (*b*) surface deformation schematic for physically unacceptable sliding.

where $L_\pm$ can be obtained from the formulas

$$2V_1+V_2(L_++L_-)=0 \tag{23a}$$

$$\frac{\pi c_{44}V_2RL^2}{8A(a+b)}=N. \tag{23b}$$

In view of (10), unilateral constraint (*a*) requires that $\sigma(x)\leqslant 0$ which, in light of (22), means that $V_2R>0$. Study of (*A*7) and (*A*9) in the Appendix shows that $R<0 (0<c<c_R)$ and $R>0 (c_R<c<1)$, where c_Rv_r is the Rayleigh speed associated with the x-direction. The former case implies that $V_2<0$, i.e., the sliding indentor is concave upward (Fig. 1), while the latter requires that $V_2>0$, i.e., the indentor is concave downward. Although the former situation seems more plausible, imposition of unilateral constraint (*b*) provides closure.

It is noted that (22) is both bounded and continuous (vanishes) at the contact zone edges $x=L_\pm$. Study of (*A*13) shows that this, in turn, guarantees continuity of ($u_{x,x},u_{y,x}$) there. However, taking the derivative of (*A*13*b*) and letting $y=0$ gives, in view of (23*b*), an integration that can be performed with standard tables ([18]):

$$u_{y,xx}=V_2\left[1-\frac{1}{2}\left(\sqrt{\frac{x-L_+}{x-L_-}}+\sqrt{\frac{x-L_-}{x-L_+}}\right)\right]\quad (x\notin L). \tag{24}$$

This is approximately the half-space surface curvature outside the contact zone. For $V_2<0$, (24) behaves as ($-\infty,x\to L_\pm$), which suggests the schematic Fig. 2(*a*) for the deformed surface. For $V_2>0$, however, (24) behaves as ($\infty,x\to L_\pm$), which suggests interpenetration—and violation of constraint (*b*)—unless the artifice depicted in Fig. 2(*b*) is adopted. This essentially requires that the indentor dimensions be defined by the contact zone parameters $L_\pm$.

In summary, then, subsonic rapid sliding can exist only below the Rayleigh speed associated with the x-direction, i.e., $0<v<c_Rv_r$. This behavior is analogous to that for the isotropic case ([5]).

Problem Solution: Transonic Case

For transonic ($v_r<v<\sqrt{\beta}v_r$) sliding speeds, the form (21) is again used to illustrate the solution identification process. In this instance, (17)–(19) and standard tables ([18]) give the bounded and continuous formula

$$\sigma(x)=-\frac{c_{44}V_2\sqrt{K}}{A\sqrt{a^2-b^2}}(L_+-x)^{1+\upsilon}(x-L_-)^{-\upsilon}\quad (x\in L). \tag{25}$$

Clearly, constraint (*a*) can be satisfied only if $V_2>0$. Thus, it would appear that constraint (*b*) cannot be satisfied for transonic sliding unless the situation depicted in Fig. 2(*b*) is allowed.

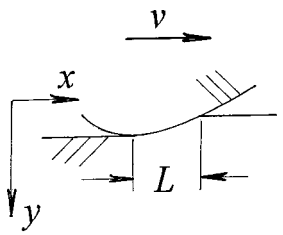


Fig. 3 Surface deformation schematic for supersonic sliding and for the single allowable transonic speed

However, (18*b*) shows that $v=0$ when $c=c^*(1<c^*<\sqrt{\beta})$, where

$$c^*=\sqrt{\frac{\alpha\beta-(m-1)^2}{\alpha+m-1}} \tag{26}$$

and (8) guarantees a real value. Then (17)–(19) reduce to the degenerate case

$$\sigma(x)=c_{44}\sqrt{\frac{\alpha}{\beta}}\frac{\alpha\beta-(m-1)^2}{\sqrt{m-1+\beta}\sqrt{m-1+\alpha}}\frac{dV(x)}{dx}\quad(x\in L). \tag{27}$$

For (21) it can be shown that dV/dx in (27) is replaced by $V_2(x-L_-)$, and both unilateral constraints can be satisfied for $V_2<0$. The auxiliary conditions

$$V_1+V_2L_-=0, \tag{28a}$$

$$-\frac{c_{44}V_2}{2}\sqrt{\frac{\alpha}{\beta}}\frac{\alpha\beta-(m-1)^2}{\sqrt{m-1+\beta}\sqrt{m-1+\alpha}}L^2=N \tag{28b}$$

define the parameters $L_\pm$. Unlike their subsonic counterparts (23*a*,*b*), these conditions do not guarantee continuity of $\sigma(x)$ at the leading edge $x=L_+$ of the contact zone. Indeed, a finite discontinuity occurs there. The surface deformation schematic in Fig. 3 shows that trailing edge $x=L_-$ is also the point of zero indentor slope, i.e., point of maximum normal displacement under the indentor.

This limit result is also consistent with isotropic work ([5]); there, a value that would correspond to $c^*=\sqrt{2}$ at which transonic sliding could occur was derived. In the isotropic limit $\alpha=\beta=1+m$, and (26) reduces to the same value.

Problem Solution: Supersonic Sliding

For supersonic ($v>\sqrt{\beta}v_r$) sliding speeds, combining (20) and (21) and then imposing both unilateral constraints gives

$$\sigma(x)=\frac{c_{44}R'V_2}{A'(a'+b')}(x-L_-)\quad(x\in L) \tag{29}$$

where $V_2<0$ and $L_\pm$ are determined from

$$V_1+V_2L_-=0, \tag{30a}$$

$$-\frac{c_{44}R'V_2L^2}{2A'(a'+b')}=N \tag{30b}$$

Equation (29) shows that, as in the transonic case $v=c^*v_r(1<c^*<\sqrt{\beta})$, the contact zone traction loses continuity at the leading edge. Again, Fig. 3 depicts the surface deformation, and (30) locates the trailing edge at the point of maximum normal surface displacement under the indentor.

Comments

Table 1 lists relevant properties for the five materials identified earlier as being representative of the orthotropic/transversely isotropic class defined by (8). It is seen that, despite differences in the parameters (α,β,m,c_{44}), the Rayleigh speeds are, as fractions c_R of the corresponding rotational speed, similar in value.

Table 1 Properties of orthotropic/transversely isotropic materials

	α	β	m	c_{44} (GPa)	c_R
beryl	3.62	4.11	2.01	68.6	0.956
cobalt	4.74	4.07	2.37	75.5	0.962
ice	4.57	4.26	2.64	3.17	0.959
magnesium	3.74	3.61	2.3	16.4	0.943
titanium	3.88	3.47	2.48	46.7	0.936

To illustrate the effects of sliding speed and material properties on the contact zone generated, consider the special case $V_1=0$ in (21): For subcritical $(0<c<c_R)$ sliding, (23) yields the expressions

$$L_\pm=\pm\frac{L}{2}, \tag{31a}$$

$$\sqrt{\frac{|V_2|}{N}}L=2\sqrt{\frac{2A(a+b)}{\pi c_{44}|R|}} \tag{31b}$$

for the contact zone edge locations $L_\pm$ with respect to the translating indentor, and the zone width L. Calculations of (31*b*) for various subcritical values of c are given in Table 2. The entries show that, for a given parabolic indentor profile and compressive load, the zone width decreases with sliding speed for all five materials at low speeds, but becomes unbounded as the critical (Rayleigh) speed is approached.

In summary, then, this article has presented exact solutions for steady dynamic sliding at any constant speed by a smooth rigid indentor on a class of orthotropic/transversely isotropic half-spaces. As in corresponding isotropic elastic analyses for linear isothermal ([5]), linear thermoelastic ([6]), and hyperelastic isothermal ([7]) materials, physically acceptable solutions have been found to exist for sliding speeds below the Rayleigh value. In the transonic range, only one allowable sliding speed exists, and contact zone traction is lost at the leading edge. Physically acceptable solutions exist for all supersonic sliding speeds but, again, the contact zone is discontinuous at the leading edge.

The solution procedure makes use of mixed boundary value problem solution techniques from classical elasticity ([17]), and variations of Cauchy singular integral equation solutions employed in [6,7,19]. In addition, convenient factorizations in integral transform space, after [16], allow a simplification of solution forms, despite the problem anisotropy. In particular, a compact exact formula for the Rayleigh speed associated with the material symmetry axis aligned with the surface emerges.

Solution results demonstrate that, for a given indentor profile and compression force, contact zone widths achieve minimum values at high but sub-Rayleigh values of the sliding speed. While insight into the problem of wear/high-speed mechanism operation is possible from such a result, the effects of friction and temperature have been neglected here. This work, therefore, along with results such as [6,7], is currently being used as the basis for studies that include both effects.

Table 2 Contact zone width parameter $\sqrt{\frac{|V_2|}{N}}L$(1/GPa$^{1/2}$) versus dimensionless crack speed

	$c=0.001$	$c=0.1$	$c=0.3$	$c=0.5$	$c=0.7$	$c=0.9$
beryl	0.2206	0.1716	0.16	0.1712	0.1989	0.3384
cobalt	0.1954	0.138	0.142	0.1518	0.175	0.286
ice	0.9817	0.6857	0.7084	0.755	0.8739	1.484
magnsium	0.4395	0.3199	0.3297	0.354	0.4155	0.785
titanium	0.2577	0.1895	0.1953	0.2104	0.2479	0.502

Appendix

Consider the bilateral Laplace transform ([20]) and its inverse

$$\hat{F}=\int_{-\infty}^{\infty}F(x)e^{-px}dx,\quad F(x)=\frac{1}{2\pi i}\int\hat{F}e^{px}dp \quad (A1a,b)$$

where p is imaginary, and integration in (A1*b*) is along a Bromwich contour in the p-plane. Application of (A1*a*) to (6) in view of (7), (10) and the boundedness/continuity conditions gives a coupled set of linear ordinary differential equations in y that can be solved to give

$$p\hat{u}_x=C_a e^{-ay\sqrt{p}\sqrt{-p}}+\frac{\psi}{mb}C_b\frac{\sqrt{-p}}{\sqrt{p}}e^{-by\sqrt{p}\sqrt{-p}} \quad (A2a)$$

$$p\hat{u}_y=\frac{\phi}{m\alpha a}C_a\frac{\sqrt{-p}}{\sqrt{p}}e^{-ay\sqrt{p}\sqrt{-p}}+C_b e^{-by\sqrt{p}\sqrt{-p}} \quad (A2b)$$

where the dimensionless coefficients are given by the formulas

$$C_a=\frac{mB^2b}{D}Q\frac{\hat{\sigma}}{c_{44}},\quad C_b=\frac{mab}{D}P\frac{\sqrt{-p}}{\sqrt{p}}\frac{\hat{\sigma}}{c_{44}} \quad (A3a)$$

$$D=aP^2-bB^2Q^2 \quad (A3b)$$

in view of the definitions

$$P=\psi+mB^2,\quad Q=m+\psi,\quad \phi=\alpha a^2-A^2,\quad \psi=\alpha b^2-B^2 \quad (A4a)$$

$$\sqrt{\alpha}(a,b)=\frac{1}{\sqrt{2}}\sqrt{T\pm\sqrt{T^2-4A^2B^2}}=\frac{1}{2}(\sqrt{T+2AB}\pm\sqrt{T-2AB}) \quad (A4b)$$

$$\sqrt{\alpha}(a+b)=\sqrt{(A+B)^2-m^2} \quad (A4c)$$

$$T=A^2+B^2-m^2,\quad A=\sqrt{\alpha}\sqrt{\beta-c^2},\quad B=\sqrt{1-c^2}. \quad (A4d)$$

Introduction of the branch cuts $\mathrm{Im}(p)=0$, $\mathrm{Re}(p)<0$ and $\mathrm{Im}(p)=0$, $\mathrm{Re}(p)>0$ for $(\sqrt{p},\sqrt{-p})$, respectively, guarantees that their real parts are non-negative in their cut planes. Therefore, for (a, b) real and positive, boundedness of (A2) for $y>0$ is assured.

Introduction of the branch cuts $\mathrm{Im}(c)=0$, $|\mathrm{Re}(c)|>\sqrt{\beta}$ and $\mathrm{Im}(c)=0$, $|\mathrm{Re}(c)|>1$ for the quantities (A, B), respectively, guarantees, similarly, that they have non-negative real parts in a cut complex c-plane. The quantities (a, b) share the branch cuts of (A, B), respectively, but the restrictions (8) guarantee that (a, b) also exhibit branch points defined by

$$(a-1)^2c^2=\gamma(1+\alpha)-2\alpha(1+\beta)\pm i2m\sqrt{\alpha}\sqrt{\gamma-\alpha-\beta}. \quad (A5)$$

These also define roots of b^2-a^2 and, indeed, the denominator term D in (A3*b*) vanishes when $b=a$. Even though c takes on only positive real values here, this apparently singular behavior does not in any event effect $(\hat{u}_x,\hat{u}_y)$: When $b=a$ the exponential terms in (A2) are identical and, by using (A4) and the related formulas

$$\alpha a^2\psi=-B^2\phi,\quad \alpha ab=AB,\quad \phi+\psi+m^2=0 \quad (A6)$$

it can be shown that the resulting numerator terms in (A2) also vanish when $b=a$. This behavior can be made more explicit, and the expressions (A2) simplified, by allowing $a+b$ to be continuous across branch cuts associated with (A5), even though $(a,b,b-a)$ remain multivalued there. Then, (A4) and (A6) can be used to show that

$$D=AB\psi(b-a)R,\quad R=c^2A+BC,\quad C=(m-1)^2-A^2. \quad (A7)$$

The quantity R has only the branch cuts of (A, B) along the $\mathrm{Re}(c)$-axis. In view of (A7), (A2) can be written as

$$p\hat{u}_x=\frac{a\hat{\sigma}}{c_{44}}\left[\frac{mAQ}{\phi(b-a)R}e^{-ay\sqrt{p}\sqrt{-p}}-\frac{P}{B(b-a)R}e^{-by\sqrt{p}\sqrt{-p}}\right] \quad (A8a)$$

$$p\hat{u}_y=\frac{A\hat{\sigma}}{\alpha c_{44}}\frac{\sqrt{-p}}{\sqrt{p}}\left[\frac{Q}{(b-a)R}e^{-ay\sqrt{p}\sqrt{-p}}+\frac{mP}{\psi(b-a)R}e^{-by\sqrt{p}\sqrt{-p}}\right]. \quad (A8b)$$

In light of previous remarks, the coefficients of the exponential terms in (A8) exhibit the branch cuts of (A, B), and are not singular when $b=a$. Indeed, because the exponential terms also become identical (unity) when $y=0$, evaluation of (A8) on the half-space surface leads to an explicit cancellation of terms proportional to $b-a$.

In particular, R is analytic in the c-plane cut along $\mathrm{Im}(c)=0$, $|\mathrm{Re}(c)|>1$, and exhibits the real roots $c=\pm c_R(0<c_R<1)$. Rationalization of the relation $R=0$ gives a cubic equation in c^2 that corresponds exactly to one obtained for the roots of a transversely isotropic Rayleigh function ([15]). That, is, $c_R v_r$ is the Rayleigh speed parallel to the x-axis of material symmetry for the class of solids treated here, and R is itself the essential factor of the Rayleigh function. The form of R is, however, simpler in form than standard functions, whether for isotropic ([10]) or anisotropic ([15]) materials. As an alternative to the cubic equation solution, a general approach ([21]) yields c_R in a form that is analytical to within a simple quadrature:

$$c_R=\sqrt{\frac{\alpha\beta-(m-1)^2}{\sqrt{\alpha\beta}(1+\sqrt{\alpha})}}G_R,$$

$$G_R=\frac{1}{\pi}\int_1^{\sqrt{\beta}}\tan^{-1}\frac{C\sqrt{t^2-1}}{t^2\sqrt{\alpha}\sqrt{\beta-t^2}}\frac{dt}{t}. \quad (A9)$$

Here (8) guarantees that the coefficient of G_R is real and positive. A similar approach ([22]) has also yielded a Rayleigh speed, but two more complicated quadratures arise due to use of a Rayleigh function without factorization.

In the isotropic limit, the branch points (A5) collapse into the origin as isolated real roots, so that a factorization process based on (A7) does not have the advantage of extracting complex branch points. However, the simplification indicated by (A8) could still prove convenient.

In light of (A1*a*),

$$\hat{\sigma}=\int_L\sigma(t)e^{-pt}dt \quad (A10)$$

where the symbol L affixed to the integral operator signifies that integration is over the real interval (L_-,L_+). The left-hand sides of (A8) are the transforms of first derivatives with respect to x which, in view of the rigid-body motion superposition noted in the body of this article, are sufficient for purposes of a problem solution. When $0<v<v_r$, (4) gives $0<c<1$. This ensures that (A, a, B, b) are all real and positive. Thus, $(v_r,\sqrt{\beta}v_r)$ are the rotational and dilatational speeds associated with the material symmetry axis aligned with the half-space surface, and $0<c<1$ corresponds to subsonic sliding. Like (A2), (A8) is bounded for (A, a, B, b) real and positive, so that substitution into (A1*b*) along with (A10) and interchanging the order of the (t, p)-integrations reduces the inversion process to the generic integrals

$$\frac{1}{2\pi i}\int\left(1,\frac{\sqrt{-p}}{\sqrt{p}}\right)e^{p(x-t)-k\sqrt{p}\sqrt{-p}}dp\quad (k=ay,by). \quad (A11)$$

The integrands are both analytic for $\mathrm{Re}(p)=0$, so that the Bromwich contour can be the entire $\mathrm{Im}(p)$-axis. Performing the integration yields, with the help of standard tables ([18]),

$$\frac{1}{\pi}\frac{(k,x-t)}{(x-t)^2+k^2}. \quad (A12)$$

The results

$$c_{44}u_{x,x}=\frac{Ay}{\pi\alpha(b-a)R}\int_L\sigma(t)\left[\frac{Q}{\psi}\frac{mB^2}{(t-x)^2+a^2y^2}-\frac{P}{(t-x)^2+b^2y^2}\right]dt \tag{A13a}$$

$$c_{44}u_{y,x}=\frac{A}{\pi\alpha(b-a)R}\int_L\sigma(t)(t-x)\left[\frac{Q}{(t-x)^2+a^2y^2}+\frac{P}{\psi}\frac{m}{(t-x)^2+b^2y^2}\right]dt \tag{A13b}$$

follow for $y>0$. As implied above, although the situation $b=a$ does not arise for c real, the expressions (*A*13) are finite in any case, and the factor b-a is cancelled from (*A*13) when $y=0$.

For $1<c<\sqrt{\beta}$ (transonic sliding) the parameters (A, a) remain positive real, but (B, b) are now imaginary. Therefore, the process used for (*A*13) must now consider the behavior in the complex c-plane noted above in order that (*A*8) remains bounded for $y>0$. Application of (*A*1*b*) then yields

$$c_{44}u_{x,x}=\frac{b'B'Q}{\pi\psi}\int_L\left[\frac{c^2\alpha a^2-B^2C}{\alpha(a^2-b^2)K}Ay+\frac{B'}{a}\frac{\alpha a^2C-c^2A^2}{\alpha(a^2-b^2)K}(t-x)\right]\frac{m\sigma(t)dt}{(t-x)^2+a^2y^2}+P\left[\frac{c^2\alpha a^2-B^2C}{\alpha(a^2-b^2)K}\frac{A}{\pi B'}\int_L\frac{\sigma(t)dt}{t-x-b'y}+\frac{c^2A^2-\alpha a^2C}{\alpha(a^2-b^2)K}\sigma(x+b'y)\right] \tag{A14a}$$

$$c_{44}u_{y,x}=\frac{b'Q}{\pi}\int_L\left[\frac{\alpha a^2C-c^2A^2}{\alpha(a^2-b^2)K}ay+AB'\frac{B^2C-c^2\alpha a^2}{\alpha(a^2-b^2)K}(t-x)\right]\frac{\sigma(t)dt}{(t-x)^2+a^2y^2}+\frac{mP}{\alpha a\psi}\left[\frac{c^2\alpha a^2-B^2C}{\alpha(a^2-b^2)K}\frac{A}{\pi}\int_L\frac{\sigma(t)dt}{t-x-b'y}+\frac{c^2A^2-\alpha a^2C}{\alpha(a^2-b^2)K}B'\sigma(x-b'y)\right] \tag{A14b}$$

for $y>0$. Here (*A*4) is augmented by the definitions

$$K=c^4A^2-B^2C^2 \tag{A15a}$$

$$B'=\sqrt{c^2-1},\quad \sqrt{2\alpha}b'=\sqrt{\sqrt{T^2-4A^2B^2}-T}. \tag{A15b}$$

It is understood that the nonintegral terms in (*A*14) appear only when $x+b'y\in L$, and the integrals that are grouped with these terms must then be viewed in the Cauchy principal value sense. As in the case of (*A*13), a cancellation of terms occurs in (*A*14) when $y=0$.

For $c>\sqrt{\beta}$ (supersonic sliding), (A,aB,b) are now all imaginary, and application of (*A*1*b*) to (*A*8) gives

$$c_{44}u_{x,x}=\frac{-mB'Q}{(b'-a')\psi R'}\sigma(x+a'y)-\frac{a'P}{B'(b'-a')R'}\sigma(x+b'y) \tag{A16a}$$

$$c_{44}u_{y,x}=\frac{A'Q}{\alpha(b'-a')R'}\sigma(x+a'y)+\frac{mA'P}{\alpha(b'-a')\psi R'}\sigma(x+b'y) \tag{A16b}$$

for $y>0$. Here the first and second terms, respectively, in both (*A*16*a*,*b*) vanish unless $x+a'y\in L$ and $x+b'y\in L$. This behavior indicates that the half-space surface is undisturbed unless $x\in L$. Moreover, depending on the form of σ, these terms, and the nonintegral terms in (*A*15), could exhibit lines of discontinuity $x+a'y=L_\pm$ or $x+b'y=L_\pm$ that radiate from the moving contact zone edges. In (*A*16), the additional definitions

$$A'=\sqrt{\alpha}\sqrt{c^2-\beta},\quad \sqrt{2\alpha}a'=\sqrt{-T-\sqrt{T^2-4A^2B^2}},\quad R'=c^2A'+B'C \tag{A17}$$

arise.

It can be shown that (*A*13) and (*A*14) are continuous at $c=1$, and that (*A*14) and (*A*16) are continuous when $c=\sqrt{\beta}$. A key step in the demonstration is the use of the standard ([23]) result

$$\frac{k}{(t-x)^2+k^2}\to\pi\delta(t-x)\quad(k\to0+) \tag{A18}$$

where δ is the Dirac function.

References

[1] Craggs, J. W., and Roberts, A. M., 1967, "On the Motion of a Heavy Cylinder Over the Surface of an Elastic Half-Space," ASME J. Appl. Mech., **24**, pp. 207–209.

[2] Gerstle, F. B., and Pearsall, G. W., 1974, "The Stress Response of an Elastic Surface to a High-Velocity, Unlubricated Punch," ASME J. Appl. Mech., **41**, pp. 1036–1040.

[3] Brock, L. M., 1981, "Sliding and Indentation by a Rigid Half-Wedge With Friction and Displacement Coupling," Int. J. Eng. Sci., **19**, pp. 33–40.

[4] Brock, L. M., 1996, "Some Analytical Results for Heating due to Irregular Sliding Contact," Indian J. Pure Appl. Math., **27**, pp. 1257–1278.

[5] Georgiadis, H. G., and Barber, J. R., 1993, "On the Super-Rayleigh/Subseismic Elastodynamic Indentation Problem," J. Elast., **31**, pp. 141–161.

[6] Brock, L. M., and Georgiadis, H. G., 2000, "Sliding Contact With Friction of a Thermoelastic Solid at Subsonic, Transonic and Supersonic Speeds," J. Therm. Stresses, **23**, pp. 629–656.

[7] Brock, L. M., 1999, "Sliding Contact With Friction at Arbitrary Constant Speeds on a Pre-stressed Highly Elastic Half-Space," J. Elast., **57**, pp. 105–132.

[8] Green, A. E., and Zerna, W., 1968, *Theoretical Elasticity*, 2nd Ed., Oxford University Press, London.

[9] Beatty, M. F., and Usmani, S. A., 1975, "On the Indentation of a Highly Elastic Half-Space," Q. J. Mech. Appl. Math., **28**, pp. 47–62.

[10] Achenbach, J. D., 1973, *Wave Propagation in Elastic Solids*, North-Holland, Amsterdam.

[11] Sokolnikoff, I. S., 1983, *Mathematical Theory of Elasticity*, Krieger, Malabar, FL.

[12] Scott, R. A., and Miklowitz, J., 1967, "Transient Elastic Waves in Anisotropic Plates," ASME J. Appl. Mech., **34**, pp. 104–110.

[13] Nye, J. F., 1957, *Physical Properties of Crystals, Their Representation by Tensors and Matrices*, Clarendon Press, Oxford.

[14] Theocaris, P. S., and Sokolis, D. P., 2000, "Invariant Elastic Constants and Eigentensors of Orthorhombic, Tetragonal, Hexagonal and Cubic Crystalline Media," Acta Crystallogr., **A56**, pp. 319–331.

[15] Payton, R. G., 1983, *Elastic Wave Propagation in Transversely Isotropic Media*, Martinus Nijhoff, The Hague.

[16] Brock, L. M., Georgiadis, H. G., and Hanson, M. T., 2001, "Rapid Indentation of Transversely Isotropic or Orthotropic Half-Spaces," ASME J. Appl. Mech., **68**, pp. 490–495.

[17] Erdogan, F., 1976, "Mixed Boundary Value Problems in Mechanics," Mechanics Today, Vol. 4 S. Nemat-Nasser, ed., Pergamon Press, New York.

[18] Gradshteyn, I. S., and Ryzhik, I. M., 1980, *Table of Integrals, Series and Products*, Academic Press, San Diego, CA.

[19] Brock, L. M., 1999, "Rapid Sliding Indentation With Friction of a Pre-stressed Thermoelastic Material," J. Elast., **53**, pp. 161–188.

[20] van der Pol, B., and Bremmer, H., 1950, *Operational Calculus Based on the Two-Sided Laplace Integral*, Cambridge University Press, London.

[21] Brock, L. M., 1998, "Analytical Results for Roots of Two irrational Functions in Elastic Wave Propagation," J. Aust. Math. Soc. B, Appl. Math., **B40**, pp. 72–79.

[22] Norris, A. N., and Achenbach, J. D., 1984, "Elastic Wave Diffraction by a Semi-infinite Crack in a Transversely Isotropic Material," Q. J. Mech. Appl. Math., **37**, pp. 565–580.

[23] Carrier, G. F., and Pearson, C. E., 1988, *Partial Differential Equations*, 2nd Ed., Academic Press, San Diego, CA.

Moment Lyapunov Exponents of a Two-Dimensional Viscoelastic System Under Bounded Noise Excitation

W.-C. Xie
Associate Professor,
Solid Mechanics Division,
Faculty of Engineering,
University of Waterloo,
Waterloo, ON N2L 3G1, Canada

The moment Lyapunov exponents of a two-dimensional viscoelastic system under bounded noise excitation are studied in this paper. An example of this system is the transverse vibration of a viscoelastic column under the excitation of stochastic axial compressive load. The stochastic parametric excitation is modeled as a bounded noise process, which is a realistic model of stochastic fluctuation in engineering applications. The moment Lyapunov exponent of the system is given by the eigenvalue of an eigenvalue problem. The method of regular perturbation is applied to obtain weak noise expansions of the moment Lyapunov exponent, Lyapunov exponent, and stability index in terms of the small fluctuation parameter. The results obtained are compared with those for which the effect of viscoelasticity is not considered. [DOI: 10.1115/1.1445143]

1 Introduction

The study of the dynamics of many engineering structures under random loadings, such as those arising from earthquakes, wind, and ocean waves, which can be described satisfactorily only in probabilistic terms, leads to a dynamic system of the form

$$\ddot{\mathbf{x}}(t)=\mathbf{f}(\mathbf{x}(t),\dot{\mathbf{x}}(t),\boldsymbol{\xi}(t)), \tag{1}$$

where $\boldsymbol{\xi}(t)$ denotes a vector of stochastic processes characterizing the randomness of the loadings. It is of practical importance to investigate the dynamical stability of system (1) under the stochastic excitations $\boldsymbol{\xi}(t)$. For engineering applications, the stochastic excitation has been modeled as a Gaussian white noise process, a real noise process, or a bounded noise process.

A white noise process is a weakly stationary process that is delta-correlated and mean zero. Its power spectral density is constant over the entire frequency range, which is obviously an idealization. A white noise process is frequently adopted as a model for noise because of the availability of mathematical theory, such as Itô calculus, in dealing with white noise processes.

A real noise $\xi(t)$ is often characterized by an Ornstein-Uhlenbeck process and is given by

$$d\xi(t)=-\alpha_0\xi(t)dt+\sigma_0\circ dW(t), \tag{2}$$

where $W(t)$ is a standard Wiener process. It is well known that $\xi(t)$ is a normally distributed random variable, which is not bounded and may take arbitrarily large values with small probabilities, and hence may not be a realistic model of noise in many engineering application.

A bounded noise $\eta(t)$ is a more realistic and versatile model of stochastic fluctuation in engineering applications and is normally represented as

$$\eta(t)=\cos[\nu_0 t+\sigma_0 W(t)+\theta], \tag{3}$$

in which θ is a uniformly distributed random number in $(0,2\pi)$. The inclusion of the phase angle θ in Eq. (3) makes $\eta(t)$ a stationary process.

Equation (3) may be written as

$$\begin{aligned}\eta(t)&=\cos Z(t),\\ dZ(t)&=\nu_0 dt+\sigma_0\circ dW(t),\end{aligned} \tag{4}$$

where the initial condition of $Z(t)$ is $Z(0)=\theta$. The correlation function of $\eta(t)$ is given by

$$E[\eta(t+\tau)\eta(t)]=R(\tau)=\frac{1}{2}\cos\nu_0\tau\exp\left(-\frac{\sigma_0^2}{2}|\tau|\right),$$

and the spectral density function of $\eta(t)$ is

$$\begin{aligned}S(\omega)&=\int_{-\infty}^{+\infty}R(\tau)e^{i\omega\tau}d\tau\\ &=\sigma_0^2\left[\frac{1}{4(\omega-\nu_0)^2+\sigma_0^4}+\frac{1}{4(\omega+\nu_0)^2+\sigma_0^4}\right].\end{aligned}$$

It may be noted that the mean-square value of the bounded noise process $\eta(t)$ is fixed at $E[\eta^2(t)]=1/2$. The spectral density function can be made to approximate the well-known Dryden and von Karman spectra of wind turbulence by suitable choice of the parameters ν_0, σ_0, and ε_0. In the limit as σ_0 approaches infinite, the bounded noise becomes a white noise of constant spectral density. However, since the mean-square value is fixed at 1/2, this constant spectral density level reduces to zero in the limit. On the other hand, in the limit as σ_0 approaches zero, the bounded noise becomes a deterministic sinusoidal function.

The bounded noise process (4) was first employed by Stratonovich [1] and has since been applied in certain engineering applications by Dimentberg [2], Wedig [3], Lin and Cai [4], and Ariaratnam [5].

The sample or almost-sure stability of the trivial solution of system (1) is determined by the Lyapunov exponent, which characterizes the average exponential rate of growth of the solutions of system (1) for t large, defined as

$$\lambda_{\mathbf{x}(t)}=\lim_{t\to\infty}\frac{1}{t}\log\|\mathbf{x}(t)\|, \tag{5}$$

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Apr. 24, 2001; final revision, Oct. 2, 2001. Associate Editor: D. A. Siginer. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

where $\|\mathbf{x}(t)\|$ denotes the Euclidean vector norm. The trivial solution of system (1) is stable with probability one (w.p.1) if the top Lyapunov exponent is negative, whereas it is unstable w.p.1 if the top Lyapunov exponent is positive.

On the other hand, the stability of the pth moment of the trivial solution of system (1), $E[\|\mathbf{x}(t)\|^p]$, is determined by the moment Lyapunov exponent

$$\Lambda_{\mathbf{x}(t)}(p)=\lim_{t\to\infty}\frac{1}{t}\log E[\|\mathbf{x}(t)\|^p], \tag{6}$$

where $E[\cdot]$ denotes expected value. If $\Lambda_{\mathbf{x}(t)}(p)<0$, then $E[\|\mathbf{x}(t)\|^p]\to 0$ as $t\to\infty$. The pth moment Lyapunov exponent $\Lambda_{\mathbf{x}(t)}(p)$ is a convex analytic function in p with $\Lambda_{\mathbf{x}(t)}(0)=0$ and $\Lambda'_{\mathbf{x}(t)}(0)$ is equal to the the top Lyapunov exponent $\lambda_{\mathbf{x}(t)}$. The non trivial zero $\delta_{\mathbf{x}(t)}$ of $\Lambda_{\mathbf{x}(t)}(p)$, i.e. $\Lambda_{\mathbf{x}(t)}(\delta_{\mathbf{x}(t)})=0$, is called the stability index.

However, suppose the top Lyapunov exponent $\lambda_{\mathbf{x}(t)}$ is negative, implying that system (1) is sample stable, the pth moment typically grows exponentially for large enough p, implying that the pth moment of system (1) is unstable. According to the theory of large deviation, although the solution of the system $\|\mathbf{x}(t)\|\to 0$ as $t\to\infty$ w.p.1 at an exponential rate $\lambda_{\mathbf{x}(t)}$, there is a small probability that $\|\mathbf{x}(t)\|$ is large, which makes the expected value $E[\|\mathbf{x}(t)\|^p]$ of this rare event large for large enough values of p, leading to pth moment instability.

Hence, to have a complete picture of the dynamical stability of system (1), it is important to study both the sample and moment stability and to determine both the top Lyapunov exponent and the pth moment Lyapunov exponent.

A systematic study of moment Lyapunov exponents is presented in reference [6] for linear Itô systems and in reference [7] for linear stochastic systems under real noise excitations. A systematic presentation of the theory of random dynamical systems and a comprehensive list of references are presented in [8].

Although the moment Lyapunov exponents are important in the study of dynamic stability of randomly perturbed systems, the actual evaluations of the moment Lyapunov exponents are very difficult. Only a few results on the moment Lyapunov exponents have been published. Using the analytic property of the moment Lyapunov exponents, Arnold et al. [9] obtained weak noise expansions of the moment Lyapunov exponents of a two-dimensional system in terms of $\varepsilon\, p$, where ε is a small parameter, under both white noise and real noise excitations. Khasminskii and Moshchuk [10] obtained an asymptotic expansion of the moment Lyapunov exponent of a two-dimensional system under white noise parametric excitation in terms of the small fluctuation parameter ε, from which the stability index was obtained. Sri Namachchivaya and Vedula [11] obtained general asymptotic approximation for the moment Lyapunov exponent and the Lyapunov exponent for a four-dimensional system with one critical mode and another asymptotically stable mode driven by a small-intensity stochastic process.

In recent studies ([12] and [13]), Xie applied a procedure similar to that employed in Khasminskii and Moshchuk [10] to obtain weak noise expansions of the moment Lyapunov exponent, the Lyapunov exponent, and the stability index of a two-dimensional system exhibiting pitch-fork bifurcation under both real noise and bounded noise excitations in terms of the small fluctuation parameter.

There are many viscoelastic systems in engineering applications. The dynamic stability of viscoelastic systems has been investigated by many authors. Ariaratnam [14] studied the almost-sure stability of a single-degree-of-freedom linear viscoelastic system subjected to random fluctuation in the stiffness parameter by evaluating the top Lyapunov exponent and the rotation number using the method of stochastic averaging for integro-differential equations due to Larinov [15]. Ariaratnam [5] determined the top Lyapunov exponent of a linear viscoelastic system parametrically forced by a bounded noise excitation by the use of the averaging method for integro-differential equation ([15]). Potapov [16] studied the almost-sure stability of a viscoelastic column under the excitation of a random wide-band stationary process using Lyapunov's direct method. Potapov [17] considered the numerical evaluation of Lyapunov exponents of linear integro-differential equations, describing the behavior of stochastic viscoelastic systems.

However, there are no publications available on the evaluation of moment Lyapunov exponents of viscoelastic systems. This paper is the first to deal with the determination of small noise expansion of the moment Lyapunov exponent of a two-dimensional viscoelastic system under bounded noise excitation, which is a realistic model of noise in engineering applications.

2 Formulation

2.1 Equation of Motion. Consider the transverse vibration of a viscoelastic column under the excitation of a stochastic axial compressive load $P(\tau)$. The equation of motion is the partial differential equation

$$m\frac{\partial^2 v}{\partial\tau^2}+c\frac{\partial v}{\partial\tau}+EI(1-\mathcal{R})\frac{\partial^4 v}{\partial x^4}+P(\tau)\frac{\partial^2 v}{\partial x^2}=0, \tag{7}$$

where $v(x,\tau)$ is the transverse deflection of the column, x the axial coordinate, τ the time parameter, m the mass per unit length of the column, c the damping constant, EI the flexural rigidity of the column, $\mathcal{R}$ the material relaxation operator given by

$$\mathcal{R}v=\int_0^\tau R(\tau-s)v(s)ds,\quad R(\tau)=\varepsilon\gamma_0 e^{-\varepsilon\kappa_0\tau}, \tag{8}$$

in which γ_0 is a constant characterizing a measure of the material relaxation, κ_0 a constant representing the material relaxation rate, $\varepsilon>0$ a small parameter. If the column is simply supported, the transverse deflection may be taken as

$$v(x,\tau)=q(\tau)\sin\frac{\tau x}{L}, \tag{9}$$

where L is the length of the column. Substituting Eq. (9) into (7) yields the equation of motion for $q(\tau)$

$$\frac{d^2q(\tau)}{d\tau^2}+2\varepsilon^2c_0\frac{dq(\tau)}{d\tau}+\omega_0^2[1-\mathcal{R}-p(\tau)]q(\tau)=0,$$

or

$$\frac{d^2q(\tau)}{d\tau^2}+2\varepsilon^2c_0\frac{dq(\tau)}{d\tau}+\omega_0^2[1-p_0-\varepsilon\mu_0\xi(\tau)]q(\tau)-\varepsilon\gamma_0\omega_0^2\int_0^\tau e^{-\varepsilon\kappa_0(\tau-s)}q(s)ds=0, \tag{10}$$

where

$$2\varepsilon^2c_0=\frac{c}{m},\quad \omega_0^2=\frac{EI}{m}\left(\frac{\pi}{L}\right)^2,$$

$$p(\tau)=\frac{1}{EI}\left(\frac{L}{\pi}\right)^2P(\tau)=p_0+\varepsilon\mu_0\xi(\tau),$$

in which $\xi(\tau)$ is a stochastic process of mean zero. In this study, $\xi(\tau)=\cos z(\tau)$ is a bounded noise process, in which

$$dz(\tau)=\nu_0 d\tau+\sigma_0\circ dW(\tau), \tag{11}$$

or

$$z(\tau)=\nu_0\tau+\sigma_0 W(\tau)+\theta, \tag{12}$$

where $W(\tau)$ is a standard Wiener process in time τ, and θ is a random phase angle required to make the bounded noise $\xi(\tau)$ stationary.

Equation (10) can be simplified by removing the damping term. Letting $q(\tau)=x(\tau)e^{-\varepsilon^2 c_0\tau}$ and substituting into Eq. (10) results in

$$\frac{d^2x(\tau)}{d\tau^2}+\omega^2[1-\varepsilon\mu\cos z(\tau)]x(\tau)-\varepsilon\tilde{\gamma}_0\int_0^\tau e^{-\varepsilon\tilde{\kappa}_0(\tau-s)}x(s)ds=0, \tag{13}$$

where

$$\omega^2=\omega_0^2\left(1-p_0-\frac{\varepsilon^4 c_0^2}{\omega_0^2}\right),\quad \mu=\frac{\mu_0}{\omega^2},\quad \tilde{\kappa}_0=\kappa_0-\varepsilon^2 c_0,$$

$$\tilde{\gamma}_0=\gamma_0\omega_0^2.$$

By applying the time scaling $t=\omega\tau$, Eq. (13) can be further simplified to yield

$$\frac{d^2x(t)}{dt^2}+[1-\varepsilon\mu\cos\zeta(t)]x(t)-\varepsilon\gamma\int_0^t e^{-\varepsilon\kappa(t-s)}x(s)ds=0, \tag{14}$$

where $\zeta(t)$ is a random process given by

$$d\zeta(t)=\nu dt+\sigma\circ dW(t), \tag{15}$$

in which $W(t)$ is a standard Wiener process in time t, and

$$\gamma=\frac{\gamma_0\omega_0^2}{\omega^3},\quad \kappa=\frac{\kappa_0-\varepsilon^2 c_0}{\omega},\quad \nu=\frac{\nu_0}{\omega},\quad \sigma=\frac{\sigma_0}{\sqrt{\omega}}.$$

From the definitions of the Lyapunov exponents and the moment Lyapunov exponents, it can be easily shown that the Lyapunov exponents and moment Lyapunov exponents of systems (10), (13), and (14) are related as follows:

$$\lambda_{q(\tau)}=-\varepsilon^2 c_0+\lambda_{x(\tau)}=-\varepsilon^2 c_0+\omega\lambda_{x(t)},$$
$$\Lambda_{q(\tau)}(p)=-\varepsilon^2 c_0 p+\Lambda_{x(\tau)}(p)=-\varepsilon^2 c_0 p+\omega\Lambda_{x(t)}(p). \tag{16}$$

Without loss of generality, the moment Lyapunov exponent of system (14) is studied in the remaining of this paper.

2.2 Eigenvalue Problem for the Moment Lyapunov Exponent. The approach employed in this section was first applied by Wedig [18] to derive the eigenvalue problem for the moment Lyapunov exponent of a two-dimensional linear Itô stochastic system.

Letting

$$x_3(t)=\int_0^t e^{-\varepsilon\kappa(t-s)}x(s)ds, \tag{17}$$

the two-dimensional viscoelastic system (14) and (15) may be considered as a four-dimensional system as follows:

$$d\begin{Bmatrix}x_1\\x_2\\x_3\\\zeta\end{Bmatrix}=\begin{Bmatrix}x_2\\[-1+\varepsilon\mu\cos\zeta(t)]x_1+\varepsilon\gamma x_3\\x_1-\varepsilon\kappa x_3\\\nu\end{Bmatrix}dt+\begin{Bmatrix}0\\0\\0\\\sigma\end{Bmatrix}dW. \tag{18}$$

Apply the transformation

$$x_1=a\cos\varphi,\quad x_2=a\sin\varphi,\quad x_3=ab,\quad x_4=\zeta,$$
$$a=\sqrt{x_1^2+x_2^2},\quad P=a^p=(x_1^2+x_2^2)^{p/2},\quad \varphi=\tan^{-1}\left(\frac{x_2}{x_1}\right), \tag{19}$$

which is an extension of the Khasminskii transformation [19]. The Itô equations for P, φ, and b can be obtained using Itô's lemma

$$d\begin{Bmatrix}P\\\varphi\\b\\\zeta\end{Bmatrix}=\begin{Bmatrix}\varepsilon pP\sin\varphi(\mu\cos\zeta\cos\varphi+\gamma b)\\-1+\varepsilon\cos\varphi(\mu\cos\zeta\cos\varphi+\gamma b)\\\cos\varphi-\varepsilon b(\kappa+\mu\cos\zeta\sin\varphi\cos\varphi+\gamma b\sin\varphi)\\\nu\end{Bmatrix}dt+\begin{Bmatrix}0\\0\\0\\\sigma\end{Bmatrix}dW. \tag{20}$$

For small values of ε, $da/dt=O(\varepsilon)$ and $d\varphi/dt=-1+O(\varepsilon)$, and hence $a(t)=a_0+\varepsilon a_1(t)+\ldots$, $\varphi(t)=-t+\varepsilon\varphi_1(t)+\ldots$. From Eq. (17),

$$\begin{aligned}x_3(t)&=\int_0^t e^{-\varepsilon\kappa(t-s)}a(s)\cos\varphi(s)ds\\&=\int_0^t e^{-\varepsilon\kappa(t-s)}[a_0\cos s+O(\varepsilon)]ds\\&=\frac{a_0\sin t}{1+\varepsilon^2\kappa^2}+\frac{\varepsilon\kappa a_0(\cos t-e^{-\varepsilon\kappa t})}{1+\varepsilon^2\kappa^2}+O(\varepsilon)\\&=a_0\sin t+O(\varepsilon),\end{aligned}$$

and

$$b(t)=\frac{x_3(t)}{a(t)}=\frac{a_0\sin t+O(\varepsilon)}{a_0+O(\varepsilon)}=\sin t+O(\varepsilon).$$

Hence, for small values of ε, the range of $b(t)$ may be taken as $-1\leq b(t)\leq 1$.

Applying a linear stochastic transformation

$$S=T(\varphi,b,\zeta)P,\quad P=T^{-1}(\varphi,b,\zeta)S,$$
$$0\leq\varphi<2\pi,\quad -1\leq b\leq 1,\quad -\infty<\zeta<\infty,$$

the Itô equation for the new pth norm process S is given by, from Itô's lemma,

$$\begin{aligned}dS=\Big\{&\frac{1}{2}\sigma^2 T_{\zeta\zeta}+\nu T_\zeta-T_\varphi+\cos\varphi T_b+\varepsilon[(\mu\cos\zeta\cos\varphi+\gamma b)\\&\times(p\sin\varphi T+\cos\varphi T_\varphi)-b(\kappa+\mu\cos\zeta\sin\varphi\cos\varphi\\&+\gamma b\sin\varphi)T_b]\Big\}Pdt+\sigma T_\zeta P dW.\end{aligned} \tag{21}$$

For bounded and nonsingular transformation $T(\varphi,b,\zeta)$, both stochastic processes P and S are expected to have the same stability behavior. Therefore, $T(\varphi,b,\zeta)$ is chosen so that the drift term of the Itô differential Eq. (21) is independent of the processes $\varphi(t)$,$b(t)$, and $\zeta(t)$, so that

$$dS=\Lambda S dt+\sigma T_\zeta T^{-1}S dW. \tag{22}$$

Comparing Eqs. (21) and (22), it is seen that such a transformation $T(\varphi,b,\zeta)$ is given by the following equation:

$$(L_0+\varepsilon L_1)T=\Lambda T, \tag{23}$$

where

$$L_0 T=\frac{1}{2}\sigma^2 T_{\zeta\zeta}+\nu T_\zeta-T_\varphi+\cos\varphi T_b,$$

$$L_1 T=L_{10}T+\cos\zeta\cdot L_{11}T,$$

$$L_{10}T = b[p\gamma \sin\varphi T + \gamma \cos\varphi T_\varphi - (\kappa + \gamma b \sin\varphi)T_b],$$

$$L_{11}T = \mu \cos\varphi(p \sin\varphi T + \cos\varphi T_\varphi - b \sin\varphi T_b).$$

Equation (23) defines an eigenvalue problem for a second-order differential operator of three independent variables, in which Λ is the eigenvalue and $T(\varphi,b,\zeta)$ the associated eigenfunction. From Eq. (22), the eigenvalue Λ is seen to be the Lyapunov exponent of the pth moment of system (14) or (18), i.e., $\Lambda = \Lambda_{x(t)}(p)$.

In the following section, the method of regular perturbation is applied to the eigenvalue problem (23) to obtain a weak noise expansion of the moment Lyapunov exponent for system (18).

3 Weak Noise Expansion of the Moment Lyapunov Exponent

In this section, the method of regular perturbation is applied to obtained a weak noise expansion of the moment Lyapunov exponent $\Lambda_{x(t)}(p)$, i.e., the eigenvalue Λ of the eigenvalue problem (23) for small ε.

Applying the method of regular perturbation, both the eigenvalue $\Lambda_{x(t)}(p)$ and the eigenfunction $T(\varphi,b,\zeta)$ are expanded in power series of ε as

$$\Lambda_{x(t)}(p) = \sum_{n=0}^{\infty} \varepsilon^n \Lambda_n(p), \quad T(\varphi,b,\zeta) = \sum_{n=0}^{\infty} \varepsilon^n T_n(\varphi,b,\zeta), \tag{24}$$

in which $T_n(\varphi,b,\zeta)$, $n=0,1,\ldots,$ are periodic functions in φ of period 2π. Substituting the perturbation series (24) into the eigenvalue problem (23) and equating terms of equal power of ε yields the nth-order equation, $n=0,1,\ldots,$

$$L_0 T_n + L_1 T_{n-1} = \sum_{k=0}^{\infty} \Lambda_k T_{n-k}. \tag{25}$$

To simplify the expressions of the results in this paper, the following notations are employed:

$$\nu_{+1} = (\nu-1)^2 + \frac{1}{4}\sigma^4, \quad \nu_{-1} = (\nu+1)^2 + \frac{1}{4}\sigma^4,$$

$$\nu_{+2} = (\nu-2)^2 + \frac{1}{4}\sigma^4, \quad \nu_{-2} = (\nu+2)^2 + \frac{1}{4}\sigma^4,$$

$$\nu_{\pm 2}^{+} = \nu^2 - 4 + \frac{1}{4}\sigma^4, \quad \nu_{\pm 2}^{-} = \nu^2 - 4 - \frac{1}{4}\sigma^4,$$

$$\nu_{\pm 1}^{-} = \nu^2 - 1 - \frac{1}{4}\sigma^4, \quad \nu_{\pm 2i} = \nu^2 + 4 + \frac{1}{4}\sigma^4.$$

3.1 Zeroth-Order Perturbation. The equation for the zeroth-order perturbation is $L_0 T_0 = \Lambda_0 T_0$, or

$$\frac{\sigma^2}{2}\frac{\partial^2 T_0}{\partial \zeta^2} + \nu \frac{\partial T_0}{\partial \zeta} - \frac{\partial T_0}{\partial \varphi} + \cos\varphi \frac{\partial T_0}{\partial b} = \Lambda_0 T_0. \tag{26}$$

From the property of the moment Lyapunov exponent, it is known that $\Lambda_{x(t)}(0) = \Sigma_{n=0}^{\infty} \varepsilon^n \Lambda_n(0) = 0$, which results in $\Lambda_n(0) = 0$ for $n=0,1,\ldots$. Since the eigenvalue problem (26) does not contain p, the eigenvalue $\Lambda_0(p)$ is independent of p. Hence, $\Lambda_0(0) = 0$ leads to $\Lambda_0(p) = 0$. The solution of Eq. (26) is taken as

$$\Lambda_0(p) = 0, \quad T_0(\varphi,b,\zeta) = 1. \tag{27}$$

Since $\Lambda_0(p) = 0$, the associated adjoint differential equation of (26) is

$$\frac{\sigma^2}{2}\frac{\partial^2 T_0^*}{\partial \zeta^2} - \nu \frac{\partial T_0^*}{\partial \zeta} + \frac{\partial T_0^*}{\partial \varphi} - \cos\varphi \frac{\partial T_0^*}{\partial b} = 0. \tag{28}$$

Applying the method of separation of variables and letting $T_0^*(\varphi,b,\zeta) = F_0^*(\varphi,b)Z_0^*(\zeta)$ leads to

$$\frac{1}{F_0^*}\left(-\frac{\partial F_0^*}{\partial \varphi} + \cos\varphi \frac{\partial F_0^*}{\partial b}\right) = \frac{1}{Z_0^*}\left(\frac{\sigma^2}{2}\frac{d^2 Z_0^*}{d\zeta^2} - \nu \frac{dZ_0^*}{d\zeta}\right) = k.$$

For the $F_0^*(\varphi,b)$ equation, letting $k=0$ yields $F_0^*(\varphi,b) = \text{constant}$. Hence, $F_0^*(\varphi,b)$ can be taken as $F_0^*(\varphi,b) = \Phi_0^*(\varphi)B_0^*(b)$, where

$$\Phi_0^*(\varphi) = \frac{1}{2\pi}, \quad 0 \leqslant \varphi < 2\pi, \tag{29}$$

which is the probability density function of a uniformly distributed random variable φ between 0 and 2π, and

$$B_0^*(b) = \frac{1}{2}, \quad -1 \leqslant \varphi \leqslant 1, \tag{30}$$

which is the probability density function of a uniformly distributed random variable b between -1 and 1.

The Z_0^* equation becomes, for $k=0$,

$$\frac{1}{2}\sigma^2 \ddot{Z}_0^* - \nu \dot{Z}_0^* = 0, \tag{31}$$

and the solution is given by

$$Z_0^*(\zeta) = C_1 + C_2 \exp\left(\frac{2\nu}{\sigma^2}\zeta\right).$$

For $Z_0^*(\zeta)$ to be bounded, $C_2 = 0$ and $Z_0^*(\zeta) = C_1 = \text{constant}$.

Note that $\zeta(t) = \nu t + \sigma W(t) + \theta$, which is a linear function νt with superimposed noise $\sigma W(t)$, and $\zeta(t)$ appears as an angle of a sinusoidal function $\cos\zeta$, which is a periodic function of period 2π. Hence, after folding, the angle $\zeta(t)$ may be considered as taking values between 0 and 2π. $Z_0^*(\zeta)$ may then be chosen as

$$Z_0^*(\zeta) = \frac{1}{2\pi}, \quad 0 \leqslant \zeta < 2\pi, \tag{32}$$

which is the probability density function of a uniformly distributed random variable between 0 and 2π.

Hence $T_0^*(\varphi,b,\zeta) = \Phi_0^*(\varphi)B_0^*(b)Z_0^*(\zeta)$ represents the joint stationary probability density function of the independent random variables φ, b, and ζ, in which φ is uniformly distributed between 0 and 2π, b is uniformly distributed between -1 and 1, and ζ is uniformly distributed between 0 and 2π.

3.2 First-Order Perturbation. The first-order perturbation equation is

$$L_0 T_1 = \Lambda_1 T_0 - L_1 T_0. \tag{33}$$

Since the homogeneous equation $L_0 T_0 = 0$ has a nontrivial solution as given by Eq. (27), for Eq. (33) to have a solution it is required that, from the Fredholm alternative,

$$(\Lambda_1 T_0 - L_1 T_0, T_0^*) = 0, \tag{34}$$

where $T_0^*(\varphi,b,\zeta)$ is the solution of the adjoint Eq. (28) as obtained in Section 3.1, and (F_1,F_2) denotes the inner product of functions $F_1(\varphi,b,\zeta)$ and $F_2(\varphi,b,\zeta)$ defined by

$$(F_1,F_2) = \int_{\zeta=0}^{2\pi}\int_{b=-1}^{1}\int_{\varphi=0}^{2\pi} F_1(\varphi,b,\zeta)F_2(\varphi,b,\zeta) \times \Phi_0^*(\varphi)B_0^*(b)Z_0^*(\zeta)\,d\varphi\,db\,d\zeta.$$

For the simplicity of analysis, define the inner product for functions $F_1(\varphi,b)$ and $F_2(\varphi,b)$ as

$$\langle F_1, F_2\rangle = \int_{b=-1}^{1}\int_{\varphi=0}^{2\pi} F_1(\varphi,b)F_2(\varphi,b)\Phi_0^*(\varphi)B_0^*(b)\,d\varphi\,db.$$

Because $(T_0, T_0^*)=1$ and

$$L_1T_0=(L_{10}+\cos\zeta\cdot L_{11})T_0=f_0^{(1)}(\varphi,b)+\cos\zeta\cdot f_{\cos,1}^{(1)}(\varphi,b),$$

where $f_0^{(1)}(\varphi,b)=L_{10}T_0=p\gamma b\sin\varphi$, and $f_{\cos,1}^{(1)}(\varphi,b)=L_{11}T_0=p\mu\sin\varphi\cos\varphi$, the first-order perturbation of the moment Lyapunov exponent is, using Eqs. (29), (30), and (34),

$$\begin{aligned}\Lambda_1&=(L_1T_0,T_0^*)\\&=\langle f_0^{(1)}(\varphi,b),F_0^*(\varphi,b)\rangle+E[\cos\zeta]\cdot\langle f_{\cos,1}^{(1)}(\varphi,b),F_0^*(\varphi,b)\rangle\\&=0,\end{aligned}\tag{35}$$

in which $E[h(\zeta)]$ denotes the expected value of the random function $h(\zeta)$ with ζ being the uniformly distributed random variable between 0 and 2π.

Hence, Eq. (33) becomes

$$L_0T_1=g_0^{(1)}(\varphi,b)+\cos\zeta\cdot g_{\cos,1}^{(1)}(\varphi,b),\tag{36}$$

where $g_0^{(1)}(\varphi,b)=-f_0^{(1)}(\varphi,b)$, $g_{\cos,1}^{(1)}(\varphi,b)=-f_{\cos,1}^{(1)}(\varphi,b)$. Equation (36) is in the form of Eq. (50) and the solution is given in the Appendix by Eq. (66).

From the Appendix, the solution of Eq. (36) is obtained as

$$T_1(\varphi,b,\zeta)=G_0^{(1)}(\varphi,b)+\cos\zeta\cdot G_{\cos,1}^{(1)}(\varphi,b)+\sin\zeta\cdot G_{\sin,1}^{(1)}(\varphi,b),\tag{37}$$

where

$$G_0^{(1)}(\varphi,b)=\int_0^s g_0^{(1)}[\psi-r,\beta-\sin(\psi-r)]\,dr,$$

$$G_{\cos,1}^{(1)}(\varphi,b)=\int_0^s g_{\cos,1}^{(1)}[\psi-r,\beta-\sin(\psi-r)]\cdot c_1(r-s)\,dr,$$

$$G_{\sin,1}^{(1)}(\varphi,b)=-\int_0^s g_{\cos,1}^{(1)}[\psi-r,\beta-\sin(\psi-r)]\cdot s_1(r-s)\,dr,$$

in which $c_1(r-s)$ and $s_1(r-s)$ are as defined in Eq. (65), and the substitutions $\psi-s=\varphi$, $\beta=b+\sin\varphi$, and $s\to-\infty$ have been employed after the integration.

3.3 Second-Order Perturbation. The equation for the second-order perturbation is

$$L_0T_2=\Lambda_2T_0-L_1T_1.\tag{38}$$

From the Fredholm alternative, for Eq. (38) to have a solution, it is required that

$$(\Lambda_2T_0-L_1T_1,T_0^*)=0.\tag{39}$$

Since, from Eq. (37),

$$\begin{aligned}L_1T_1&=(L_{10}+\cos\zeta\cdot L_{11})[G_0^{(1)}(\varphi,b)+\cos\zeta\cdot G_{\cos,1}^{(1)}(\varphi,b)+\sin\zeta\\&\quad\cdot G_{\sin,1}^{(1)}(\varphi,b)]\\&=f_0^{(2)}(\varphi,b)+\sum_{k=1}^{2}[\cos k\zeta\cdot f_{\cos,k}^{(2)}(\varphi,b)+\sin k\zeta\cdot f_{\sin,k}^{(2)}(\varphi,b)],\end{aligned}$$

where

$$f_0^{(2)}(\varphi,b)=L_{10}G_0^{(1)}(\varphi,b)+\frac{1}{2}L_{11}G_{\cos,1}^{(1)}(\varphi,b),$$

$$f_{\cos,1}^{(2)}(\varphi,b)=L_{10}G_{\cos,1}^{(1)}(\varphi,b)+L_{11}G_0^{(1)}(\varphi,b),$$

$$f_{\sin,1}^{(2)}(\varphi,b)=L_{10}G_{\sin,1}^{(1)}(\varphi,b),$$

$$f_{\cos,2}^{(2)}(\varphi,b)=\frac{1}{2}L_{11}G_{\cos,1}^{(1)}(\varphi,b),\quad f_{\sin,2}^{(2)}(\varphi,b)=\frac{1}{2}L_{11}G_{\sin,1}^{(1)}(\varphi,b),$$

and noting that $E[\cos k\zeta]=E[\sin k\zeta]=0$, $k=1,2,\ldots$, one obtains the second-order perturbation of the moment Lyapunov exponent

$$\Lambda_2=(L_1T_1,T_0^*)=\langle f_0^{(2)},F_0^*\rangle=\frac{p(p+2)\mu^2\sigma^2\nu_{\pm 2i}}{32\nu_{+2}\nu_{-2}}.\tag{40}$$

In Section 2.2, it is shown, although not in a vigorous manner, that the range of the variable $b(t)$ is $-1\leqslant b(t)\leqslant 1$. To ascertain this conclusion, assume that the range of $b(t)$ is $-b_0\leqslant b(t)\leqslant b_0$. Following the same procedure, one obtains the second-order perturbation of the moment Lyapunov exponent as

$$\Lambda_2=\frac{p(p+2)b_0\mu^2\sigma^2\nu_{\pm 2i}}{32\nu_{+2}\nu_{-2}}.$$

Ignoring the viscoelastic effect, i.e., setting $\gamma=0$ in Eq. (14), system (14) becomes a two-dimensional system under bounded noise excitation. The resulting moment Lyapunov exponent should reduce to that obtained in Xie [13], which is the same as Eq. (40). It is obvious that the value of b_0 should be taken as 1.

Equation (38) becomes

$$L_0T_2=g_0^{(2)}(\varphi,b)+\sum_{k=1}^{2}[\cos k\zeta\cdot g_{\cos,k}^{(2)}(\varphi,b)+\sin k\zeta\cdot g_{\sin,k}^{(2)}(\varphi,b)],\tag{41}$$

where $g_0^{(2)}(\varphi,b)=\Lambda_2-f_0^{(2)}(\varphi,b)$, $g_{\cos,k}^{(2)}(\varphi,b)=-f_{\cos,k}^{(2)}(\varphi,b)$, $g_{\sin,k}^{(2)}(\varphi,b)=-f_{\sin,k}^{(2)}(\varphi,b)$, $k=1,2$. From the Appendix, the solution of Eq. (41) given by Eq. (66) is

$$\begin{aligned}T_2(\zeta,\varphi)&=G_0^{(2)}(\varphi,b)+\sum_{k=1}^{2}[\cos k\zeta\cdot G_{\cos,k}^{(2)}(\varphi,b)\\&\quad+\sin k\zeta\cdot G_{\sin,k}^{(2)}(\varphi,b)],\end{aligned}\tag{42}$$

where, for $k=1,2$,

$$G_0^{(2)}(\varphi,b)=\int_0^s g_0^{(2)}[\psi-r,\beta-\sin(\psi-r)]\,dr,$$

$$G_{\cos,k}^{(2)}(\varphi,b)=\int_0^s[g_{\cos,k}^{(2)}\cdot c_k(r-s)+g_{\sin,k}^{(2)}\cdot s_k(r-s)]\,dr,$$

$$G_{\sin,k}^{(2)}(\varphi,b)=\int_0^s[-g_{\cos,k}^{(2)}\cdot s_k(r-s)+g_{\sin,k}^{(2)}\cdot c_k(r-s)]\,dr,$$

in which the functions $g_0^{(2)}$, $g_{\cos,k}^{(2)}$, $g_{\sin,k}^{(2)}$ in the integrands are all functions of $[\psi-r,\beta-\sin(\psi-r)]$, and the substitutions $\psi-s=\varphi$, $\beta=b+\sin\varphi$, and $s\to-\infty$ are taken after integration.

3.4 Third-Order Perturbation. From Eq. (25), the equation for the third-order perturbation is

$$L_0T_3=\Lambda_2T_1+\Lambda_3T_0-L_1T_2.\tag{43}$$

According to the Fredholm alternative, for Eq. (43) to have a solution, it is required that

$$(\Lambda_2T_1+\Lambda_3T_0-L_1T_2,T_0^*)=0.$$

Since, from Eq. (42),

$$\begin{aligned}L_1T_2&=(L_{10}+\cos\zeta\cdot L_{11})\left\{G_0^{(2)}(\varphi,b)+\sum_{k=1}^{2}[\cos k\zeta\cdot G_{\cos,k}^{(2)}(\varphi,b)\right.\\&\quad\left.+\sin k\zeta\cdot G_{\sin,k}^{(2)}(\varphi,b)]\right\}=f_0^{(3)}(\varphi,b)+\sum_{k=1}^{3}[\cos k\zeta\\&\quad\cdot f_{\cos,k}^{(3)}(\varphi,b)+\sin k\zeta\cdot f_{\sin,k}^{(3)}(\varphi,b)],\end{aligned}$$

where

$$f_0^{(3)}(\varphi,b)=L_{10}G_0^{(2)}+L_{11}G_{\cos,1}^{(2)},$$

$$f_{\cos,1}^{(3)}(\varphi,b)=L_{10}G_{\cos,1}^{(2)}+L_{11}G_0^{(2)}+\frac{1}{2}L_{11}G_{\cos,2}^{(2)},$$

$$f_{\sin,1}^{(3)}(\varphi,b)=L_{10}G_{\sin,1}^{(2)}+\frac{1}{2}L_{11}G_{\sin,2}^{(2)},$$

$$f_{\cos,2}^{(3)}(\varphi,b)=L_{10}G_{\cos,2}^{(2)}+\frac{1}{2}L_{11}G_{\cos,1}^{(2)},$$

$$f_{\sin,2}^{(3)}(\varphi,b)=L_{10}G_{\sin,2}^{(2)}+\frac{1}{2}L_{11}G_{\sin,1}^{(2)},$$

$$f_{\cos,3}^{(3)}(\varphi,b)=\frac{1}{2}L_{11}G_{\cos,2}^{(2)},\quad f_{\sin,3}^{(3)}(\varphi,b)=\frac{1}{2}L_{11}G_{\sin,2}^{(2)},$$

and noting $E[\cos k\zeta]=E[\sin k\zeta]=0$, $k=1, 2, 3$, the third-order perturbation of the moment Lyapunov exponent is

$$\begin{aligned}\Lambda_3&=(L_1T_2,T_0^*)-\Lambda_2(T_1,T_0^*)=\langle f_0^{(3)},F_0^*\rangle-\Lambda_2\langle G_0^{(1)},F_0^*\rangle\\&=-p\gamma^2(9p^2\gamma-6p\gamma-16p\kappa+16\kappa)/48\\&\quad+p\gamma\{32p^2\mu^2\nu_+\nu_{-1}\nu_{\pm2}^--p\mu^2[64(\nu^2-1)(\nu^2-4)(5\nu^2\\&\quad+1)+16\sigma^4(5\nu^4-56\nu^2-15)-4\sigma^8(5\nu^2+24)-5\sigma^{12}]\\&\quad+256\mu^2\nu_{+2}\nu_{-2}\nu_{\pm1}^-\}/(2048\nu_{+1}\nu_{-1}\nu_{+2}\nu_{-2}).\end{aligned}\tag{44}$$

3.5 Moment Lyapunov Exponent and Lyapunov Exponent. The weak noise expansion of the moment Lyapunov exponent is obtained as

$$\Lambda_{x(t)}(p)=\varepsilon^2\Lambda_2+\varepsilon^3\Lambda_3+O(\varepsilon^4),\tag{45}$$

where Λ_2 and Λ_3 are given by Eqs. (40) and (44), respectively.

The procedure of regular perturbation presented in Sections 3.1–3.4 can be extended easily to higher-order terms and can be carried out using a symbolic computation software such as *Maple*. However, the numbers of terms involved in higher-order expansions increase drastically, and the higher-order terms obtainable are limited by the computer systems.

When the effect of viscoelasticity is absent, i.e., when $\gamma=0$, system (14) is a two-dimensional dynamic system under bounded noise parametric excitation. The moment Lyapunov exponents $\Lambda_{x(t)}(p)$ can be obtained from Eqs. (40) and (44) by setting $\gamma=0$:

$$\Lambda_2=\frac{p(p+2)\mu^2\sigma^2\nu_{\pm2i}}{32\nu_{+2}\nu_{-2}},\quad \Lambda_3=0,$$

which are the same as those obtained in [13].

The Lyapunov exponent for system (14) can be obtained from Eq. (45) by using the property of the moment Lyapunov exponent

$$\lambda_{x(t)}=\left.\frac{d\Lambda_{x(t)}(p)}{dp}\right|_{p=0}=\varepsilon^2\lambda_2+\varepsilon^3\lambda_3+O(\varepsilon^4),\tag{46}$$

where

$$\lambda_2=\frac{\mu^2\sigma^2\nu_{\pm2i}}{16\nu_{+2}\nu_{-2}},\quad \lambda_3=-\frac{1}{3}\gamma^2\kappa+\frac{\gamma\mu^2\nu_{\pm1}^-}{8\nu_{+1}\nu_{-1}}.$$

3.6 Stability Index. By definition, the stability index is the nontrivial zero of the moment Lyapunov exponent. For system (10), the moment Lyapunov exponent is given by

$$\Lambda_{q(\tau)}(p)=-\varepsilon^2c_0p+\omega\Lambda_{x(t)}(p),$$

and the stability index $\delta_{q(\tau)}$ is given by $\Lambda_{q(\tau)}(\delta_{q(\tau)})=0$, or

$$\Lambda_{x(t)}(\delta_{q(\tau)})-\varepsilon^2c\,\delta_{q(\tau)}=0,\tag{47}$$

where $c=c_0/\omega$.

Expanding the stability index $\delta_{q(\tau)}$ in power series of ε as $\delta_{q(\tau)}=\Sigma_{k=0}^{\infty}\varepsilon^k\delta_k$, substituting it into Eq. (47) along with Eq. (45), expanding, and equating terms of equal power of ε yields the equations

$$\varepsilon^2:\quad \delta_0\left\{-c+\frac{\mu^2\sigma^2(\delta_0+2)\nu_{\pm2i}}{32\nu_{+2}\nu_{-2}}\right\}=0,$$

$$\begin{aligned}\varepsilon^3:\quad &-c\,\delta_1+\frac{\delta_1(\delta_0+1)\mu^2\sigma^2\nu_{\pm2i}}{16\nu_{+2}\nu_{-2}}-\frac{1}{48}\delta_0\gamma^2(9\delta_0^2\gamma-6\delta_0\gamma\\&-16\delta_0\kappa+16\kappa)-\delta_0\gamma\mu^2\{-256\nu_{+2}\nu_{-2}\nu_{\pm1}^-+\delta_0[64(\nu^2-1)\\&\times(\nu^2-4)(5\nu^2+1)+16\sigma^4(5\nu^4-56\nu^2-15)-4\sigma^8(5\nu^2\\&+24)-5\sigma^{12}]-32\delta_0^2\nu_{+1}\nu_{-1}\nu_{\pm2}^-\}/(2048\nu_{+1}\nu_{-1}\nu_{+2}\nu_{-2})\\&=0,\end{aligned}\tag{48}$$

$$\vdots\quad\vdots$$

Using a symbolic computation software such as *Maple*, these equations can be easily manipulated and solved for $\delta_i, i=0,1,\ldots,$ to result in

$$\delta_0=-2+\frac{32c\nu_{+2}\nu_{-2}}{\mu^2\sigma^2\nu_{\pm2i}},\tag{49}$$

$$\begin{aligned}\delta_1&=8\gamma\{96\mu^4\sigma^4\nu_{\pm2i}\nu_{+1}\nu_{-1}[-\mu^2\nu_{\pm2}^-+2\gamma(\gamma+\kappa)\nu_{+2}\nu_{-2}]\\&\quad-c\mu^2\sigma^2\nu_{\pm2i}\nu_{+2}\nu_{-2}\cdot\{-3\mu^2[448\nu^6-16(144-7\sigma^4)\nu^4\\&\quad+4(528-256\sigma^4-7\sigma^8)\nu^2-(256+528\sigma^4+144\sigma^8+7\sigma^{12})]\\&\quad+256\gamma(21\gamma+8\kappa)\nu_{+1}\nu_{-1}\nu_{+2}\nu_{-2}\}\\&\quad+3072c^2\nu_{+1}\nu_{-1}(\nu_{+2})^2(\nu_{-2})^2(-\mu^2\nu_{\pm2}^-\\&\quad+12\gamma^2\nu_{+2}\nu_{-2})\}/[3\mu^6\sigma^6(\nu_{\pm2i})^3(16\nu^4-32\nu^2+8\sigma^4\nu^2+16\\&\quad+8\sigma^4+\sigma^8)].\end{aligned}$$

4 Numerical Results and Conclusions

In this paper, the moment Lyapunov exponents of a two-dimensional viscoelastic system under bounded noise parametric excitation are studied. An extension of the Khasminskii transformation is employed to convert the two-dimensional viscoelastic system under bounded noise excitation into a four-dimensional system in terms of P, φ, b, and ζ. An approach original applied by Wedig to two-dimensional Itô's system is used to set up the eigenvalue problem for the moment Lyapunov exponent. The method of regular perturbation is applied to obtain a weak noise expansion of the moment Lyapunov exponent in terms of the small fluctuation parameter. Weak noise expansions of the Lyapunov exponent and stability index are also obtained.

When the viscoelastic effect is not considered, i.e., when $\gamma=0$, system (14) is a two-dimensional system under bounded noise excitation. In the absence of noise perturbation, i.e., when $\sigma=0$, the two-dimensional system (14) under bounded noise parametric excitation (15) is further reduced to the Mathieu's equation. It is well-known that parametric resonance occurs when the nondimen-

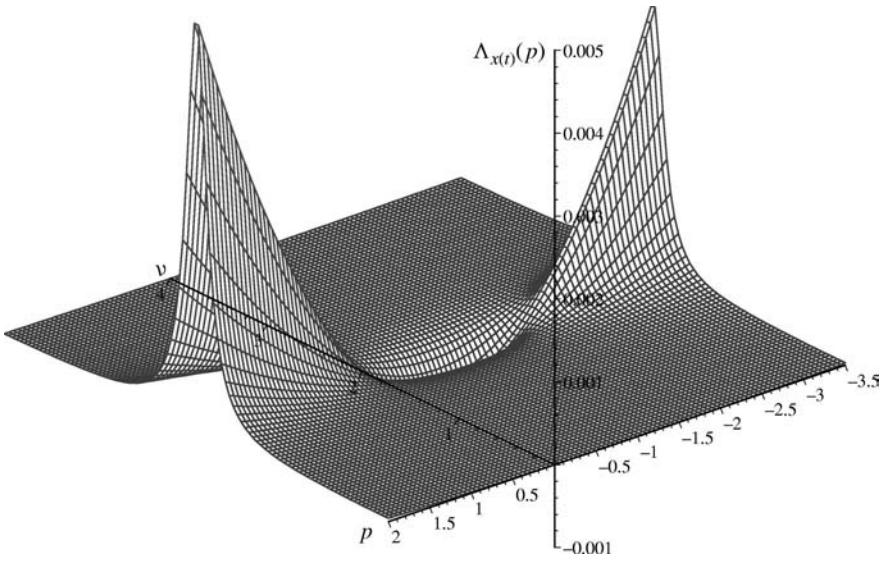


Fig. 1 Moment Lyapunov exponent $\Lambda_{x(t)}(p)$. ε=0.05, μ=1, κ=5, γ=0, σ=0.5.

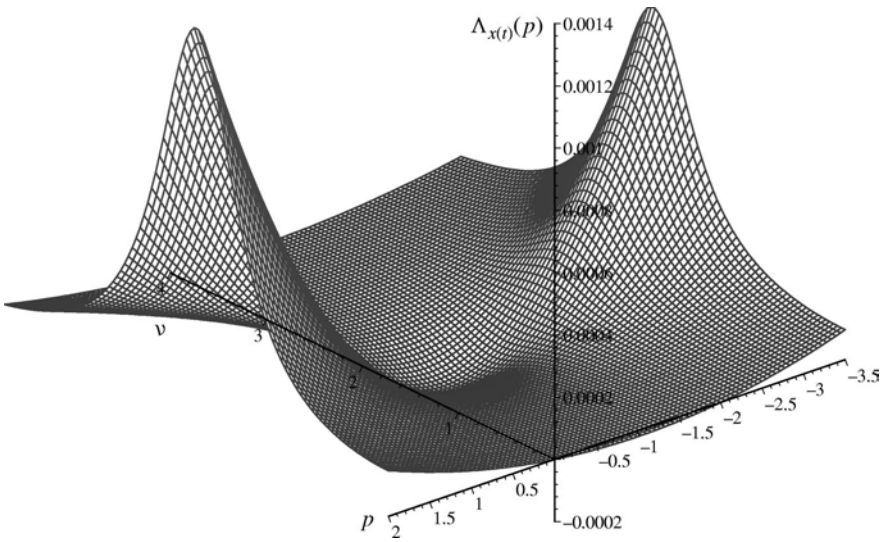


Fig. 2 Moment Lyapunov exponent $\Lambda_{x(t)}(p)$. ε=0.05, μ=1, κ=5, γ=0, σ=1.

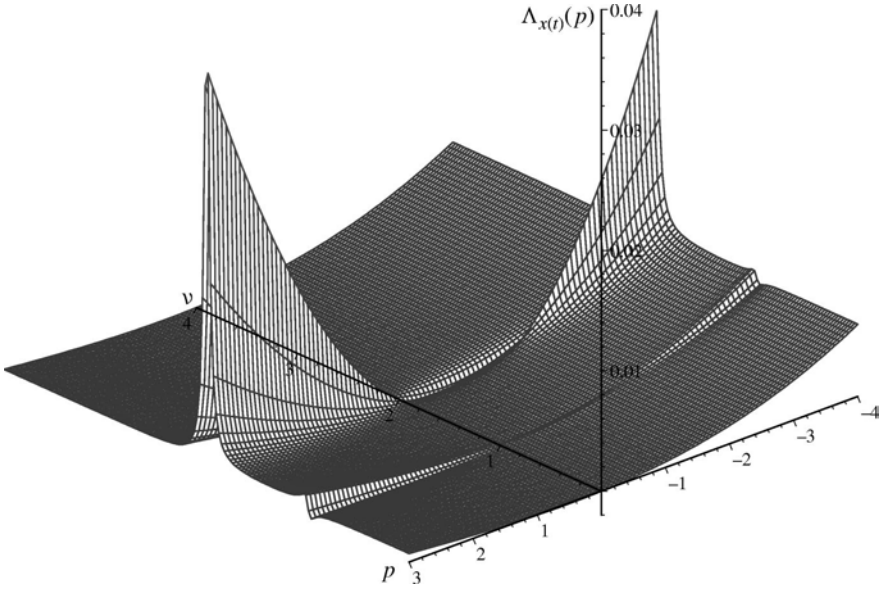


Fig. 3 Moment Lyapunov exponent $\Lambda_{x(t)}(p)$. ε=0.05, μ=1, κ=5, γ=1, σ=0.25.

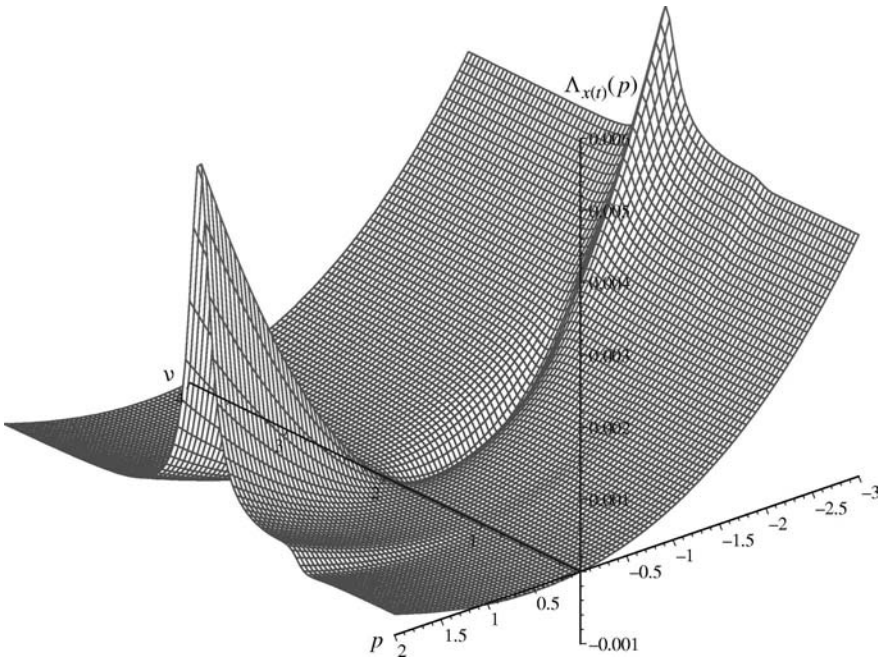


Fig. 4 Moment Lyapunov exponent $\Lambda_{x(t)}(p)$. ε=0.05, μ=1, κ=5, γ=1, σ=0.5.

sional excitation frequency $\nu/2$ is in the vicinity of 1, 1/2, 1/3, 1/4, ... For system (14) without the viscoelastic effect and noise fluctuation, the primary parametric resonance occurs in the vicinity of $\nu=2$, while the secondary parametric resonance occurs in the vicinity of $\nu=1$.

When the noise fluctuation parameter σ is not zero, the bounded noise is a sinusoidal function with noise superimposed. The larger the value of σ, the noisier the bounded noise $\cos\zeta(t)$, resulting in a smaller effect of the parametric resonance. The effect of primary parametric resonance can be clearly observed from

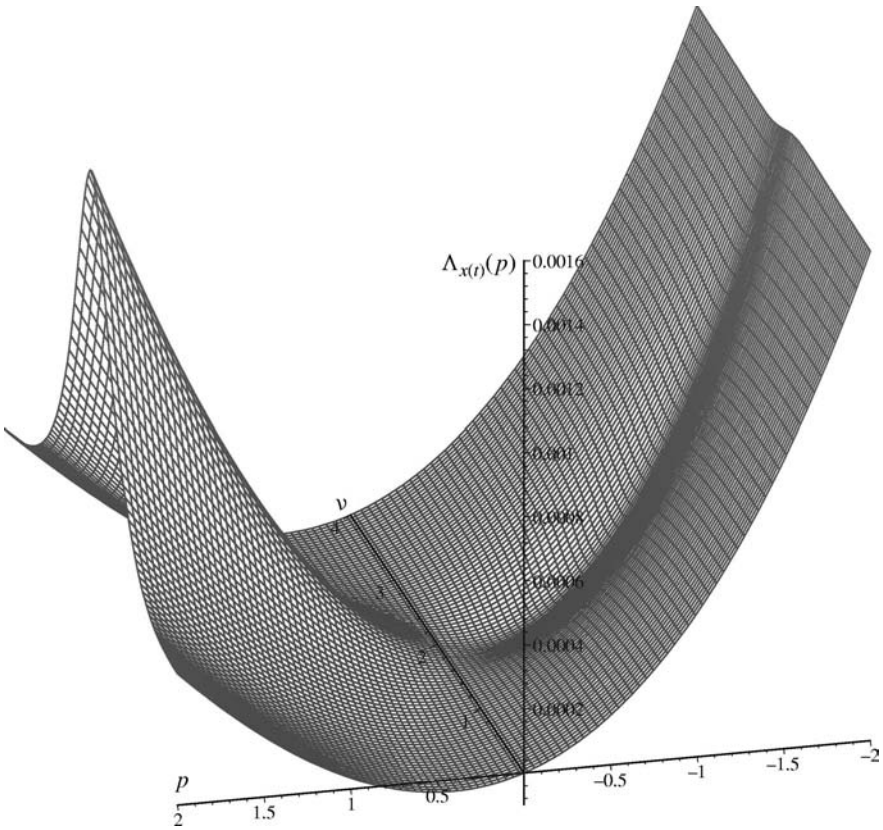


Fig. 5 Moment Lyapunov exponent $\Lambda_{x(t)}(p)$. ε=0.05, μ=1, κ=5, γ=1, σ=1.

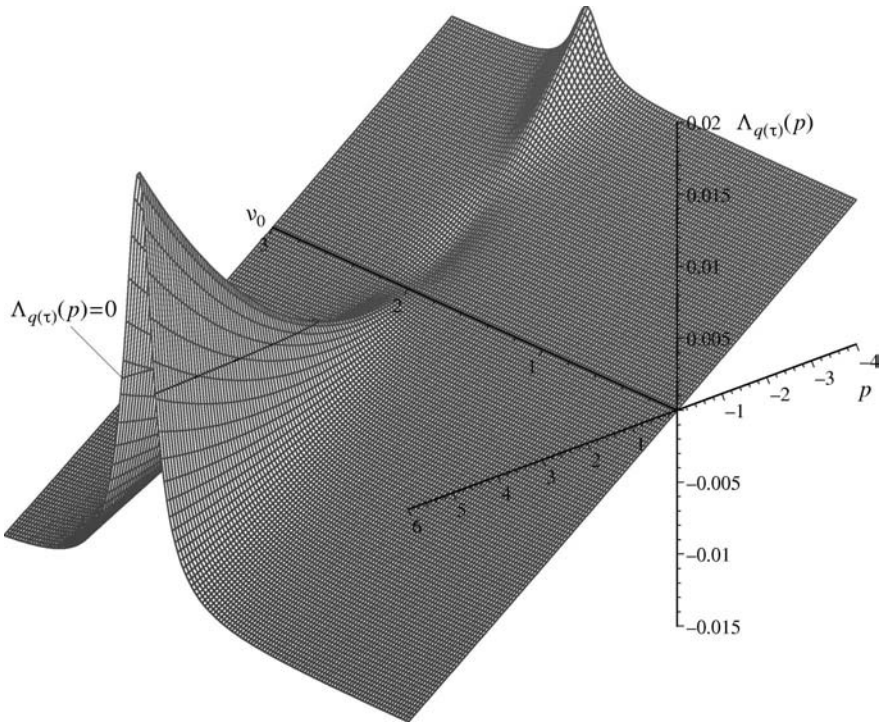


Fig. 6 Moment Lyapunov exponent $\Lambda_{q(\tau)}(p)$. $\varepsilon=0.05$, $c_0=1$, $p_0=0$, $\omega_0=1$, $\mu_0=1$, $\kappa_0=5$, $\gamma_0=0$, $\sigma_0=0.5$.

the typical plots (Figs. 1 and 2) of the moment Lyapunov exponent $\Lambda_{x(t)}(p)$ for system (14) given by Eq. (45) without the viscoelastic effect.

Typical results of the moment Lyapunov exponents for system (14) are plotted in Figs. 3–5, while those for system (10) are shown in Figs. 6 and 7. The viscoelastic term has a significant effect on dynamic stability of the system. This is clearly seen by comparing Figs. 1–2 and Figs. 3–5, Figs. 6 and 7. In general, the viscoelasticity has a stabilizing effect, which may be seen more clearly by plotting the stability index later in this section.

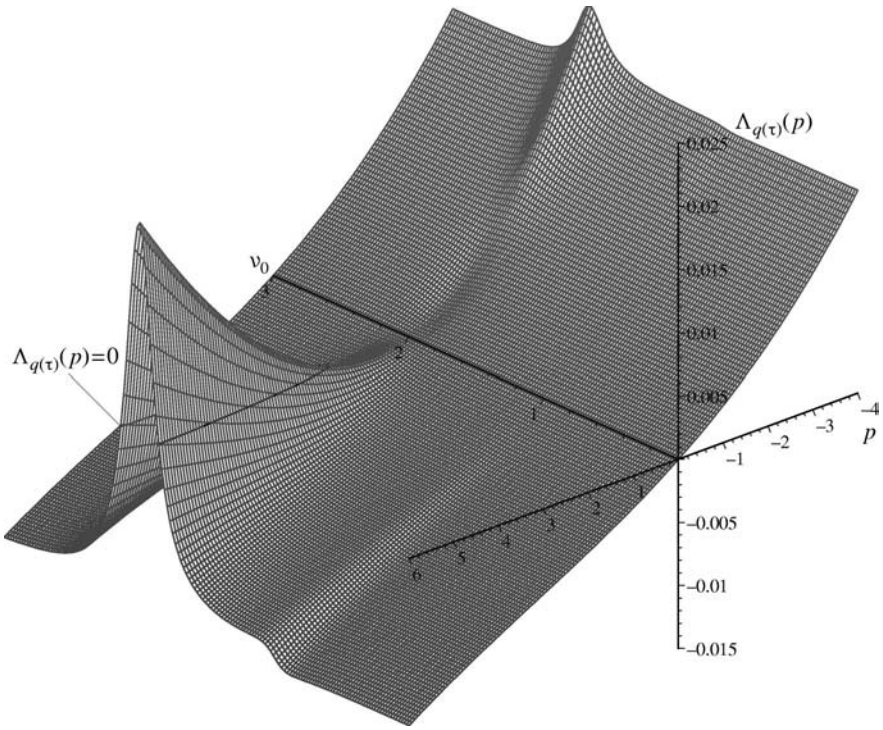


Fig. 7 Moment Lyapunov exponent $\Lambda_{q(\tau)}(p)$. $\varepsilon=0.05$, $c_0=1$, $p_0=0$, $\omega_0=1$, $\mu_0=1$, $\kappa_0=5$, $\gamma_0=1$, $\sigma_0=0.5$.

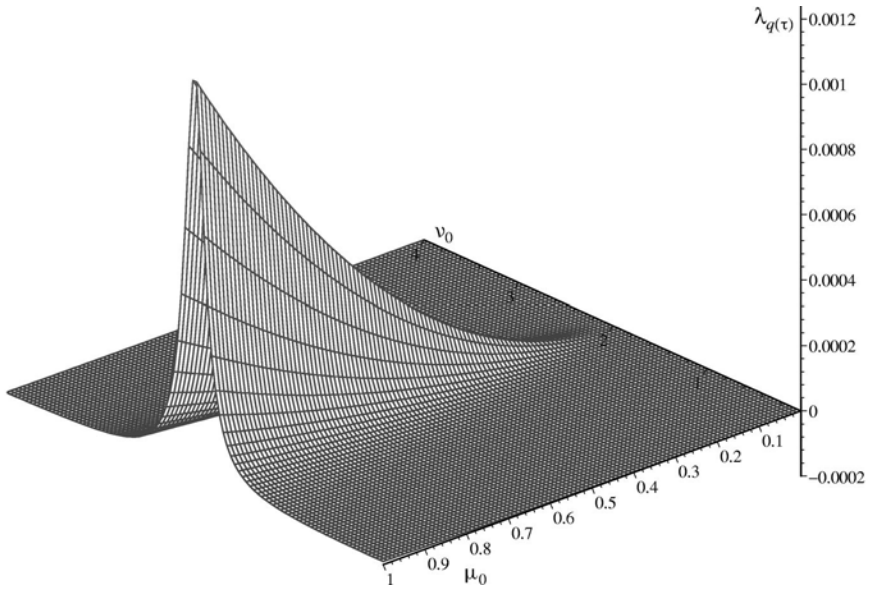


Fig. 8 Lyapunov exponent $\lambda_{q(\tau)}$. ε=0.05, c_0=0, p_0=0, ω_0=1, κ_0=5, γ_0=0, σ_0=0.5.

For the viscoelastic system (10), typical results of the Lyapunov exponent $\lambda_{q(\tau)}$ are shown in Figs. 8–10 for $\gamma_0=0$, 0.5, and 1, respectively. The effect of parametric resonance when ν_0 is in the vicinity of 2 can be clearly seen. It can also be observed that the viscoelastic term has a stability effect in the sense that the Lyapunov exponent $\lambda_{q(\tau)}$ is reduced with the increase of the value of γ_0.

The stability index is one of the single most important parameters in characterizing the dynamic stability of a stochastic system. From the definition of the stability index, it is clear that the larger the value of the stability index, the more stable of the system in the sense of moment stability.

For the viscoelastic system (10), it can be concluded that the viscoelasticity renders the effect of parametric resonance more prominent in the vicinity of $\nu_0=2$. The noise term has a destabilizing effect in the sense that the stability index $\delta_{q(\tau)}$ is reduced with the increase of the noise intensity parameter σ_0. It is also observed that the effect of parametric resonance diminishes with the increase of the noise intensity parameter σ_0. As expected, the damping term has a stabilizing effect. It is found that the effect of the relaxation rate is not very significant, although it reduces the effect of parametric resonance when the values of κ_0 is increased. The viscoelasticity has a stabilizing effect in the sense that the stability index is increased with the increase of the parameter γ_0.

It should be noted that the application of the method of regular perturbation in determining the moment Lyapunov exponent is based on the assumption that the noise fluctuation parameter σ is not small so that the eigenvalue problem (23) is not singular. Hence, the results obtained in this research cannot be used to deduce the results for the Mathieu's equation by setting σ to zero. In the case of small noise fluctuation, i.e., σ is small, a method of singular perturbation has to be employed to determine the moment Lyapunov exponent, which will be studied in future research.

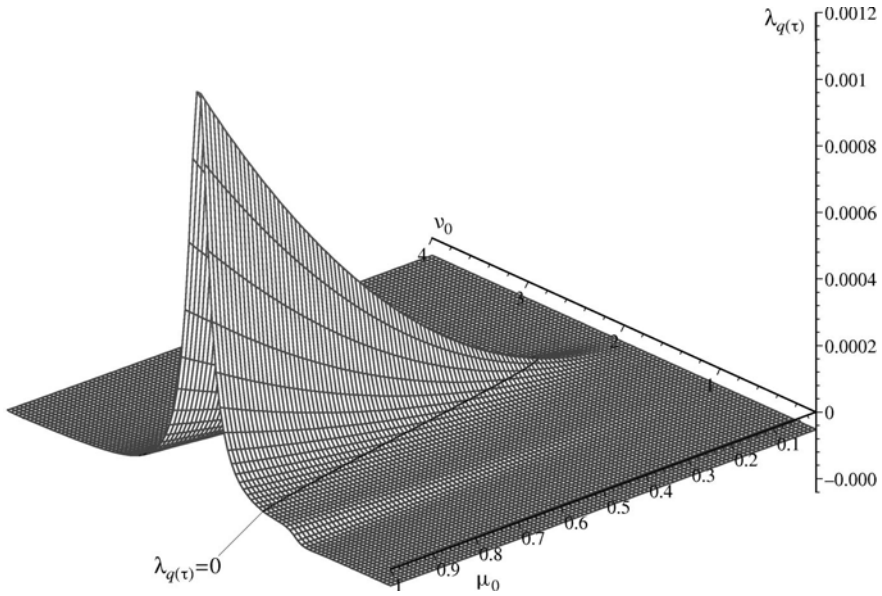


Fig. 9 Lyapunov exponent $\lambda_{q(\tau)}$. ε=0.05, c_0=0, p_0=0, ω_0=1, κ_0=5, γ_0=0.5, σ_0 =0.5.

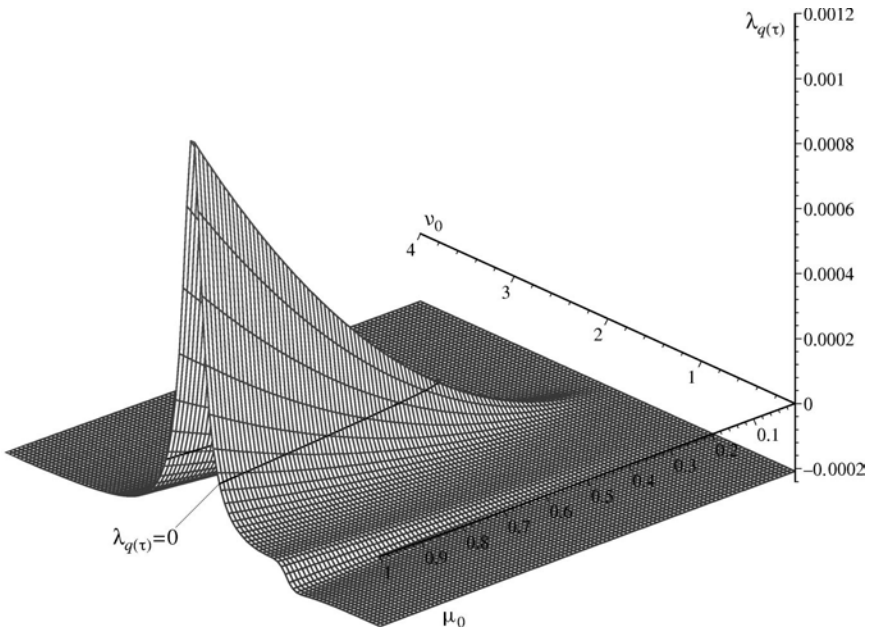


Fig. 10 Lyapunov exponent $\lambda_{q(\tau)}$. $\varepsilon=0.05$, $c_0=0$, $p_0=0$, $\omega_0=1$, $\kappa_0=5$, $\gamma_0=1$, $\sigma_0=0.5$.

Acknowledgment

The research for this paper was supported, in part, by the Natural Sciences and Engineering Research Council of Canada through Grant No. OGPO131355.

Appendix

Solution of $L_0T(\varphi,b,\zeta)=f(\zeta)g(\varphi,b)$. Consider the partial differential equation $L_0T(\varphi,b,\zeta)=f(\zeta)g(\varphi,b)$, or

$$\left(\frac{\sigma^2}{2}\frac{\partial^2}{\partial\zeta^2}+\nu\frac{\partial}{\partial\zeta}-\frac{\partial}{\partial\varphi}+\cos\varphi\frac{\partial}{\partial b}\right)T(\varphi,b,\zeta)=f(\zeta)g(\varphi,b). \tag{50}$$

Introducing an auxiliary time t' to Eq. (50) leads to

$$\left(\frac{\partial}{\partial t'}+\frac{\sigma^2}{2}\frac{\partial^2}{\partial\zeta^2}+\nu\frac{\partial}{\partial\zeta}-\frac{\partial}{\partial\varphi}+\cos\varphi\frac{\partial}{\partial b}\right)T(\varphi,b,\zeta;t') =f(\zeta)g(\varphi,b). \tag{51}$$

Applying the transformation

$$\psi=\frac{1}{2}(t'+\varphi),\quad s=\frac{1}{2}(t'-\varphi),\quad \beta=b+\sin\varphi,$$

or the inverse transformation

$$t'=\psi+s,\quad \varphi=\psi-s,\quad b=\beta-\sin(\psi-s),$$

Eq. (51) becomes

$$\left(\frac{\partial}{\partial s}+\frac{\sigma^2}{2}\frac{\partial^2}{\partial\zeta^2}+\nu\frac{\partial}{\partial\zeta}\right)T(\psi,\beta,\zeta;s)=f(\zeta)g[\psi-s,\beta-\sin(\psi-s)]. \tag{52}$$

Applying Duhamel's principle ([20]), the solution $T(\psi,\beta,\zeta;s)$ to Eq. (52) is given by

$$T(\psi,\beta,\zeta;s)=\int_0^s V(\psi,\beta,\zeta;s;r)dr, \tag{53}$$

where $V(\psi,\beta,\zeta;s;r)$ is the solution of the homogeneous equation

$$\left(\frac{\partial}{\partial s}+\frac{\sigma^2}{2}\frac{\partial^2}{\partial\zeta^2}+\nu\frac{\partial}{\partial\zeta}\right)V(\psi,\beta,\zeta;s;r)=0,\qquad \text{for } s>r,$$
$$V(\psi,\beta,\zeta;r;r)=f(\zeta)g[\psi-r,\beta-\sin(\psi-r)],\qquad \text{for } s=r. \tag{54}$$

To solve Eq. (54), consider the equation

$$\left(\frac{\partial}{\partial s}+\frac{\sigma^2}{2}\frac{\partial^2}{\partial\zeta^2}+\nu\frac{\partial}{\partial\zeta}\right)P(s,\zeta;t,z)=0,\quad s<t,$$
$$P(t,\zeta;t,z)=\lim_{s\uparrow t}P(s,\zeta;t,z)=\delta(z-\zeta). \tag{55}$$

Equation (55) is the Kolmogorov's backward equation for the transition probability function $P(s,\zeta;t,z)$. It is well known ([21]) that the transition probability $P(s,\zeta;t,z)$ is also the fundamental solution of the forward or Fokker-Planck equation, i.e., for the initial condition s and ζ fixed,

$$\left[\frac{\partial}{\partial t}-\frac{\sigma^2}{2}\frac{\partial^2}{\partial z^2}+\nu\frac{\partial}{\partial z}\right]P(s,\zeta;t,z)=0,\quad t>s,$$
$$P(s,\zeta;s,z)=\lim_{t\downarrow s}P(s,\zeta;t,z)=\delta(z-\zeta). \tag{56}$$

Applying the Fourier transformation ([20])

$$\tilde{P}(s,\zeta;t,k)=\mathcal{F}[P(s,\zeta;t,z)]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{ikz}P(s,\zeta;t,z)dz$$

to Eq. (56) leads to

$$\frac{\partial\tilde{P}}{\partial t}+\left(\frac{1}{2}\sigma^2k^2-i\nu k\right)\tilde{P}=0,$$
$$\tilde{P}(s,\zeta;s,k)=\frac{1}{\sqrt{2\pi}}e^{ik\zeta}. \tag{57}$$

Equation (57) can be solved using the method of characteristics to give

$$\tilde{P}(s,\zeta;t,k)=\frac{1}{\sqrt{2\pi}}\exp\left[ik\zeta-\left(\frac{1}{2}\sigma^2k^2-i\nu k\right)(t-s)\right]. \tag{58}$$

Applying the inverse Fourier transformation

$$P(s,\zeta;t,z)=\mathcal{F}^{-1}[\tilde{P}(s,\zeta;t,k)]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-ikz}\tilde{P}(s,\zeta;t,k)dk$$

to Eq. (58) leads to

$$P(s,\zeta;t,z)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp\left\{i[(\zeta-z)+\nu(t-s)]k-\frac{1}{2}\sigma^2k^2(t-s)\right\}dk.$$

Using the integral formula

$$\int_{-\infty}^{+\infty}\exp(\pm qx-p^2x^2)dx=\frac{\sqrt{\pi}}{p}\exp\left(\frac{q^2}{4p^2}\right),\quad p>0,$$

as in Formula 3.323(2) of reference [22], one obtains

$$P(s,\zeta;t,z)=\frac{1}{\sqrt{2\pi}\sigma_{z(t)}}\exp\left\{-\frac{[z-\mu_{z(t)}]^2}{2\sigma_{z(t)}^2}\right\},\tag{59}$$

where

$$\mu_{z(t)}=\zeta+\nu(t-s),\quad \sigma_{z(t)}^2=\sigma^2(t-s).\tag{60}$$

Hence, for the initial condition $\zeta(s)$ fixed, $z(t)$ is a normally distributed random variable with mean $\mu_{z(t)}$ and standard deviation $\sigma_{z(t)}$, the probability density function of which is given by Eq. (59).

From Eqs. (54) and (55), the solution $V(\psi,\beta,\zeta;s;r)$ to Eq. (54) is given by

$$V(\psi,\beta,\zeta;s;r)=g[\psi-r,\beta-\sin(\psi-r)]\int_{-\infty}^{\infty}f(z)P(s,\zeta;r,z)dz,\tag{61}$$

where

$$E[f(z(r))]=\int_{-\infty}^{\infty}f(z)P(s,\zeta;r,z)dz,$$

is the expected value of the random variable $f(z(r))$ with $z(r)$ being the normally distributed random variable as defined in (59) and (60).

Combining Eqs. (53) and (61), the solution to Eq. (52) is given by

$$T(\psi,\beta,\zeta;s)=\int_0^s g[\psi-r,\beta-\sin(\psi-r)]E[f(z(r))]dr.\tag{62}$$

The solution $T(\varphi,b,\zeta)$ to Eq. (50) is obtained by replacing $\varphi=\psi-s, b=\beta-\sin(\psi-s)$, and passing the limit $s\to-\infty$.

For the special cases when $f(\zeta)=\sin k\zeta$ or $\cos k\zeta$, one has

$$E\left[\begin{Bmatrix}\sin kz(r)\\ \cos kz(r)\end{Bmatrix}\right]=\frac{1}{\sqrt{2\pi}\sigma_{z(r)}}\int_{-\infty}^{+\infty}\begin{Bmatrix}\sin kz\\ \cos kz\end{Bmatrix}\times\exp\left\{-\frac{[z-\mu_{z(r)}]^2}{2\sigma_{z(r)}^2}\right\}dz=\exp\left[-\frac{1}{2}k^2\sigma_{z(r)}^2\right]\begin{Bmatrix}\sin k\mu_{z(r)}\\ \cos k\mu_{z(r)}\end{Bmatrix},\tag{63}$$

in which the integral formulas

$$\int_{-\infty}^{+\infty}\exp(-q^2x^2)\begin{Bmatrix}\sin[p(x+\lambda)]\\ \cos[p(x+\lambda)]\end{Bmatrix}dx=\frac{\sqrt{\pi}}{q}\exp\left(-\frac{p^2}{4q^2}\right)\begin{Bmatrix}\sin p\lambda\\ \cos p\lambda\end{Bmatrix},$$

as given in 3.896(1) and (2) of reference [22] have been employed. Substituting Eq. (60) into (63) results in

$$E\left[\begin{Bmatrix}\sin kz(r)\\ \cos kz(r)\end{Bmatrix}\right]=c_k(r-s)\begin{Bmatrix}\sin k\zeta\\ \cos k\zeta\end{Bmatrix}+s_k(r-s)\begin{Bmatrix}\cos k\zeta\\ -\sin k\zeta\end{Bmatrix},\tag{64}$$

in which the following notations are used:

$$\begin{Bmatrix}s_k(r-s)\\ c_k(r-s)\end{Bmatrix}=\exp\left[-\frac{1}{2}k^2\sigma_{z(r)}^2\right]\begin{Bmatrix}\sin k\nu(r-s)\\ \cos k\nu(r-s)\end{Bmatrix}.\tag{65}$$

Substituting Eqs. (64) into Eq. (62), one obtains the solution of Eq. (50) as, when $f(\zeta)=\sin k\zeta$ and $\cos k\zeta$, respectively,

$$T(\psi,\beta,\zeta;s)=\int_0^s\begin{Bmatrix}\sin k\zeta\cdot c_k(r-s)+\cos k\zeta\cdot s_k(r-s)\\ \cos k\zeta\cdot c_k(r-s)-\sin k\zeta\cdot s_k(r-s)\end{Bmatrix}\times g[\psi-r,\beta-\sin(\psi-r)]dr.\tag{66}$$

References

[1] Stratonovich, R. L., 1967, *Topics in the Theory of Random Noise*, Vol. II, Gordon and Breach, New York.

[2] Dimentberg, M. F., 1988, *Statistical Dynamics of Nonlinear and Time-Varying Systems*, John Wiley and Sons, New York.

[3] Wedig, W., 1989, "Analysis and Simulation of Nonlinear Stochastic Systems," *Nonlinear Dynamics in Engineering Systems*, W. Schiehlen, ed., Springer-Verlag, Berlin, pp. 337–344.

[4] Lin, Y. K., and Cai, G. Q., 1995, *Probabilistic Structural Dynamics: Advanced Theory and Applications*, McGraw-Hill, New York.

[5] Ariaratnam, S. T., 1996, "Stochastic Stability of Viscoelastic Systems Under Bounded Noise Excitation," *IUTAM Symposium on Advances in Nonlinear Stochastic Mechanics*, A. Naess and S. Krenk, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 11–18.

[6] Arnold, L., Oeljeklaus, E., and Pardoux, E., 1986, "Almost Sure and Moment Stability for Linear Itô Equations," *Lyapunov Exponents* (Lecture Notes in Mathematics, 1186), L. Arnold and V. Wihstutz, eds., Springer-Verlag, Berlin, pp. 85–125.

[7] Arnold, L., Kliemann, W., and Oeljeklaus, E., 1986, "Lyapunov Exponents of Linear Stochastic Systems," *Lyapunov Exponents* (Lecture Notes in Mathematics, 1186), L. Arnold and V. Wihstutz, eds., Springer-Verlag, Berlin, pp. 129–159.

[8] Arnold, L., 1998, *Random Dynamical Systems*, Springer-Verlag, Berlin.

[9] Arnold, L., Doyle, M. M., and Sri Namachchivaya, N., 1997, "Small Noise Expansion of Moment Lyapunov Exponents for Two-Dimensional Systems," Dynamics and Stability of Systems, **12**, No. 3, pp. 187–211.

[10] Khasminskii, R. Z., and Moshchuk, N., 1998, "Moment Lyapunov Exponent and Stability Index for Linear Conservative System With Small Random Perturbation," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **58**, No. 1, pp. 245–256.

[11] Sri Namachchivaya, N., and Vedula, L., 2000, "Stabilization of Linear Systems by Noise: Application to Flow Induced Oscillations," Dyn. Stab. Syst., **15**, No. 2, pp. 185–208.

[12] Xie, W.-C., 2001, "Moment Lyapunov Exponents of a Two-Dimensional System Under Real Noise Excitation," J. Sound Vib., **239**, No. 1, pp. 139–155.

[13] Xie, W.-C., 2000, "Moment Lyapunov Exponents of a Two-Dimensional System under Bounded Noise Parametric Excitation," J. Sound Vib., submitted for publication.

[14] Ariaratnam, S. T., 1993, "Stochastic Stability of Linear Viscoelastic Systems," Probabl. Eng. Mech., **8**, pp. 153–155.

[15] Larinov, G. S., 1969, "Investigation of the Vibration of Relaxing Systems by the Averaging Method," Mech. Polym., **5**, pp. 714–720 (English translation).

[16] Potapov, V. D., 1994, "On Almost Sure Stability of a Viscoelastic Column Under Random Loading," J. Sound Vib., **173**, No. 3, pp. 301–308.

[17] Potapov, V. D., 1997, "Numerical Method for Investigation of Stability of Stochastic Integro-Differential Equations," Appl. Numer. Math., **24**, pp. 191–201.

[18] Wedig, W., 1988, "Lyapunov Exponent of Stochastic Systems and Related Bifurcation Problems," *Stochastic Structural Dynamics—Progress in Theory and Applications*, S. T. Ariaratnam, G. I. Schuëller, and I. Elishakoff, eds., Elsevier, New York, pp. 315–327.

[19] Khasminskii, R. Z., 1967, "Necessary and Sufficient Conditions for the Asymptotic Stability of Linear Stochastic Systems," Theor. Probab. Appl., **12**, pp. 144–147 (English translation).

[20] Zauderer, E., 1989, *Partial Differential Equations of Applied Mathematics*, 2nd Ed., John Wiley and Sons, New York.

[21] Arnold, L., 1974, *Stochastic Differential Equations: Theory and Applications*, John Wiley and Sons, New York.

[22] Gradshteyn, I. S., and Ryzhik, I. M., 1994, *Table of Integrals, Series, and Products*, 5th Ed., Academic Press, San Diego, CA.

Width-Wise Variation of Magnetic Tape Pack Stresses

Y. M. Lee
Assoc. Mem. ASME

J. A. Wickert
Fellow ASME

Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213-3890

A model is developed for predicting the stress and displacement fields within a magnetic tape pack, where those quantities are allowed to vary in both the pack's radial and transverse (cross-tape) directions. As has been the case in previous analyses based upon one-dimensional wound roll models, the present approach accounts for the anisotropic and nonlinear constitutive properties of the layered tape, and the incremental manner in which the pack is wound. Further, such widthwise variation effects as differential hub compliance and nonuniform winding tension, which can be significant in data cartridge design, are also treated in the model. The pack is analyzed through a two-dimensional axisymmetric finite element model that couples individual representations of the hub/flange and layered tape substructures. The bulk radial elastic modulus of the tape, which depends on the in-pack radial stress, is measured for a variety of media samples, and a reduced-order model is developed to capture the nonlinear modulus-stress correlation. The stiffness matrix of the hub/flange at its interface with the media provides a mixed boundary condition to the tape substructure. In this manner, design-specific hubs can be readily analyzed, and criteria for their optimization explored. Simulations of several cartridge designs are presented, and the roles of hub compliance and wound-in tension gradient in setting the pack's stress field and cross-tape width change are discussed. [DOI: 10.1115/1.1460911]

1 Introduction

Magnetic data storage on flexible media has played an important role in computer systems since the early 1950s. Tape systems currently span a broad range of consumer and enterprise storage applications, including audio, video, instrumentation, and computer systems. Within the next decade, compact tape cartridges for computer data storage are projected to provide several terabytes of capacity through a nearly 50-fold growth in areal density and substantial reduction in media thickness. In the rapidly moving computer industry, tape storage systems remain competitive through their low storage cost per megabyte of data and their high capacity. The technologies of magnetic and optical disk recording, at least presently, do not offer the same combination of advantages, and while other emerging technologies do offer strengths in one or more areas, overall they do not meet the entire set of performance needs that tape cartridge drives address.

One application of magnetic tape systems is backup/restore missions on medium-to-large scale computers and networks. The number of cartridges required for such applications can grow to the hundreds or thousands for systems in the graphics, medical imaging, and entertainment industries. For instance, a large system with a thousand gigabytes of on-line disk capacity can require about 3000 to 5000 gigabytes to hold several full and daily incremental backups. Although a primary use of tape systems is for off-line and archival data storage, mass storage systems are today supplemented by automated robotic libraries that make tape storage a nearly on-line computer peripheral. With growth in network-based business and information retrieval systems, the accessibility and reliability of vast data libraries on magnetic tape remains an important technical issue.

As tape technology continues to migrate towards thinner media substrates ($\sim$4–5 μm), lower winding tensions ($\sim$0.28–0.56 N) and higher speeds ($\sim$7–10 m/s) to increase capacity and reduce access time, tape mechanics problems, for which solutions had been engineered, often recur. For instance, reductions of tape thickness and data track width provide the direct means to increase volumetric storage density, and therefore the capacity of a given sized cartridge. On the other hand, as the media becomes thinner, its handling and mechanical stability become more problematic. Over roughly the past decade, media substrates have been reduced in thickness from some 30 μm to 5 μm, and as track widths fall to several microns, the dimensional stability of the media in the presence of aggressive in-pack stresses likewise becomes critical. Uneven or excessive compressive stress within a cartridge can cause tape layers, individually or together, to buckle in the pack's radial direction, axial direction, or both. The photograph of Fig. 1 depicts a so-called spoking defect in which the tape layers buckled locally to such an extent that an internal gap formed. Such buckling is facilitated by the combined conditions of low radial stress, and high compressive circumferential or transverse stresses.

Traditionally, research on the stress field in magnetic tape packs or other types of wound rolls has emphasized one dimensional models wherein the hub and tape are each treated as being infinitely wide, and with uniform mechanical properties and tension across the width. Based on the constitutive properties specified for the media, the one-dimensional models can be categorized broadly into four groups: linearly elastic ([1–3]), linearly viscoelastic ([4–6]), nonlinearly elastic ([7–11]), and nonlinearly viscoelastic ([12]). Such studies consider only radial and circumferential stresses, treat the media as being orthotropic, specify that the hub has uniform cross-tape compliance, and ignore effects associated axial and shear stress components ([13]).

The tape pack itself is generally modeled as a layered structure which is developed incrementally as a succession of pretensioned hoops that are shrunk-fit onto the underlying pack. That process is a nonlinear one to the extent that the effective radial elastic modulus of the tape region, at the bulk level, is a function of the interlayer radial stress which in turn changes as the pack is formed. In short, the mechanical properties of a wound tape pack are often quite different from those of a single isolated layer because of entrained air and surface asperity contact. Accordingly, the bulk radial modulus plays an important role in pack stress modeling

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. 1, 2001; final revision, Nov. 22, 2001. Associate Editor: R. C. Benson. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

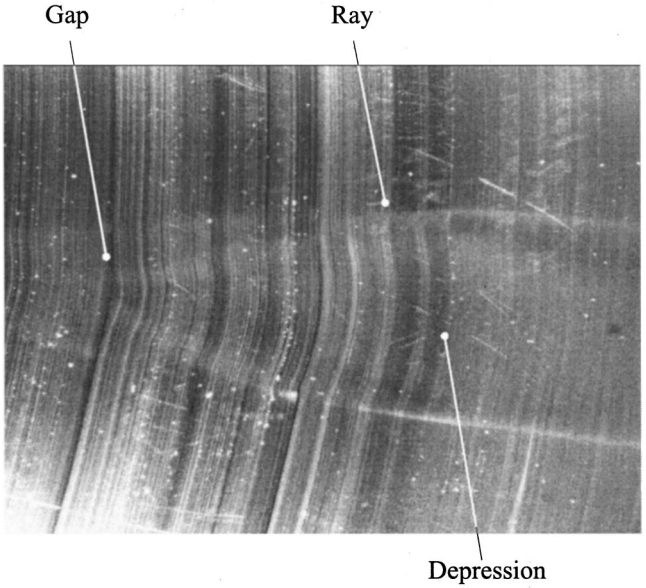


Fig. 1 Magnified view of a region in a magnetic tape pack exhibiting interlayer buckling

([8,9,14]). Since the media's substrate is typically a polymer such as PET (polyethylene terephthalate) or PEN (polyethylene naphthalate), viscoelastic effects, although not considered here, set the manner in which wound-in stresses relax over time. Other effects including wound-on tension losses, temperature, humidity, and finite deformation have also been studied in various applications ([3,10,11]).

Certain two-dimensional stress models have been considered with a view towards understanding stress dependencies which arise due to thickness variations across the width ([15–17]). The nonuniform widthwise winding tension was simulated by stacking web layers having varying widthwise thickness. Each of these approaches was based on the assumptions that the pack could be partitioned across its width into a discrete number of segments that do not couple, and that the stresses and displacements so developed are width-independent. Notably, shear-extension and bending-extension coupling were ignored in those treatments.

In what follows, a two-dimensional axisymmetric finite element model is developed in order to analyze the magnetic tape pack stress problem. The internal stresses are allowed to vary both in the pack's radial and axial directions, and four stress components—radial, circumferential, axial, and shear—as well as two displacements—radial and axial—are determined. The present approach is appropriate for treating realistic cartridge hub designs, nonuniform widthwise tension profiles which can result from guiding imperfections, and the tape's cross-track width change. Experiments are further discussed in which the bulk compressive radial modulus is measured for several media specimens. A reduced-order treatment of individual layer and nonlinear interface compression is discussed in order to model the experimental data and incorporate it into the stress analysis.

2 Finite Width Tape Pack Model

A magnetic tape pack is formed by winding a continuous stream of media, having specified tension and speed, onto a hub. In general, the hub will have a radial stiffness that varies across its width, and the hub's geometry and materials are designed and chosen to meet various functionality requirements. Shown illustratively in Fig. 2(*a*) is the so-called 9840-style cartridge hub which has a plane of symmetry with respect to the media's centerline. The central rib at the hub's centerline reinforces the structure and allows more tape to be wound at higher tension. Since the

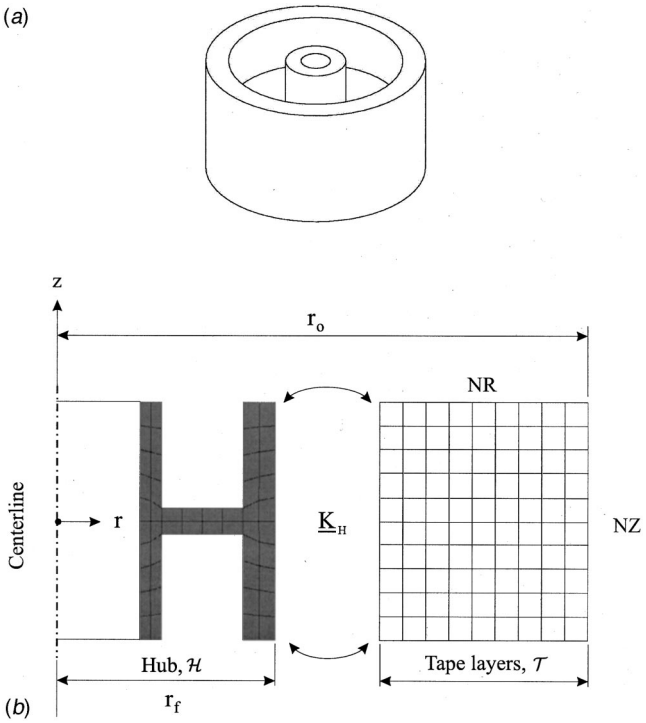


Fig. 2 (*a*) An axisymmetric hub that is also symmetric with respect to the pack's midplane, and (*b*) model of the hub and tape layer substructures

flanges and the hub connect only through a press-fit snap lock, the stiffness afforded to the hub by the flanges is neglected in this case. Figure 3(*a*) depicts a second type of hub design, termed 3480-style, which is a common format in the tape storage industry and which has no particular midplane symmetry. The cartridge's two flanges are formed from different materials, and the structure's cross-sectional geometry is generally step-shaped in order to facilitate attaching the hub to the drive motor's spindle.

By way of motivation, Fig. 4 indicates how the collocated point radial compliance of these two hub designs changes across the media's width. The ratio of the maximum-to-minimum compliance values for the symmetric hub design is roughly 5:1, and the ratio is roughly 4:1 for the asymmetric design. The cross-tape compliance variation is significant in each case, and sets the boundary condition afforded to the layered tape region at its interface with the hub. In what follows, the effects of differential hub compliance are explicitly treated in the pack stress model in order to develop a stress simulation of greater fidelity than is available through existing conventional one-dimensional models.

In Figs. 2 and 3, the tape has thickness h, and it is wound layer by layer into a nearly cylindrical shape having outer radius r_o and inner radius r_f common with the hub's winding face surface. Formed from N individual layers which are conceptually shrunk-fit onto one another in an incremental quasi-static manner, the tape region is treated as being a composite material having bulk anisotropic and nonlinear material properties. In general, the hub is an integral solid structure, and it is often made of fiber-reinforced plastic. In short, the material properties of the hub and tape regions can be substantially different.

In order for the analysis to properly account for various hub geometries, the tape pack is conceptually separated into the hub and tape substructures following the approach developed in [13]. The respective subdomains are denoted $\mathcal{H}=\{(r,\theta,z): r<r_f, 0<\theta<2\pi, -w/2<z<w/2\}$ over the hub, and $\mathcal{T}=\{(r,\theta,z): r_f<r<r_o, 0<\theta<2\pi, -w/2<z<w/2\}$ over the tape. For the symmetric hub shown in Fig. 2, the hub's winding face and outer radius are the same. For the asymmetric hub of Fig. 3, the winding face is

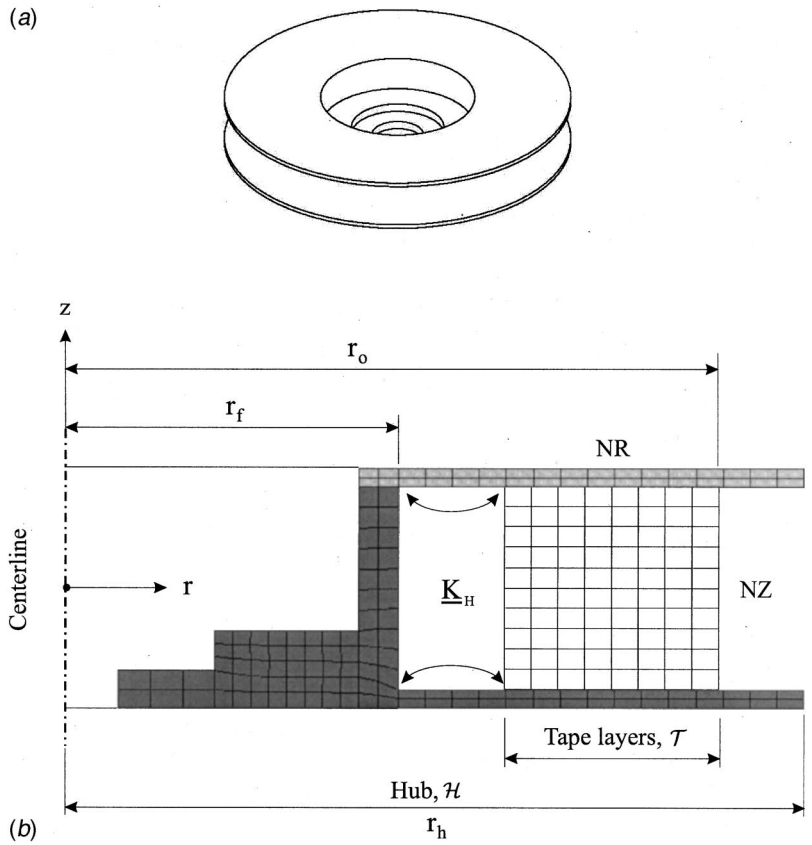


Fig. 3 (*a*) **An axisymmetric hub having no particular midplane symmetry, and** (*b*) **model of the hub and tape layer substructures**

located at a different radius than the hub's outer radius r_h owing to presence of the flanges. The substructures couple through the interfacial hub stiffness matrix $\mathbf{K}_H$ which affords a mixed boundary condition to the tape substructure. In order to effectively deal with the potential variety of hub materials and geometry, the hub itself is analyzed through a commercial finite element package, and $\mathbf{K}_H$ is extracted by sequentially applying unit loads along the hub-tape interface, and inverting the flexibility matrix so obtained.

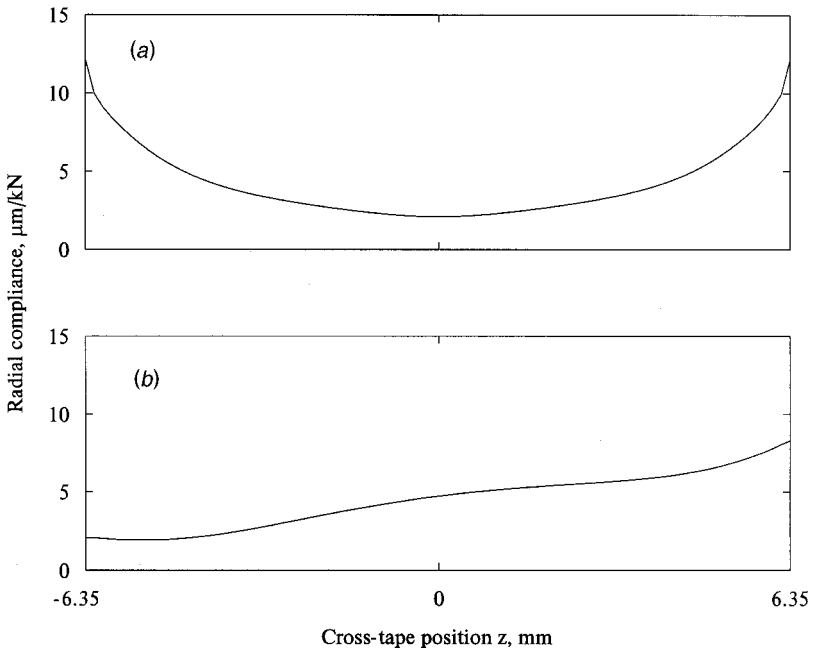


Fig. 4 Collocated point radial compliance of (*a*) **the symmetric hub of Fig. 2, and** (*b*) **the asymmetric hub of Fig. 3**

The stress field and displacements in $\mathcal{T}$ develop as each layer, a pre-tensioned cylindrical shell, is fit to the underlying pack. As the nth layer is added, for instance, the cumulative stresses $\boldsymbol{\sigma}_n$ within the pack are expressed $\boldsymbol{\sigma}_{n-1}+\Delta\boldsymbol{\sigma}_n$, where $\boldsymbol{\sigma}_{n-1}$ is the stress developed upon winding the first through $(n-1)$st layers, and $\Delta\boldsymbol{\sigma}_n$ is the incremental stress associated with addition of the final nth layer.

As the nth layer is added, $\mathcal{T}$ is discretized with bilinear axisymmetric elements. The governing equations comprise equilibrium, constitutive, and compatibility conditions, and they are functions of nodal displacements $\mathbf{a}=\{\mathbf{uw}\}^T$, where $\mathbf{u}(r,z)$ and $\mathbf{w}(r,z)$ are the radial and transverse components. The stiffness matrix in $\mathcal{T}$ is $\mathbf{K}_T=\Sigma_{i=1}^{NE}\mathbf{k}_i^e$, where $\mathbf{k}_i^e$ is the stiffness matrix for each element i, and NE is the number of elements in $\mathcal{T}$'s representation. The 8×8 elemental stiffness matrix becomes

$$\mathbf{k}_i^e=2\pi\int_{\mathcal{A}^e}\mathbf{B}_i^T\mathbf{E}_i\mathbf{B}_i dA^e, \tag{1}$$

where A^e is the element's $r-z$ cross-sectional area, $\mathbf{B}_i$ is the derivative of the strain-displacement relation, and $\mathbf{E}_i$ is the material's elasticity matrix. In turn, the structure-level stiffness matrix of the entire pack becomes $\mathbf{K}_P=\mathbf{K}_H+\mathbf{K}_T$, where proper assembly of the matrices accounting for the interfacial nodes at $r=r_h$ is implied. In addition, the nodal loads for the entire pack are evaluated as $\mathbf{F}=\Sigma_{i=1}^{NE}\mathbf{f}_i^e$ where the elemental nodal load is given by

$$\mathbf{f}_i^e=2\pi\int_{\delta A^e}\mathbf{N}_i^T\mathbf{t}_i r d\delta A^e+2\pi\int_{A^e}\mathbf{B}_i^T\mathbf{D}_i\boldsymbol{\epsilon}_{0i}d\mathcal{A}^e-2\pi\int_{A^e}\mathbf{B}_i^T\boldsymbol{\sigma}_{0i}d\mathcal{A}^e. \tag{2}$$

Here $\mathbf{N}_i$ is the bilinear shape function, and $\boldsymbol{\epsilon}_{0i}$ and $\boldsymbol{\sigma}_{0i}$ are the initial strain and stress for each element. The governing equations become $\mathbf{K}_P(\mathbf{a})\mathbf{a}=\mathbf{F}$ and are written in terms of the nodal displacements $\mathbf{a}$.

Solutions are subject to specified displacement and traction boundary conditions. For instance, the boundary conditions for a cartridge with the symmetric hub of Fig. 2 include vanishing transverse displacement at $(r_i,0)$ in order to suppress the rigid body motion. The general traction boundary conditions include

- vanishing traction at all hub surfaces excluding the positions with specified displacement boundary conditions,
- $\sigma_z=\sigma_{rz}=0$ over the upper and lower tape surfaces $z=\pm w/2$ and $r\in[r_f,r_o]$,
- $\sigma_{rz}=0$ and $\sigma_r=T(z)/(w(r_f+(n-1)h))$ over the outer tape surface $r=r_o$ and $z\in[-w/2,w/2]$

where r_i and w are the hub's inner radius and the tape's width, and the winding tension $T(z)$ is specified.

As the media's bulk radial modulus is known to depend strongly on stress, as each layer or group of layers is added to an existing pack, a truncated Taylor expansion is used to linearize the governing equations about either an initial estimate or a converged result obtained from calculation at the preceding state. Computation through Newton-Raphson iteration begins by evaluating $\mathbf{K}_P$ at an initial estimate $\mathbf{a}^*$. In the first iteration, the nodal displacements become $\mathbf{a}_1=\mathbf{K}_P^{-1}(\mathbf{a}^*)\mathbf{F}$. The vector of imbalanced nodal loads in the second iteration becomes $\Delta\mathbf{f}_2=\mathbf{F}-\mathbf{K}_P(\mathbf{a}_1)\mathbf{a}_1$. The incremental nodal displacements $\Delta\mathbf{a}$ in the second iteration are $\Delta\mathbf{a}_2=\mathbf{K}_P^{-1}(\mathbf{a}_1)\Delta\mathbf{f}_2$, and the cumulative displacements at that stage become $\mathbf{a}_2=\mathbf{a}_1+\Delta\mathbf{a}_2$. Iteration proceeds until the solution satisfies a specified convergence criterion expressed in terms of the norm $\eta^2=\Sigma\Delta a_j^2/\Sigma a_j^2$. When η falls below a specified tolerance, say 10^{-3} as in the case studies below, the solutions are said to have converged relatively, and iteration is terminated. With the nodal displacements so obtained, the stresses over element i are incremented by $\Delta\boldsymbol{\sigma}_{ni}=\mathbf{D}_i(\mathbf{B}_i\mathbf{a}_i-\boldsymbol{\epsilon}_{0i})+\boldsymbol{\sigma}_{0i}$, and the cumulative stresses advance to $\boldsymbol{\sigma}_{(n-1)}+\Delta\boldsymbol{\sigma}_n$.

Table 1 Baseline parameter values used in the pack winding case studies

Property	Symmetric	Asymmetric
	Hub	
Modulus, E	2.5 GPa	3.5 GPa
Poisson ratio, ν	0.43	0.43
Outer radius, r_f	11.43 mm	25.00 mm
	Tape	
Tension, T_0		1.0 N
Number of layers, NL		3000 —
Width, w		12.7 mm
Thickness, h		10.0 μm
Bulk radial modulus, E_r		$7000/(1+10.7\sigma_r^{-0.77})$ MPa
Circumferential modulus, E_θ		7 GPa
Transverse modulus, E_z		9 GPa
Shear modulus, G_{rz}		100 MPa
Poisson ratio, $\nu_{\theta z}=\nu_{\theta r}=\nu_{zr}$		0.3 —

3 Media Bulk Radial Modulus

Region $\mathcal{T}$ is well approximated as being orthotropic and axisymmetric, and so some four elastic moduli (E_r, E_θ, E_z, and G_{rz}) and six Poisson ratios ($\nu_{r\theta}$, ν_{rz}, $\nu_{\theta z}$, $\nu_{\theta r}$, ν_{zr}, and $\nu_{z\theta}$) must be specified to complete the formulation. The values for these material properties, exclusive of E_r, are developed following the discussion presented in reference [13], and they are listed in Table 1.

The media's substrate, magnetic and tribological coatings, and interfaces each contribute to E_r. The radial modulus was measured through compression tests conducted with a conventional materials testing machine. Each stack of experimental media had dimensions 102 mm×12.7 mm×12.7 mm. When such samples are extracted from an existing pack or larger format pancake, it is possible for irregular or distorted edges to be present, and such effects are minimized by having the compression plates be smaller in diameter than the samples' lengths. A 5 kN load cell was used to measure the applied force, and an extensometer with gage length 25.4 mm and 20 percent extension to measure the displacement across the sample. During manufacturing, a wide web of magnetic media is slit into many individual streams of tape, which in turn are wound on large format hubs so as to store the media temporarily prior to forming data cartridges. The diameters of such pancakes typically vary from 200–300 mm, and the samples used in these studies were cut from such pancakes.

For a particular experimental media, Fig. 5 illustrates a representative stress-strain response over five successive tests. The behavior during the first load cycle differs somewhat from the subsequent ones as most of the air entrapped between adjacent layers is expelled during that first load-unload cycle. The experimental data are fit to a polynomial through least squares regression, and the bulk modulus is determined by differentiating the stress-strain expression. Figure 6 depicts measured response of the specimen over a full load-unload cycle, and the behavior is clearly path-dependent. The area enclosed by the loading and unloading curves represents the energy dissipated during the process as caused by the internal material damping and interfacial friction, among other factors. Such path-dependent behavior can be used to simulate the full winding and unwinding process.

In order to examine potential variability in E_r for media samples extracted from a single pancake, and namely from the same tape stream and widthwise position on the manufacturing web, samples were removed from the pancake from three different radii, each at a different circumferential location. Because the pancake is substantially larger than a typical data cartridge and the stress field within it varies substantially with position, each sample was presumably produced and stored under a different stress history. It is an objective here to assess variability in E_r associated with the production process, quite aside from the wound roll model itself. Measured moduli for the nine media

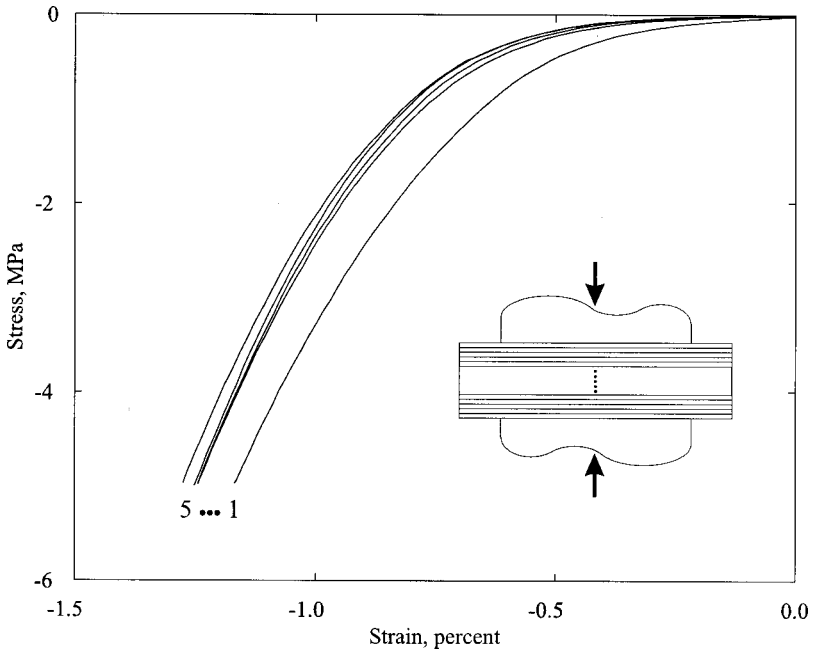


Fig. 5 Measured stress-strain response in compression of a magnetic tape stack over five loading cycles; sample dimensions: 102 mm×12.7 mm×12.7 mm

samples are shown in Fig. 7(*a*), where some variability among the samples depending on radial location within the pancake is evident. At a fixed value of stress, say 2 MPa, E_r varies between roughly 800 and 1200 MPa, or ± 20 percent about its mean value. The three samples taken from the outer periphery of the pancake consistently exhibited lower E_r than those extracted from the inner periphery, consistent with the notion of the media undergoing slight strain hardening as asperities are plastically deformed during the pancake's formation.

In considering a media stack under uniform compression, its bulk deformation arises both from compression of the substrate layers and compliance of the surfaces. Considering N layers, there are $N-1$ interfaces formed by adjacent contact of the magnetic surface on one layer, and the backcoat surface on another. Both the tape layers and the interfaces contribute to the macroscopic stiffness, and this assemblage is modeled accordingly as a series of elastic springs. Specifying that the substrate deforms linearly during compression, the elastic constant contributed by the tape's

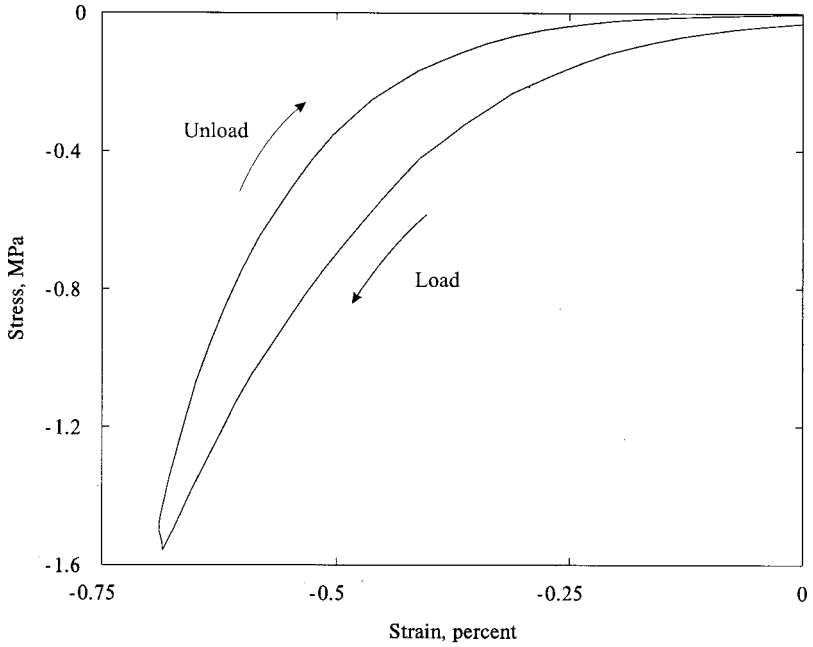


Fig. 6 Typical measured stress-strain response of a magnetic tape stack over a single load–unload cycle; sample dimensions: 102 mm×12.7 mm×12.7 mm

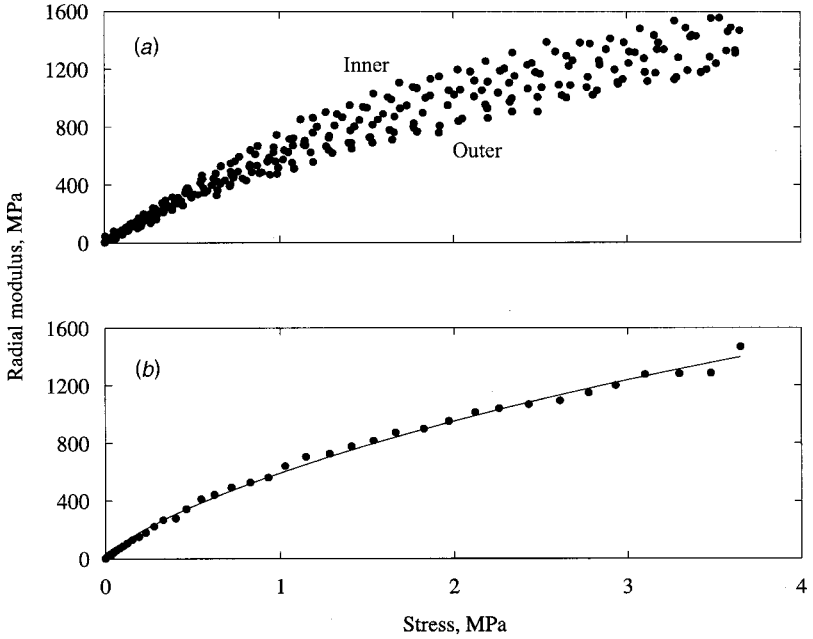


Fig. 7 (*a*) Variability of the measured bulk radial modulus for nine nominally identical media samples. (*b*) Measurements averaged over all samples (● ● ● ●), and the least-squares fit to the modulus model (——); $c=1.61\times10^{7}$ N/mm$^{4.26}$ and $m=3.26$.

substrate for each layer is $k_s=EA/h$, where E and A are the elastic modulus and apparent contact area for a single layer. The real area of contact, in turn, increases with the compressive load ([18]). The interfacial stiffness changes in response to its deformation with constant $k_i=(c\Delta_i)^m$, where c and m are constants determined subsequently through a fit to the data. For pure Hertzian contact of similar materials, for instance, $c=16E^2\beta/9$ and $m=1/2$, where β is the radius of an individual asperity's summit. Extension to rough but nominally flat surfaces in contact is discussed by [18].

Under compressive load P, the stack's total deformation becomes $\Delta=N\Delta_s+(N-1)\Delta_i\approx N(\Delta_s+\Delta_i)\approx N(P/k_s+P/k_i)$ for large N. By definition, the strain in the media's stack is $\epsilon=\Delta/(Nh)=(P/k_s+P/k_i)/h=P/(E_rA)$. Following algebraic manipulation, the bulk radial modulus is expressed

$$E_r=E\left(1+\frac{EA}{h}\frac{1}{\sqrt[m+1]{c^mP^m}}\right)^{-1} \tag{3}$$

as a function of applied stress. The unknown parameters c and m are determined by fitting the experimental data through a least square method. Figure 7(*b*) demonstrates the manner in which Eq. (3) captures the measured data.

4 Stress Field With Symmetric Hub Design

Figures 8 and 9 depict the predicted radial and circumferential stress distributions as functions of r and z for the hub design of Fig. 2 and parameter values as given in Table 1. Pack formation was simulated with a winding tension of 1 N having uniform stress 7.87 MPa over the tape's cross section. In Fig. 8, σ_r increases from zero at the pack's outer periphery to reach its maximum compressive value of -2.45 MPa along the hub-tape interface at the location $(r,z)=(11.43,0)$ mm of the hub's central reinforcement rib. The radial stress is less compressive near the tape's edges where the hub has greater compliance. For the first layer on the hub, σ_r varies from -1.48 MPa at the tape's edges to -2.45 MPa at the centerline, an increase of some 70 percent. Such cross-width variation of in-pack stresses is significant in Figs. 8–9 over only about one-third of the pack, as indicated by the contour diagrams shown as insets in the figures. With respect to the circumferential stress, the maximum value occurs at r_f on the tape's edges at the position $(11.43,\pm6.35)$ mm. Further, σ_θ varies from -7.73 MPa at the edges to -1.21 MPa at the centerline for the pack's first layer. Notably, σ_θ is compressive at the pack's innermost radial positions, grows to become tensile in the outer layers, and precisely equals the winding stress at the outermost layer. The bold demarcation line in the contour diagram inset in Fig. 9 indicates the loci of points where $\sigma_\theta=0$, in order to aid in understanding the locations of such pack buckling defects as seen in Fig. 1. In this simulation, the transverse σ_z and shear σ_{rz} stresses are significant only near the hub-tape interface and arise from the strain mismatch between the different materials forming $\mathcal{H}$ and $\mathcal{T}$. Since the edges of the pack model are traction free and coupling ν_{rz} between the r and z directions is almost zero, σ_z and σ_{rz} are likewise negligible away from r_f.

4.1 Cross-Track Media Width Change. The in-pack stresses in turn cause the tape's width to change slightly, and as the pack would be subsequently unwound, those dimensional changes would be reflected as variations in the spacing of data tracks. Figure 10 shows the predicted tape width change Δw as a function of position along the tape's length based on the stress fields of Figs. 8–9. To the extent that the viscoelastic relaxation time of the media is long when compared to the time required for the pack to unwind, Δw would be measured and compensated by the read-write head's servo system. The width change is calculated by subtracting the tape's transverse displacements at the upper and lower edges, and the difference increases gradually from the outer periphery with a rapid increase near $\mathcal{H}$. The maximum value reached in this case study is about 14 μm or 1100 ppm, which is in fact greater than the roughly 5 μm data track width on a modern drive. As shown in Fig. 10, more than 35 percent of the pack is subjected to a width change greater than a single track width.

4.2 Cross-Tape Tension Gradient. Figures 11 and 12 illustrate the radial and circumferential stress fields as functions of r and z for a pack wound under tension that varies linearly across

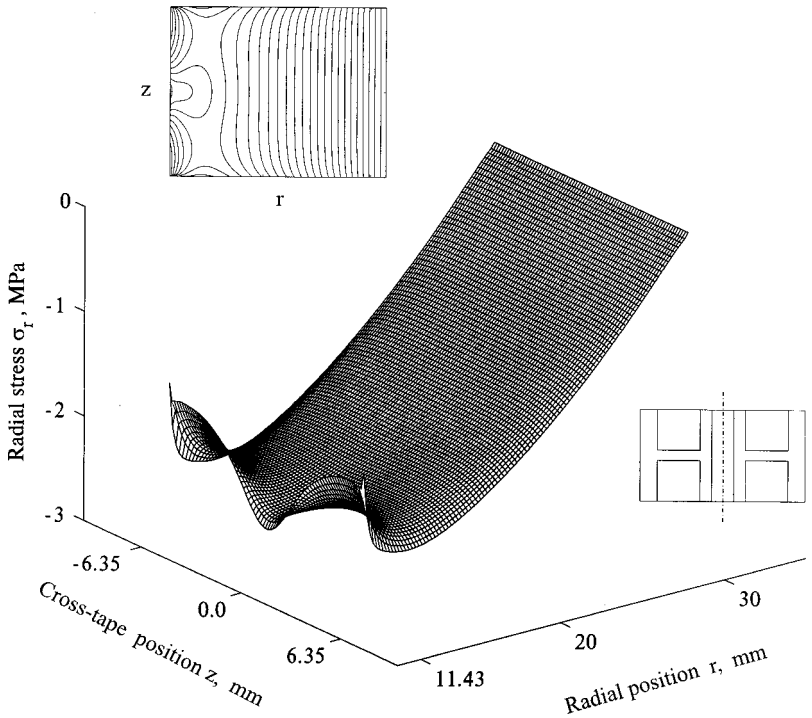


Fig. 8 Radial stress field in the symmetric hub case study. The insets depict the hub's cross section and a contour representation of σ_r over the $r-z$ plane; constant tension, $NR=100$, $NZ=80$.

the tape's width. Such a situation arises when the tape bends in the transport path either as a result of guide misalignment or tape lateral motion. The tension profile is approximated by a quadratic function as $T(z)=c_2z^2+c_1z+c_0$, where c_2, c_1, and c_0 are set by specified values at the tape's edges and centerline.

In Figs. 11 and 12, the tension is specified to vary from $0.75\,T_0$ at the bottom edge $z=-6.35$ mm to $1.25\,T_0$ at the top edge $z=6.35$, where T_0 is the nominal tension. The resulting tension profile becomes $T(z)=(z/2w+1)T_0$ over $z=[-w/2,w/2]$. Cross-tape variations of σ_r and σ_θ are more prominent in Figs. 11

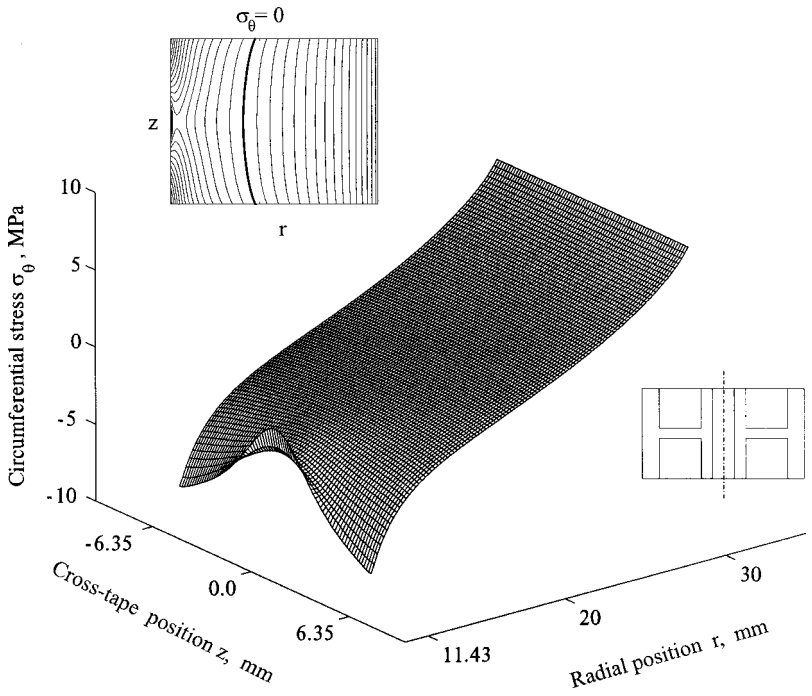


Fig. 9 Circumferential stress field in the symmetric hub case study. The insets depict the hub's cross section and a contour representation of σ_θ in the $r-z$ plane; constant tension, $NR=100$, $NZ=80$

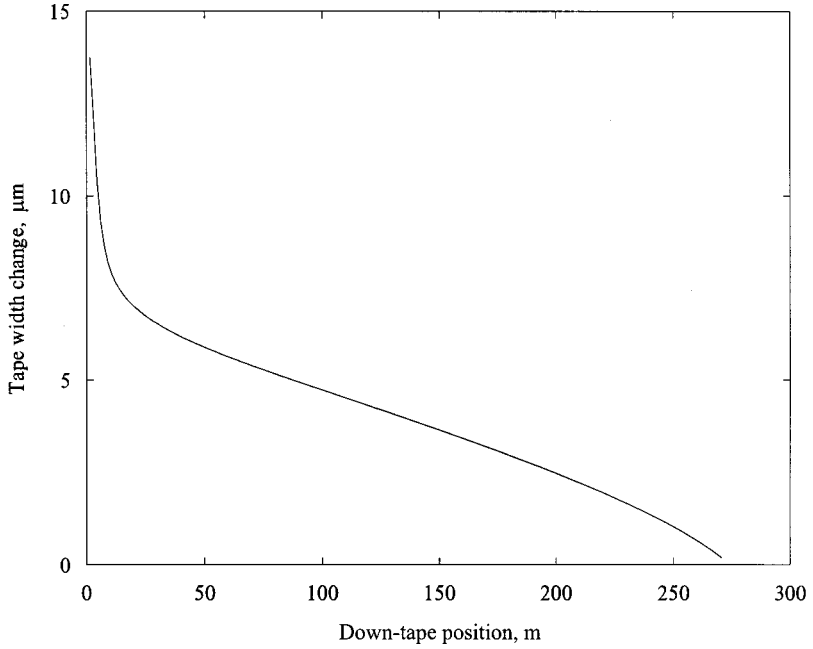


Fig. 10 Predicted change in tape width. The maximum value at the hub-tape interface is about 14 μm, or 1100 ppm.

and 12 than for the case of a uniform tension profile in Figs 8 and 9. Along the tape's higher tension edge, both the radial and circumferential stresses in Figs. 11 and 12 have higher magnitudes. As was the case in Figs. 8 and 9, the maximal compressive radial stress occurs at the location of the central reinforcement rib is located. With respect to the extreme values of σ_r at the first layer, the ratio of the maximum and minimum values is 2.4:1, somewhat larger than 1.7:1 as in the uniform tension case study. Component σ_θ has a more uneven distribution, and the ratio of the extreme values here is 9.4:1, with the maximum compressive value of -11.1 MPa occurring at (11.43,6.35) mm.

In some circumstances, the winding tension is known to roll-off at both edges of the tape in a manner well approximated as $T(z)=(-(z/w)^2+1)T_0$. The radial and circumferential stresses

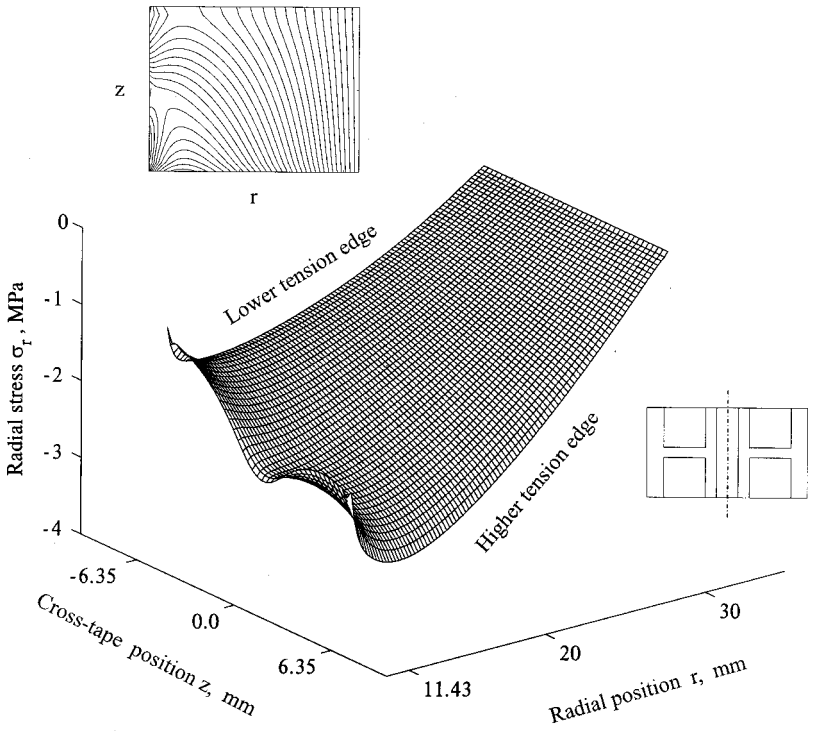


Fig. 11 Radial stress field in the symmetric hub case study. The insets depict the hub's cross section and a contour representation of σ_r in the $r-z$ plane; linear cross-tape tension gradient, $NR=100$, $NZ=80$.

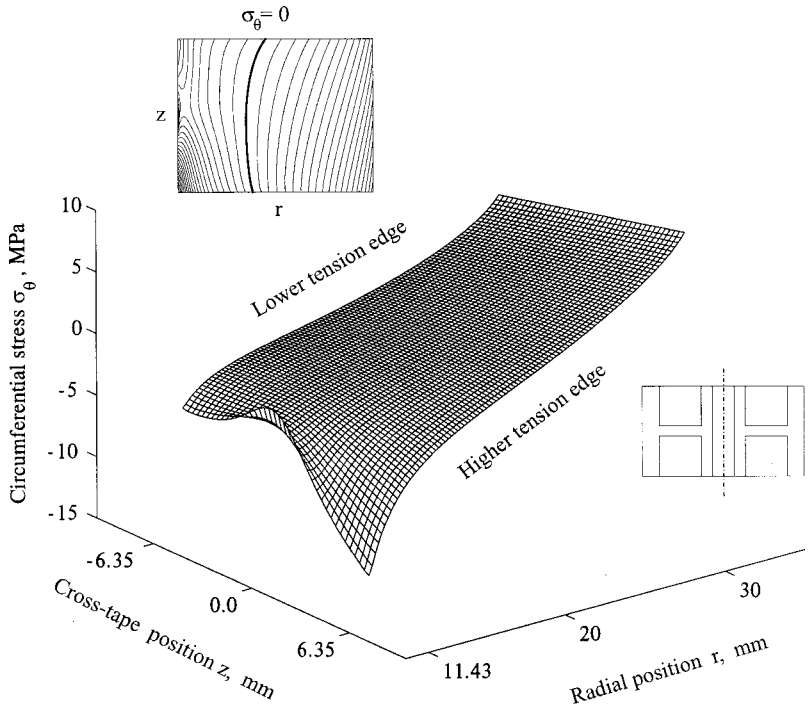


Fig. 12 Circumferential stress field in the symmetric hub case study. The insets depict the hub's cross section and a contour representation of σ_θ in the $r-z$ plane; linear cross-tape tension gradient, $NR=100$, $NZ=80$.

resulting from this tension profile are shown in Figs. 13 and 14. Here σ_r and σ_θ are less compressive than in the case of uniform tension, but they do vary to a greater extent across the tape's width. Here the ratios of the extreme values at the first layer for σ_r and σ_θ are 2.2:1 and 6.2:1, respectively.

5 Stress Field With Asymmetric Hub Design

In the case of the asymmetric hub format of Fig. 3, the in-pack stress field is expected to be singular at the intersecting corners of the hub's winding face and the flanges, to the extent that those

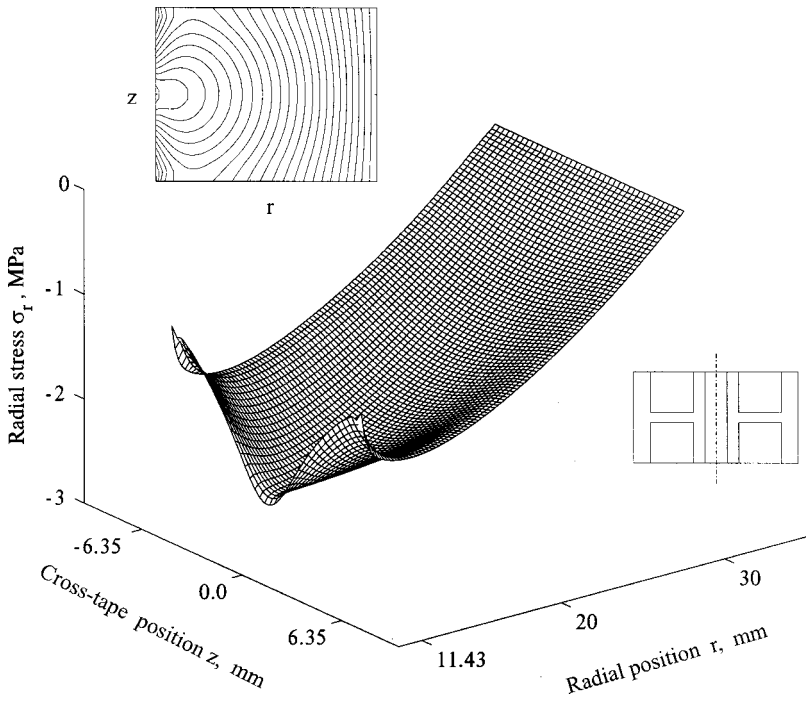


Fig. 13 Radial stress field in the symmetric hub case study. The insets depict the hub's cross section and a contour representation of σ_r in the $r-z$ plane; parabolic cross-tape tension gradient, $NR=100$, $NZ=80$.

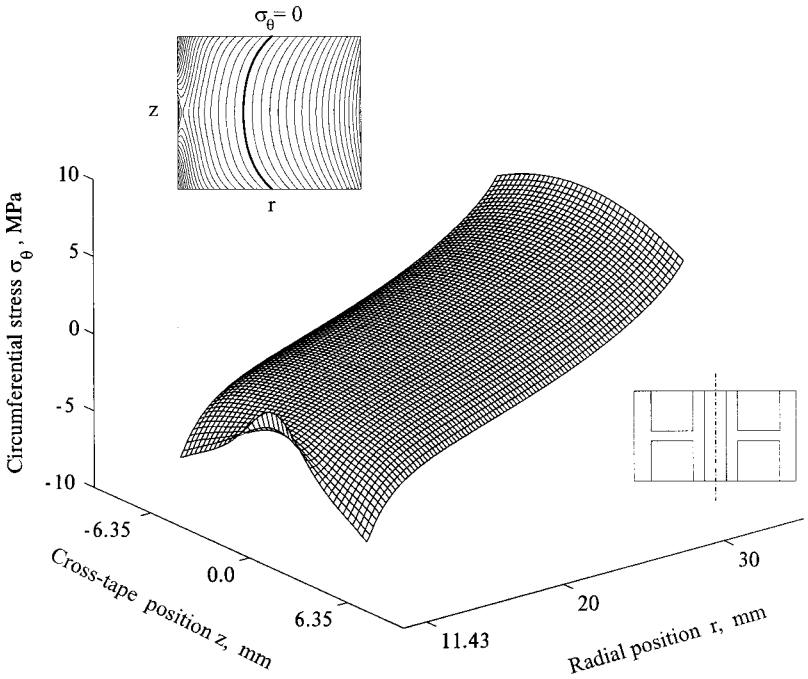


Fig. 14 Circumferential stress field in the symmetric hub case study. The insets depict the hub's cross section and a contour representation of σ_θ in the $r-z$ plane; parabolic cross-tape tension gradient, $NR=100$, $NZ=80$.

components are formed of dissimilar materials ([13,19]). Stress behavior at those regions was investigated through a mesh refinement convergence study. Figure 15(a) shows the predicted radial stresses along the pack's centerline and edges as functions of radial position for $NZ=40$ and 80. Stress σ_r converges well along the centerline and throughout most of the pack, except for the narrow regions near the hub-tape interface. Figure 15(*b*) shows the distribution of σ_r in the first layer as a function of cross-tape position. The radial stress converges over some 90 percent of the cross section at $NZ=80$ except for regions adjacent to the corners. The domain of convergence expands with increasing NZ as the singularity further localizes.

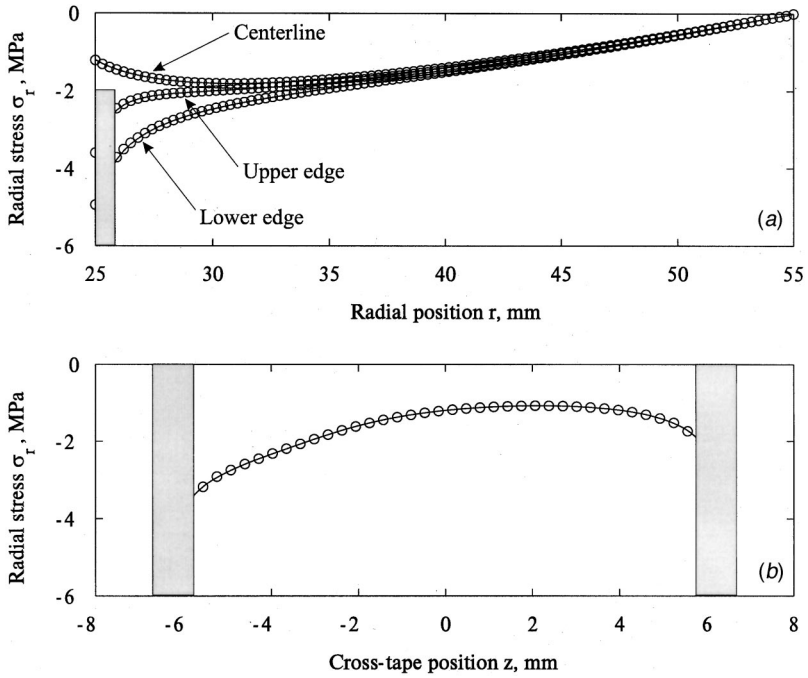


Fig. 15 Radial stress distribution (*a*) in the down-tape direction at three positions across the width, and (*b*) across the tape's width at the hub-tape interface; $NZ=40$ (○ ○ ○ ○) and $NZ=80$ (——). Shaded zones indicate where the solution did not converge to three significant figures.

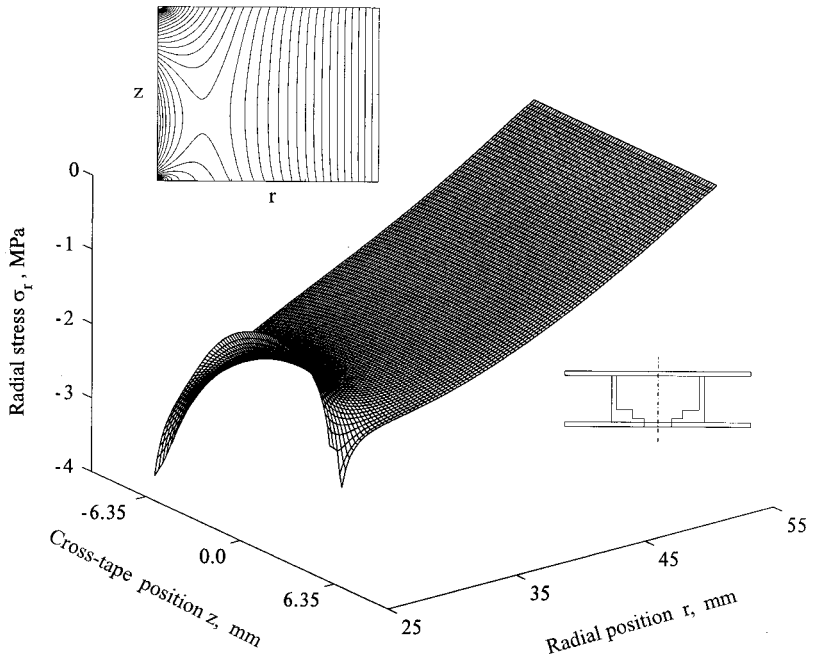


Fig. 16 Radial stress field in the asymmetric hub case study. The insets depict the hub's cross-section and a contour representation of σ_r in the $r-z$ plane; constant tension, $NR=100$, $NZ=80$.

Figures 16 and 17 show the radial and circumferential stresses as functions of r and z only in the domain over which the solution has fully converged. The singular behavior at the corners strongly affects σ_r over the first several layers of tape, and it exhibits a sudden increase in compression near the corners. Even though the upper portion of the hub is comparatively compliant, the high gradients dominate the σ_r distribution. Likewise, stress σ_θ varies significantly along the hub-tape interface from 2.30 MPa at the pack's bottom edge to -8.97 MPa at the top edge. The bold demarcation line in the inset contour diagram of Fig. 17 represents the loci of points where $\sigma_\theta=0$. It is interesting to note that the pack's bottom edge is not in circumferential compression, evidently suggesting that it is more stable from the defect formation perspective than regions near the top surface.

6 Summary

A finite width model for predicting the stresses and displacements within a magnetic tape cartridge has been discussed. Such widthwise variations as differential hub stiffness and nonuniform

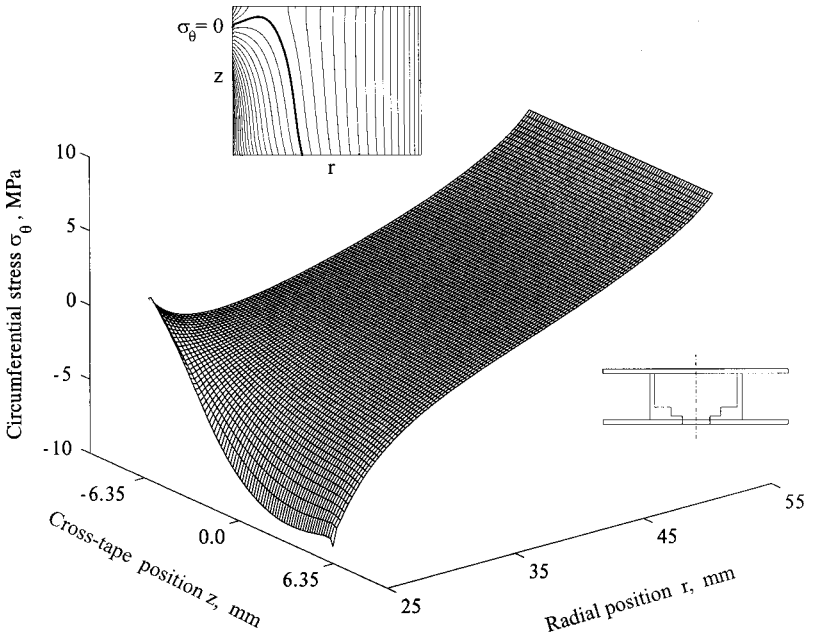


Fig. 17 Circumferential stress field in the asymmetric hub case study. The insets depict the hub's cross section and a contour representation of σ_θ in the $r-z$ plane; constant tension, $NR=100$, $NZ=80$.

winding tension are explicitly treated in the formulation. The solution is developed through finite element and substructuring methods in order to handle a potential variety of hub geometries and materials. The cross-tape tension gradient as is caused by guide misalignment, scatterwinding, and tension roll-off is also modeled with a view towards understanding the in-pack stress field as a precursor to defect formation.

Measurements of the media's bulk radial modulus were performed through compression tests and demonstrate that such values do depend on the specimen's stress history and location within the manufacturing pancake. The heuristic reduced-order model for E_r was shown to represent the measured data, and aside from polynomial curve fits as have been used previously, provides a meaningful bound with increasing σ_r. In case studies with two typical hub designs, the simulations quantify the roles of differential hub compliance, hub design, and winding tension gradients in setting the pack's internal stress distribution.

Acknowledgment

This work was supported in part by Imation Corporation through the Advanced Tape Systems Program, and by the National Science Foundation. The authors appreciate the assistance and comments of Durkee Richards, William Qualls, and Chris Merton of Imation Corporation, Bill Buchan of Seagate Technology, and Ted Schwarz of Peregrine Recording Technology.

References

[1] Altmann, H. C., 1968, ''Formulas for Computing the Stresses in Center-Wound Rolls,'' Journal of the Technical Association of the Paper and Pulp Industry, **51**, pp. 176–179.
[2] Yagoda, H. P., 1980, ''Resolution of a Core Problem in Wound Rolls,'' ASME J. Appl. Mech., **47**, pp. 847–854.
[3] Connolly, D., and Winarski, D. J., 1984, ''Stress Analysis of Wound Magnetic Tape,'' ASLE Tribology and Mechanics of Magnetic Storage Media, Special Publication 16, ASLE, pp. 172–182.
[4] Tramposch, H., 1965, ''Relaxation of Internal Forces in a Wound Reel of Magnetic Tape,'' ASME J. Appl. Mech., **32**, pp. 865–873.
[5] Tramposch, H., 1967, ''Anisotropic Relaxation of Internal Forces in a Wound Reel of Magnetic Tape,'' ASME J. Appl. Mech., **34**, pp. 888–894.
[6] Heinrich, J. C., Connolly, D., and Bhushan, B., 1986, ''Axisymmetric, Finite Element Analysis of Stress Relaxation in Wound Magnetic Tapes,'' ASLE Trans., **29**, pp. 75–84.
[7] Hakiel, Z., 1987, ''Nonlinear Model for Wound Roll Stresses,'' Journal Tech. Assoc. Paper Pulp Indust., **70**, pp. 113–117.
[8] Willett, M. S., and Poesch, W. L., 1988, ''Determining the Stress Distributions in Wound Reels of Magnetic Tape Using a Nonlinear Finite-Difference Approach,'' ASME J. Appl. Mech., **55**, pp. 365–371.
[9] Zabaras, N., Liu, S., Koppuzha, J., and Donaldson, E., 1994, ''A Hypoelastic Model for Computing the Stresses in Center-Wound Rolls of Magnetic Tape,'' ASME J. Appl. Mech., **61**, pp. 290–295.
[10] Good, J. K., Pfeiffer, J. D., and Giachetto, R. M., 1992, ''Losses in Wound-on Tension in the Centerwinding of Wound Rolls,'' *Web Handling*, ASME AMD-149, ASME, New York, pp. 1–12.
[11] Benson, R. C., 1995, ''A Nonlinear Wound Roll Model Allowing for Large Deformation,'' ASME J. Appl. Mech., **62**, pp. 853–859.
[12] Qualls, W. R., and Good, J. K., 1997, ''An Orthotropic Viscoelastic Winding Model Including a Nonlinear Radial Stiffness,'' ASME J. Appl. Mech., **64**, pp. 201–208.
[13] Lee, Y. M., and Wickert, J. A., 2002, ''Stress Field in Finite Width Axisymmetric Wound Rolls,'' ASME J. Appl. Mech., submitted for publication.
[14] Bhushan, B., 1992, *Mechanics and Reliability of Flexible Magnetic Media*, Springer-Verlag, New York.
[15] Hakiel, Z., 1992, ''On the Effect of Width Direction Thickness Variations in Wound Rolls,'' *Second International Conference on Web Handling*, Oklahoma State University, pp. 79–98.
[16] Kedl, D. M., 1992, ''Using a Two Dimensional Winding Model to Predict Wound Roll Stresses that Occur due to Circumferential Steps in Core Diameter or to Cross-Web Caliper Variation,'' *Second International Conference on Web Handling*, Oklahoma State University, pp. 99–112.
[17] Cole, A., and Hakiel, Z., 1992, ''A Nonlinear Wound Roll Stress Model Accounting for Widthwise Web Thickness Nonuniformities,'' *Web Handling*, ASME AMD-149, ASME, New York, pp. 13–24.
[18] Greenwood, J. A., and Williamson, J. B. P., 1966, ''Contact of Nominally Flat Surfaces,'' Proc. R. Soc. London, Ser. A, **A295**, pp. 300–319.
[19] Bogy, D. B., 1970, ''On the Problem of Edge-Bonded Elastic Quarter-Planes Loaded at the Boundary,'' Int. J. Solids Struct., **6**, pp. 1287–1313.

Finite Element Investigation of Quasi-Static Crack Growth in Functionally Graded Materials Using a Novel Cohesive Zone Fracture Model

Z.-H. Jin
Mem. ASME

G. H. Paulino
Mem. ASME

R. H. Dodds, Jr.
Mem. ASME

Department of Civil and Environmental Engineering,
University of Illinois at Urbana-Champaign,
Newmark Laboratory, MC-250,
205 North Mathews Avenue
Urbana, IL 61801

This work studies mode I crack growth in ceramic/metal functionally graded materials (FGMs) using three-dimensional interface-cohesive elements based upon a new phenomenological cohesive fracture model. The local separation energies and peak tractions for the metal and ceramic constituents govern the cohesive fracture process. The model formulation introduces two cohesive gradation parameters to control the transition of fracture behavior between the constituents. Numerical values of volume fractions for the constituents specified at nodes of the finite element model set the spatial gradation of material properties with standard isoparametric interpolations inside interface elements and background solid elements to define pointwise material property values. The paper describes applications of the cohesive fracture model and computational scheme to analyze crack growth in compact tension, C(T), and single-edge notch bend, SE(B), specimens with material properties characteristic of a TiB/Ti FGM. Young's modulus and Poisson's ratio of the background solid material are determined using a self-consistent method (the background material remains linear elastic). The numerical studies demonstrate that the load to cause crack extension in the FGM compares to that for the metal and that crack growth response varies strongly with values of the cohesive gradation parameter for the metal. These results suggest the potential to calibrate the value of this parameter by matching the predicted and measured crack growth response in standard fracture mechanics specimens. [DOI: 10.1115/1.1467092]

1 Introduction

Functionally graded materials (FGMs) provide promising candidates for advanced technological applications ([1–3]). An FGM comprises a multiphase material with volume fractions of the constituent materials varying in a pre-determined profile, thus giving a nonuniform microstructure in the material with continuously graded properties. In applications involving severe thermal gradients (e.g., thermal protection structures), FGM systems exploit the heat, oxidation, and corrosion resistance typical of ceramics, and the strength and toughness typical of metals.

Cohesive fracture models have been widely used to simulate and analyze crack growth in ductile and quasi-brittle materials. In a cohesive fracture model, a narrow band termed a cohesive zone, or process zone, exists ahead of the crack front. Material behavior in the cohesive zone follows a cohesive constitutive law which relates the cohesive traction to the relative displacements of the adjacent surfaces. Crack growth occurs by progressive decohesion of the cohesive surfaces. Dugdale [4] first proposed a cohesive-type model to study ductile fracture in a thin sheet of mild steel. The Dugdale model assumes that a cracked metal sheet deforms elastically outside of the extended surfaces of the crack where a narrow band (plastic zone) of idealized zero width deforms at the constant yield stress of the material. Cohesive fracture models have been extended to study fracture processes in quasi-brittle materials such as concrete (see, e.g., [5,6]), ductile metals (see, e.g., [7,8]), and metal matrix composites ([9]).

Though cohesive fracture models have been successfully employed to simulate failure processes in homogeneous materials and conventional composites, few studies have extended the concept to FGMs. The difficulty lies in the coexistence of different failure mechanisms in an FGM as explained in the next section. Studies of crack growth through the whole FGM component require a new phenomenological model to simulate the fracture process. Jin and Batra [10] studied crack growth in the ceramic-rich region in a ceramic/metal FGM by using both a rule of mixtures and a crack bridging model (essentially a cohesive-type model). Cai and Bao [11] investigated crack growth in a ceramic/metal graded coating by using a similar, but simpler crack bridging model. Simple applications of the rule of mixtures to an FGM significantly overestimate the fracture toughness compared to estimates from crack bridging models ([10]). Thus, it appears inappropriate to employ directly the conventional rule of mixtures to formulate the cohesive parameters of FGMs. The modifications described here provide a more realistic approach to formulate a cohesive model suitable for FGMs.

This work studies crack growth in ceramic/metal FGMs using three-dimensional interface-cohesive elements. While we are not considering the ductile deformation in the graded background material, the current study focuses on presentation of the cohesive zone model and does incorporate the ductile separation of the graded cohesive material in the analysis of crack growth. Investigations of crack growth in ceramic/metal FGMs considering plasticity in the background material are in progress. The paper is organized as follows. Section 2 proposes a new phenomenological, cohesive fracture model developed specifically for ceramic/

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, June 18, 2001; final revision, Oct. 15, 2001. Associate Editor: A. Needleman. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

metal FGMs. In addition to the cohesive energy densities and the peak cohesive tractions of the metal and ceramic phases, two other parameters are introduced to account for the overall damage and other micromechanical effects in FGMs. Section 3 describes the three-dimensional finite element formulation with graded solid and interface-cohesive elements for applications to FGMs. Section 4 discusses the method of determining the material parameters of FGMs. Section 5 presents results of a parametric study of crack growth analyses for a titanium/titanium monoboride (Ti/TiB) FGM. Compact tension, C(T), and single-edge notched bend, SE(B), specimens are considered in the numerical simulations. Section 6 provides some conclusions and outlines ongoing work to extend the present study. The Appendix summarizes details of the tangent modulus matrix for the cohesive constitutive relation applicable to FGMs.

2 A Novel Cohesive Fracture Model

Generalization of the cohesive zone concept to model fracture in functionally graded materials (FGMs) represents a challenging task in view of the different failure mechanisms present in an FGM. In a typical ceramic/metal FGM, the ceramic-rich region may be regarded as a metal particle reinforced ceramic matrix composite, whereas the metal-rich region may be treated as a ceramic particle-reinforced metal matrix composite. Though models for the failure mechanisms of conventional composites may be adopted to study the fracture processes in the ceramic-rich or metal-rich region, the failure mechanisms operative in the interconnecting region which has no distinct matrix and inclusion phases remain unknown. This section thus proposes a volume fraction-based phenomenological cohesive fracture model suitable for engineering scale applications. The formulation first considers tensile mode (mode I) fracture of FGMs, and is then extended to general three-dimensional fracture including both tensile and shear deformations. Such volume fraction-based formulas have been used previously to calculate Young's modulus and the plastic tangent modulus of FGMs ([12,13]).

2.1 Mode I Fracture. Let σ_{fgm} denote the normal traction across the surfaces of the cohesive zone necessary to model the propagation of a macroscale crack. We propose that the cohesive traction σ of a two-phase FGM (e.g., ceramic/metal FGM) can be approximated by the following volume fraction-based formula having a simple functional form

$$\sigma_{\mathrm{fgm}}(\mathbf{x}) = \frac{V_{\mathrm{met}}(\mathbf{x})}{V_{\mathrm{met}}(\mathbf{x}) + \beta_{\mathrm{met}}[1 - V_{\mathrm{met}}(\mathbf{x})]}\,\sigma_{\mathrm{met}} + \frac{1 - V_{\mathrm{met}}(\mathbf{x})}{1 - V_{\mathrm{met}}(\mathbf{x}) + \beta_{\mathrm{cer}} V_{\mathrm{met}}(\mathbf{x})}\,\sigma_{\mathrm{cer}}, \tag{1}$$

where σ_{met} is the cohesive traction of the metal, σ_{cer} the cohesive traction of the ceramic, $V_{\mathrm{met}}(\mathbf{x})$ denotes the volume fraction of the metal, $\mathbf{x}=(x_1,x_2,x_3)$, and $\beta_{\mathrm{met}}(\geqslant 1)$ and $\beta_{\mathrm{cer}}(\geqslant 1)$ are two cohesive gradation parameters. The motivation to choose (1) is that the cohesive traction of the FGM will reduce to that of the metal when $V_{\mathrm{met}}=1$, and to that of the ceramic when $V_{\mathrm{met}}=0$, and the two parameters β_{met} and β_{cer}, together with the metal volume fraction (V_{met}), could describe the transition of the failure mechanism from pure ceramic to pure metal (operative in the interconnecting region which has no distinct matrix and inclusion phases).

The FGM cohesive fracture model, Eq. (1), increases the number of material-dependent parameters by two ($\beta_{\mathrm{met}},\beta_{\mathrm{cer}}$). Values for the local separation energies and peak cohesive tractions related to the pure ductile and brittle phases are obtained using standard procedures for homogeneous materials (see [8], for example). The material-dependent parameters β_{met} and β_{cer} describe approximately the overall effect of cohesive traction reduction (from the level predicted by the rule of mixtures) and the transition between the fracture mechanisms of the metal and ceramic phases. Our preliminary computations of crack growth in a TiB/Ti FGM indicate that β_{met} plays a far more significant role than β_{cer}, which can be simply set to unity. We anticipate that the parameter β_{met} may be experimentally calibrated by two different procedures. The first procedure determines β_{met} by matching the predicted and measured crack growth responses in standard fracture mechanics specimens of FGMs. Instead of using FGM specimens, the second procedure employs fracture specimens made of a monolithic composite each with a fixed volume fraction of the constitutents. This opens the potential to calibrate β_{met} for each volume fraction level of metal and ceramic, which comprise the FGM specimens, i.e., β_{met} can become a function of V_{met} in the present model. The second calibration procedure may be particularly useful if a constant β_{met} fails to generate a match between the predicted and experimentally measured crack growth responses. Experimental determination of the β_{met} parameter is presently under investigation for zirconia/stainless steel FGMs.

For the metal phase, the cohesive traction may be derived from a free-energy density function, $\phi_{\mathrm{met}}(\delta,q)$, in the form ([8,14–16])

$$\sigma_{\mathrm{met}} = \frac{\partial \phi_{\mathrm{met}}}{\partial \delta}, \tag{2}$$

where δ is the normal displacement jump across the cohesive surfaces and q is an internal variable describing the irreversible processes of decohesion. Because in general, the shape of the cohesive traction-separation curve ($\sigma-\delta$) is not as significant as the cohesive energy density and the maximum cohesive traction in simulating fracture in ductile metals ([17]), the free-energy potential, $\phi_{\mathrm{met}}(\delta,q)$, may be chosen in a computationally convenient exponential form ([8,14–16])

$$\phi_{\mathrm{met}} = e\,\sigma^c_{\mathrm{met}}\delta^c_{\mathrm{met}}\left[1 - \left(1 + \frac{\delta}{\delta^c_{\mathrm{met}}}\right)\exp\left(-\frac{\delta}{\delta^c_{\mathrm{met}}}\right)\right]. \tag{3}$$

Under loading conditions governed by q, the cohesive traction of the metal with the above energy potential is given by

$$\sigma_{\mathrm{met}} = e\,\sigma^c_{\mathrm{met}}\left(\frac{\delta}{\delta^c_{\mathrm{met}}}\right)\exp\left(-\frac{\delta}{\delta^c_{\mathrm{met}}}\right), \tag{4}$$

where $e=\exp(1)$, σ^c_{met} the maximum cohesive traction, and δ^c_{met} the value of δ at $\sigma_{\mathrm{met}}=\sigma^c_{\mathrm{met}}$. Figure 1(*a*) shows a typical curve for $\sigma_{\mathrm{met}}/\sigma^c_{\mathrm{met}}$ versus $\delta/\delta^c_{\mathrm{met}}$.

For quasi-brittle materials such as concrete and ceramics, the shape of the cohesive traction-separation curve may play a significant role in determining the peak load ([18]). In the present study of ceramic/metal FGMs, however, the failure mechanism of metal phase plays a dominant role. Thus, for simplicity, this study adopts the same exponential form as Eq. (4) to describe the cohesive response of the ceramic material

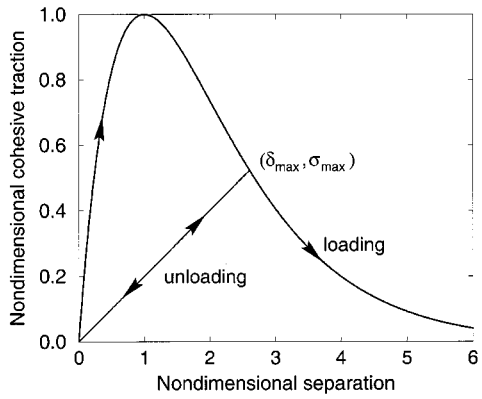
$$\sigma_{\mathrm{cer}} = e\,\sigma^c_{\mathrm{cer}}\left(\frac{\delta}{\delta^c_{\mathrm{cer}}}\right)\exp\left(-\frac{\delta}{\delta^c_{\mathrm{cer}}}\right), \tag{5}$$

where σ^c_{cer} is the maximum cohesive traction of the ceramic occurring at $\delta=\delta^c_{\mathrm{cer}}$. The free-energy potential corresponding to Eq. (5) is

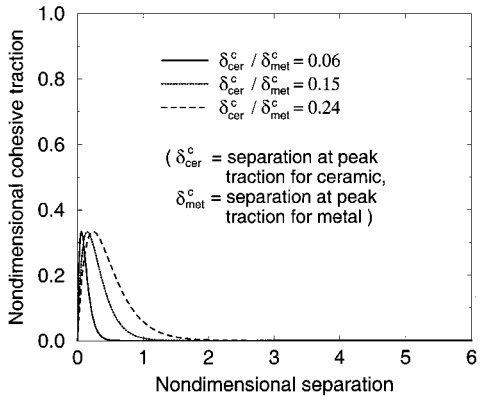
$$\phi_{\mathrm{cer}} = e\,\sigma^c_{\mathrm{cer}}\delta^c_{\mathrm{cer}}\left[1 - \left(1 + \frac{\delta}{\delta^c_{\mathrm{cer}}}\right)\exp\left(-\frac{\delta}{\delta^c_{\mathrm{cer}}}\right)\right]. \tag{6}$$

Figure 1(*b*) shows typical curves $\sigma_{\mathrm{cer}}/\sigma^c_{\mathrm{met}}$ versus $\delta/\delta^c_{\mathrm{met}}$ for various values of $\delta^c_{\mathrm{cer}}/\delta^c_{\mathrm{met}}$.

By substituting Eqs. (4) and (5) into Eq. (1), we obtain the cohesive traction of the FGM under loading conditions as



(*a*)



(*b*)

Fig. 1 Normalized cohesive traction versus nondimensional separation displacement; (*a*) for metal, $\sigma_{met}/\sigma^c_{met}$ versus δ/δ^c_{met}; (*b*) for ceramic, $\sigma_{cer}/\sigma^c_{met}$ versus δ/δ^c_{met} (where metal/ceramic strength ratio, $\sigma^c_{met}/\sigma^c_{cer}$, is taken to be 3)

$$\sigma_{\rm fgm}(\mathbf{x}) = \frac{V_{\rm met}(\mathbf{x})}{V_{\rm met}(\mathbf{x}) + \beta_{\rm met}[1 - V_{\rm met}(\mathbf{x})]} e\sigma^c_{\rm met}\left(\frac{\delta}{\delta^c_{\rm met}}\right)\exp\left(-\frac{\delta}{\delta^c_{\rm met}}\right) + \frac{1 - V_{\rm met}(\mathbf{x})}{1 - V_{\rm met}(\mathbf{x}) + \beta_{\rm cer} V_{\rm met}(\mathbf{x})} e\sigma^c_{\rm cer}\left(\frac{\delta}{\delta^c_{\rm cer}}\right)\exp\left(-\frac{\delta}{\delta^c_{\rm cer}}\right). \tag{7}$$

The free-energy density function corresponding to the above cohesive traction is

$$\phi_{\rm fgm}(\mathbf{x}, \delta, q) = \frac{V_{\rm met}(\mathbf{x})}{V_{\rm met}(\mathbf{x}) + \beta_{\rm met}[1 - V_{\rm met}(\mathbf{x})]} \times e\sigma^c_{\rm met}\delta^c_{\rm met}\left[1 - \left(1 + \frac{\delta}{\delta^c_{\rm met}}\right)\exp\left(-\frac{\delta}{\delta^c_{\rm met}}\right)\right] + \frac{1 - V_{\rm met}(\mathbf{x})}{1 - V_{\rm met}(\mathbf{x}) + \beta_{\rm cer} V_{\rm met}(\mathbf{x})} \times e\sigma^c_{\rm cer}\delta^c_{\rm cer}\left[1 - \left(1 + \frac{\delta}{\delta^c_{\rm cer}}\right)\exp\left(-\frac{\delta}{\delta^c_{\rm cer}}\right)\right]. \tag{8}$$

As often assumed for homogeneous materials, the cohesive law of the FGM also follows an irreversible path. The internal variable describing the irreversible processes, q, is chosen as $\delta_{\max}$, the maximum opening displacement attained. For updating of the cohesive stresses, the loading condition is defined by

$$\delta = \delta_{\max} \quad \text{and} \quad \dot{\delta} \geqslant 0, \tag{9}$$

where $\dot{\delta}$ denotes the rate of δ. The unloading condition is then described by

$$\delta < \delta_{\max} \quad \text{or} \quad \dot{\delta} < 0. \tag{10}$$

Following the cohesive law for homogeneous materials, the unloading curve follows the linear relation

$$\sigma_{fgm} = \left(\frac{\sigma_{\max}}{\delta_{\max}}\right)\delta, \quad \text{if } \delta < \delta_{\max} \text{ or } \dot{\delta} < 0, \tag{11}$$

where $\sigma_{\max}$ is the value of σ_{fgm} at $\delta = \delta_{\max}$ calculated from Eq. (7). We note that the irreversibility of the above cohesive law does not influence the results reported in Section 5 since we have only studied crack growth under monotonic loading conditions.

2.2 Three-Dimensional Mixed Mode Fracture. For general three-dimensional mixed mode fracture problems, an effective opening displacement jump is introduced ([19])

$$\delta_{\rm eff} = \sqrt{\delta_n^2 + \eta^2 \delta_s^2}, \tag{12}$$

where δ_n and δ_s are the normal and tangential displacement jumps across the cohesive surfaces. The parameter η assigns different weights to the opening and sliding displacements (η is usually taken as $\sqrt{2}$). Similarly, an effective cohesive traction may be introduced ([19])

$$\sigma_{\rm eff} = \sqrt{\sigma_n^2 + \eta^{-2}\sigma_s^2}, \tag{13}$$

where σ_n and σ_s are the normal and shear tractions across the cohesive surfaces. Here we assume that resistance of the cohesive surfaces to relative sliding is isotropic in the cohesive (tangent) plane so that

$$\delta_s = \sqrt{\delta_{s1}^2 + \delta_{s2}^2}, \tag{14}$$

$$\sigma_s = \sqrt{\sigma_{s1}^2 + \sigma_{s2}^2}, \tag{15}$$

where δ_{s1} and δ_{s2} are the two relative sliding displacements across the cohesive surfaces, and σ_{s1} and σ_{s2} are the two shear tractions.

With the introduction of the above effective traction and displacement, a free-energy potential in three dimensions is assumed to exist in the same form as that for the mode I case (8), i.e.,

$$\phi_{\rm fgm}(\mathbf{x}, \delta_{\rm eff}, \delta^{\max}_{\rm eff}) = \frac{V_{\rm met}(\mathbf{x})}{V_{\rm met}(\mathbf{x}) + \beta_{\rm met}[1 - V_{\rm met}(\mathbf{x})]} \times e\sigma^c_{\rm met}\delta^c_{\rm met}\left[1 - \left(1 + \frac{\delta_{\rm eff}}{\delta^c_{\rm met}}\right)\exp\left(-\frac{\delta_{\rm eff}}{\delta^c_{\rm met}}\right)\right] + \frac{1 - V_{\rm met}(\mathbf{x})}{1 - V_{\rm met}(\mathbf{x}) + \beta_{\rm cer} V_{\rm met}(\mathbf{x})} \times e\sigma^c_{\rm cer}\delta^c_{\rm cer}\left[1 - \left(1 + \frac{\delta_{\rm eff}}{\delta^c_{\rm cer}}\right)\exp\left(-\frac{\delta_{\rm eff}}{\delta^c_{\rm cer}}\right)\right], \tag{16}$$

where $\delta^{\max}_{\rm eff}$ is the maximum value of $\delta_{\rm eff}$ attained. The cohesive law for general three-dimensional deformations is then formulated as follows:

$$\sigma_n = \frac{\partial \phi_{\rm fgm}}{\partial \delta_n} = \frac{\partial \phi_{\rm fgm}}{\partial \delta_{\rm eff}}\frac{\partial \delta_{\rm eff}}{\partial \delta_n} = \left(\frac{\sigma_{\rm eff}}{\delta_{\rm eff}}\right)\delta_n,$$
$$\sigma_s = \frac{\partial \phi_{\rm fgm}}{\partial \delta_s} = \frac{\partial \phi_{\rm fgm}}{\partial \delta_{\rm eff}}\frac{\partial \delta_{\rm eff}}{\partial \delta_s} = \eta^2\left(\frac{\sigma_{\rm eff}}{\delta_{\rm eff}}\right)\delta_s, \tag{17}$$

where

$$\sigma_{\text{eff}}=\frac{\partial\phi_{\text{fgm}}}{\partial\delta_{\text{eff}}}=\frac{V_{\text{met}}(\mathbf{x})}{V_{\text{met}}(\mathbf{x})+\beta_{\text{met}}[1-V_{\text{met}}(\mathbf{x})]}e\sigma^c_{\text{met}}\left(\frac{\delta}{\delta^c_{\text{met}}}\right)\times\exp\left(-\frac{\delta}{\delta^c_{\text{met}}}\right)+\frac{1-V_{\text{met}}(\mathbf{x})}{1-V_{\text{met}}(\mathbf{x})+\beta_{\text{cer}}V_{\text{met}}(\mathbf{x})}\times e\sigma^c_{\text{cer}}\left(\frac{\delta}{\delta^c_{\text{cer}}}\right)\exp\left(-\frac{\delta}{\delta^c_{\text{cer}}}\right),$$

$$\text{if}\quad \delta_{\text{eff}}=\delta^{\max}_{\text{eff}}\quad\text{and}\quad\dot{\delta}_{\text{eff}}\geqslant 0, \tag{18}$$

for the loading case, and

$$\sigma_{\text{eff}}=\left(\frac{\sigma^{\max}_{\text{eff}}}{\delta^{\max}_{\text{eff}}}\right)\delta_{\text{eff}},\quad\text{if}\quad\delta_{\text{eff}}<\delta^{\max}_{\text{eff}}\quad\text{or}\quad\dot{\delta}_{\text{eff}}<0, \tag{19}$$

for the unloading case, where $\sigma^{\max}_{\text{eff}}$ is the value of σ_{eff} at $\delta_{\text{eff}}=\delta^{\max}_{\text{eff}}$ calculated from Eq. (18).

2.3 Cohesive Energy Density. The cohesive energy density, or the work of separation per unit area of cohesive surface, is defined by

$$\Gamma^c_{\text{fgm}}=\int_0^\infty\sigma(\delta_{\text{eff}})d\delta_{\text{eff}}. \tag{20}$$

By substituting Eq. (18) into the above equation, we obtain

$$\Gamma^c_{\text{fgm}}(\mathbf{x})=\frac{V_{\text{met}}(\mathbf{x})}{V_{\text{met}}(\mathbf{x})+\beta_{\text{met}}[1-V_{\text{met}}(\mathbf{x})]}\Gamma^c_{\text{met}}+\frac{1-V_{\text{met}}(\mathbf{x})}{1-V_{\text{met}}(\mathbf{x})+\beta_{\text{cer}}V_{\text{met}}(\mathbf{x})}\Gamma^c_{\text{cer}}, \tag{21}$$

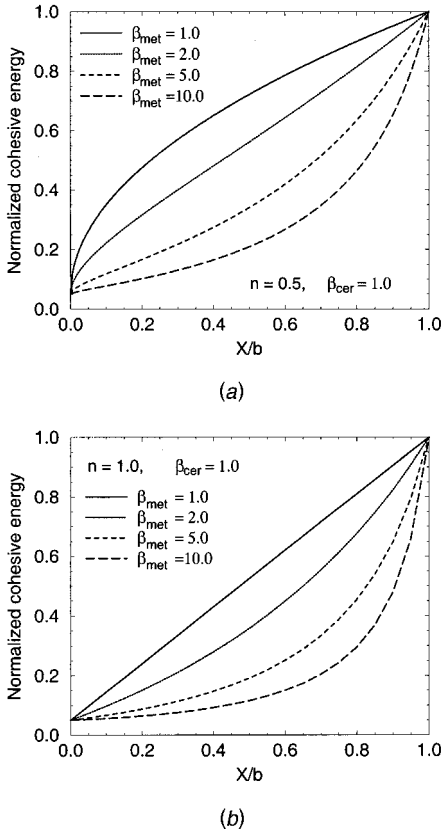


Fig. 2 Normalized cohesive energy density $\Gamma^c_{\text{fgm}}/\Gamma^c_{\text{met}}$ ($\Gamma^c_{\text{cer}}/\Gamma^c_{\text{met}}=0.05$, $V_{\text{met}}(X)=(X/b)^n$), (*a*) $n=0.5$; (*b*) $n=1.0$

where Γ^c_{met} and Γ^c_{cer} are the cohesive energy densities of the metal and ceramic phases, respectively,

$$\Gamma^c_{\text{met}}=e\sigma^c_{\text{met}}\delta^c_{\text{met}},\quad\Gamma^c_{\text{cer}}=e\sigma^c_{\text{cer}}\delta^c_{\text{cer}}. \tag{22}$$

Equation (21) shows that the cohesive energy density follows the same rule as that of the cohesive traction. Figure 2 shows the normalized cohesive energy density $\Gamma^c_{\text{fgm}}/\Gamma^c_{\text{met}}$ versus the nondimensional coordinate X/b for a ceramic/metal FGM with metal volume fraction $V_{\text{met}}=(X/b)^n$, where X is the gradation direction and b is a geometrical parameter, e.g., the thickness of the FGM specimen. In these figures, the energy ratio, $\Gamma^c_{\text{cer}}/\Gamma^c_{\text{met}}$ is assumed 0.05 with β_{cer} taken as 1.0. The cohesive energy of the FGM decreases markedly with increasing β_{met}.

3 Three-Dimensional Finite Element Modeling of Functionally Graded Materials

This section describes the small-displacement formulation of both the three-dimensional solid element and the interface-cohesive element with graded material properties. In the present study, the solid elements remain linearly elastic but the material properties (Young's modulus and Poisson's ratio) may vary within the element and thus graded elements are employed (Kim and Paulino [20]). For the cohesive element, the material properties follow the functionally graded cohesive law described in Section 2. Figure 3 illustrates the three-dimensional interface-cohesive and solid elements used in the present work. The interface-cohesive element consists of two four-node bilinear isoparametric surfaces. Nodes 1–4 lie on one surface of the element while nodes 5–8 lie on the opposite surface. The two surfaces initially occupy the same location. When the whole body deforms, the two surfaces undergo both normal and tangential displacements relative to each other. The cohesive tractions corresponding to the relative displacements follow the constitutive relations (17)–(19), and thus maintain the two surfaces in a "cohesive" state.

Now first consider the stiffness matrix of the isoparametric solid element. Denote by $N_i(\xi,\eta,\zeta)(i=1,2,\ldots,m)$ the standard shape functions of the solid element ([21]), where m is the number of the nodes of the element. The element stiffness matrix is given by

$$\mathbf{K}=\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}\mathbf{B}^T\mathbf{D}\mathbf{B}J_0 d\xi d\eta d\zeta, \tag{23}$$

where $\mathbf{B}$ is the strain-displacement matrix, J_0 is the usual Jacobian of the transformation between parametric (ξ,η,ζ) and Cartesian coordinates (x_1,x_2,x_3), and $\mathbf{D}$ is the elastic stiffness matrix. For functionally graded materials (FGMs), the $\mathbf{D}$ matrix depends on

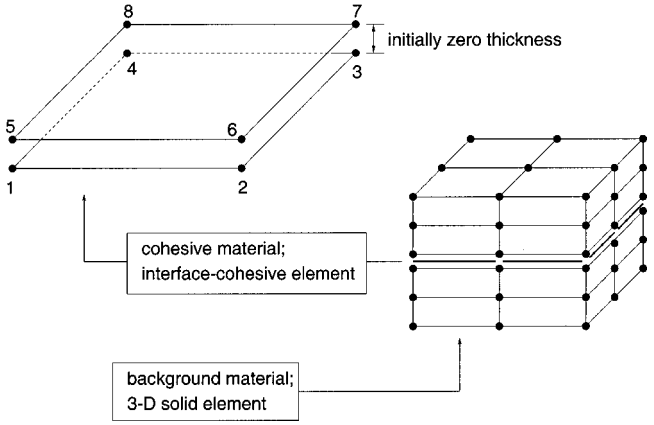


Fig. 3 Interface-cohesive and three-dimensional solid elements

spatial position. To calculate the Young's modulus and the Poisson's ratio in the solid element, we use the following interpolation:

$$E=\sum_{i=1}^{m} N_i E_i, \quad \nu=\sum_{i=1}^{m} N_i \nu_i, \tag{24}$$

where E_i and $\nu_i (i=1,2,\ldots,m)$ are the values of the Young's modulus and the Poisson's ratio at nodal points, respectively.

Turning to the cohesive element, the tangent stiffness matrix is given by ([8]),

$$\mathbf{K}_T=\int_{-1}^{1}\int_{-1}^{1} \mathbf{B}_{\text{coh}}^T \mathbf{D}_{\text{coh}} \mathbf{B}_{\text{coh}} J_0 \, d\eta \, d\zeta, \tag{25}$$

where $\mathbf{B}_{\text{coh}}$ extracts the relative displacement jumps within the cohesive element from the nodal displacements ([8]), J_0 is the Jacobian of the transformation between parametric (η,ζ) and Cartesian coordinates (s_1,s_2) in the tangent plane of the cohesive element, and $\mathbf{D}_{\text{coh}}$ is the tangent modulus matrix of the cohesive law (17)–(19) which can be found in the Appendix. For FGMs, Eq. (18) and Eqs. (34) and (35) in the Appendix show that the $\mathbf{D}_{\text{coh}}$ matrix depends on spatial position through the graded volume fraction of the metal phase, V_{met}, in a ceramic/metal FGM. In this study, V_{met} is also approximated by the standard interpolation

$$V_{\text{met}}=\sum_{i=1}^{4} N_i V_{\text{met}}^i, \tag{26}$$

where $V_{\text{met}}^i (i=1,2,3,4)$ are the values of V_{met} at the nodal points of the interface-cohesive elements. The present formulation is fully isoparametric in which the same shape functions interpolate the displacements, the geometry and the material parameters. Such a generalized isoparametric formulation has been presented by Kim and Paulino [20].

4 Functionally Graded Material Properties

This section describes the techniques adopted to obtain the properties for both the background functionally graded materials (FGM) and cohesive FGM materials. One of the advantages of the present methodology is that each model is developed separately for each material, as described below. This feature introduces significant flexibility in modeling the actual material behavior.

4.1 Background Material Properties. Consider an FGM as a two-phase composite with graded volume fractions of its constituent phases. The effective properties of an FGM should be calculated from those of the constituent materials and the volume fractions by means of a micromechanical model. Though such a model is not available as yet for FGMs, some models for conventional homogeneous composite materials, for example, the self-consistent scheme, may be used for FGMs with reasonable accuracy ([22]). In this study, we use the self-consistent scheme ([23]) to calculate the effective elastic properties of the FGM. The shear and bulk moduli μ_{fgm} and K_{fgm} of the FGM are thus calculated by the following system of equations:

$$\begin{aligned}&(\mu_{\text{fgm}}-\mu_{\text{met}})(\mu_{\text{fgm}}-\mu_{\text{cer}})\left[\frac{V_{\text{met}}K_{\text{met}}}{K_{\text{met}}+4\mu_{\text{fgm}}/3}+\frac{(1-V_{\text{met}})K_{\text{cer}}}{K_{\text{cer}}+4\mu_{\text{fgm}}/3}+2\right]\\&\quad+5[V_{\text{met}}\mu_{\text{cer}}(\mu_{\text{fgm}}-\mu_{\text{met}})+(1-V_{\text{met}})\mu_{\text{met}}(\mu_{\text{fgm}}-\mu_{\text{cer}})]\\&\quad=0,\end{aligned} \tag{27}$$

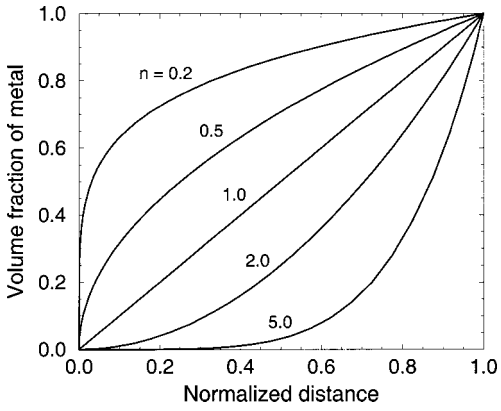


Fig. 4 Volume fraction of metallic phase in a ceramic/metal functionally graded material (FGM)

$$K_{\text{fgm}}=-\frac{4}{3}\mu_{\text{fgm}}+\frac{(K_{\text{met}}+4\mu_{\text{fgm}}/3)(K_{\text{cer}}+4\mu_{\text{fgm}}/3)}{V_{\text{met}}(K_{\text{cer}}+4\mu_{\text{fgm}}/3)+(1-V_{\text{met}})(K_{\text{met}}+4\mu_{\text{fgm}}/3)}. \tag{28}$$

The Young's modulus E_{fgm} and the Poisson's ratio ν_{fgm} of the FGM are then determined from the following relations:

$$E_{\text{fgm}}=\frac{9\mu_{\text{fgm}}K_{\text{fgm}}}{\mu_{\text{fgm}}+3K_{\text{fgm}}}, \tag{29}$$

$$\nu_{\text{fgm}}=\frac{3K_{\text{fgm}}-2\mu_{\text{fgm}}}{2(\mu_{\text{fgm}}+3K_{\text{fgm}})}. \tag{30}$$

In the present study, the volume fraction of the metal phase follows a simple power function, i.e.,

$$V_{\text{met}}(X)=\left(\frac{X-X_{\min}}{X_{\max}-X_{\min}}\right)^n, \tag{31}$$

where n is the power exponent, X is the gradation direction, and the material properties are graded in the interval $[X_{\min},X_{\max}]$. Figure 4 shows the volume fraction of the metal phase for various values of n.

The following numerical analysis of crack growth utilizes the properties of a TiB/Ti FGM system. Table 1 lists the relevant material properties of TiB (titanium monoboride) and Ti (commercially pure titanium). The company CERCOM Inc. developed this ceramic/metal FGM system in a layered structural form for armor applications ([24]).

4.2 Cohesive Material Properties. The functionally graded cohesive constitutive model (7) or (18) (three-dimensional case) has the following six independent parameters that characterize the fracture process in a ceramic/metal FGM:

Γ_{met}^c: local work of separation of metal
Γ_{cer}^c: local work of separation of ceramic
σ_{met}^c: peak cohesive traction of metal
σ_{cer}^c: peak cohesive traction of ceramic

Table 1 Material Properties of Ti and TiB

Materials	Young's modulus (GPa)	Poisson's ratio	J_c (KJ/m^2)	σ_{met}^c (MPa)	δ_{met}^c (mm)	σ_{cer}^c (MPa)	δ_{cer}^c (mm)
Ti	107	0.34	150	620	0.089		
TiB	375	0.14	0.11			4.0	0.01

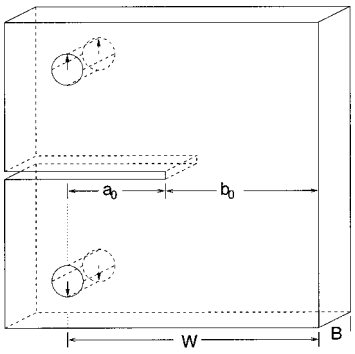


Fig. 5 C(T) specimen geometry

β_{met} and β_{cer}: cohesive gradation parameters.

The calibrated values of Γ^c_{met} and σ^c_{met} are the Griffith energy release rate (under small-scale yielding conditions) and the peak cohesive stress of the metal phase, which generally lies between two to three times the uniaxial yield stress. The first equation of (22) yields the characteristic opening displacement δ^c_{met}. We note that ductile deformations are present in the background material when the cohesive characteristic parameters of the metal are calibrated following the above procedure, however, such deformations are not considered in the present study. Nevertheless, this calibration procedure is used with emphasis on the presentation of the cohesive model and the effects of gradation parameter β_{met} and metal volume fraction V_{met} on the load versus crack growth responses. For the ceramic phase, it is natural to assign the energy release rate to Γ^c_{cer}. For this phenomenological model applicable at engineering scales, the characteristic opening displacement δ^c_{cer} is assumed to be approximately the average grain size of ceramic particles in the ceramic/metal FGM. The peak cohesive traction σ^c_{cer} is therefore determined from the second equation of (22). At smaller length scales, the local nature of the failure mechanism contributes to the characteristic parameters of the cohesive zone model, which may lead to different material parameters and different simulation results of crack growth. Calibration of the other two parameters β_{met} and β_{cer} follows by matching the predicted, with measured, fracture behavior. Table 1 lists the relevant cohesive properties for the TiB/Ti FGM, where the critical J values (J_c, as the cohesive energy) for TiB and Ti are taken from references [25,26].

5 Crack Growth in TiB/Ti Functionally Graded Materials

5.1 Finite Element Models. We performed numerical analyses of crack growth for both C(T) and SE(B) specimens, as illustrated in Figs. 5–6 and Figs. 7–8, respectively. Table 2 summarizes the geometric parameters of the C(T) specimen. The absolute size for the specimen is $W=50$ mm. The initial nondimensional crack length is $a_0/W=0.4$, the initial nondimensional ligament size is then $b_0/W=0.6$, and the thickness is 4.5 mm. Table 3 provides the geometric parameters for the SE(B) specimens. A layered functionally graded material (FGM) version of the SE(B) specimen has been recently tested in [25]. From a modeling point of view, the functionally graded material (FGM) composition varies from 100 percent TiB at the cracked surface to 100 percent Ti at the uncracked surface. Thus the volume fraction of Ti varies from zero at the cracked surface to one at the uncracked surface.

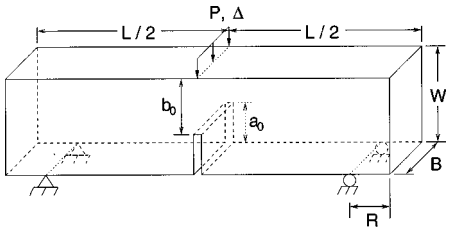


Fig. 7 SE(B) specimen geometry

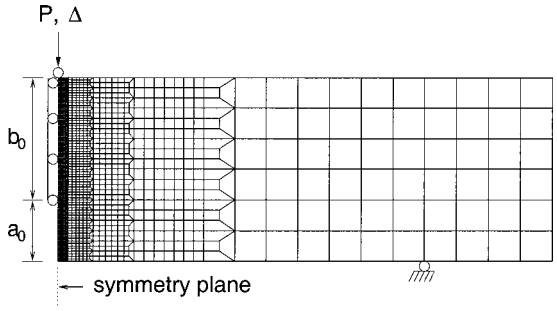


Fig. 8 Typical mesh for analyses of SE(B) specimen

The finite element models consist of eight-node isoparametric solid elements and the eight-node interface-cohesive elements. Due to symmetry considerations, we model only one-quarter of each specimen. Interface-cohesive elements are placed only over the initial uncracked ligament and have a uniform size of 0.25 mm for the C(T) specimen, and 0.1 mm for the SE(B) specimens. The finite element model has eight uniform layers of elements over the half thickness for the C(T) specimen. For the thicker SE(B) specimens, the model has ten uniform layers over the half thickness. Figure 6 shows the front view of the typical finite element mesh for the C(T) specimen and Fig. 8 shows the front view of the finite element mesh for the SE(B) specimens.

5.2 Finite Element Analysis. The FGM modeling features described in this work have been implemented in the fracture mechanics research code WARP3D ([27]). In addition to the conventional solid and interface-cohesive elements for homogeneous materials, this code also incorporates the solid element with

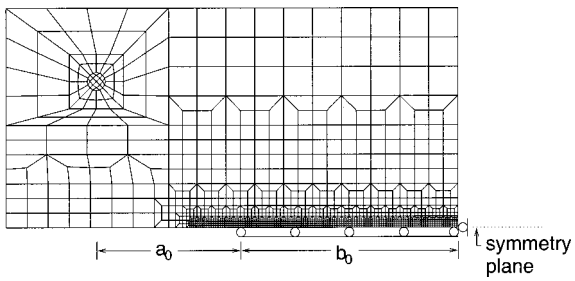


Fig. 6 Typical mesh for analyses of C(T) specimen

Table 2 Geometric parameters of C(T) specimen

Specimen	W (mm)	B (mm)	a_0/W
C(T)	50	4.5	0.4

Table 3 Geometric parameters of SE(B) specimens

Specimen	L (mm)	W (mm)	B (mm)	a_0/W	R (mm)
SE(B)	79.4	14.7	7.4	0.1, 0.3	10.2

graded elastic properties and the interface-cohesive element coupled with the functionally graded cohesive constitutive model described in Sections 2 and 3.

WARP3D supports the conventional interface-cohesive element for crack growth with adaptive load control, element extinction and other features. Such computational procedures (previously used for homogeneous materials) also prove essential in analyses of FGMs to track accurately the cohesive constitutive response. For the cohesive fracture model proposed in Section 2, the adaptive load control parameter becomes the characteristic opening displacement δ^c_{met} of the metal. The analysis uses a limit of $\Delta\delta/\delta^c_{\text{met}} = 0.2$ per load step for adaptive load control, where $\Delta\delta$ is the largest change of effective opening displacement δ experienced by interface-cohesive elements in a given load step. The element extinction occurs when the average opening displacement δ of the element reaches $5\,\delta^c_{\text{met}}$, which corresponds to a cohesive traction less than 10 percent of the peak value of the metal multiplied by the metal volume fraction. Selection of δ^c_{met} (of the metal phase) as the controlling parameter for adaptive load control and element extinction follows from the analyses demonstrating that the metal phase largely controls fracture behavior of the FGM. The cohesive fracture energy of TiB, for example, is less than 0.1 percent of that for Ti.

5.3 Crack Growth in C(T) Specimen. The specimen is loaded by opening displacements applied uniformly through the thickness at the loading pin. Crack growth is taken to occur when the interface-cohesive elements ahead of the crack front satisfy the element extinction condition. Figure 9(*a*) shows the load versus crack extension curves for the C(T) specimen for various values of β_{met}. The power exponent $n = 0.5$ (shape index of the metal volume fraction) defines an overall metal rich specimen. Because the cohesive traction of the ceramic phase is extremely small compared with that of the metal phase for the TiB/Ti FGM studied, the parameter β_{cer} plays a negligible role in determining the cohesive traction of the FGM. Consequently, we take $\beta_{\text{cer}} = 1.0$ in the current and all subsequent calculations. Figure 9(*a*) shows that for a given β_{met}, the load decreases steadily with crack extension in the present analyses which do not include plasticity in the background material. This contrasts with ductile fracture of metals which show load increases with crack extension during initial growth followed by load reductions when strain hardening no longer accommodates the decreasing ligament (see [8] for examples). The figure also shows that for a given crack extension, a larger β_{met} lowers the load. This is consistent with the cohesive fracture model (7) where a larger β_{met} reduces the peak cohesive traction. Figure 9(*b*) and 9(*c*) show similar results for the same specimen for $n = 1.0$ (a specimen with equal overall metal and ceramic volume fractions) and $n = 2.0$ (an overall metal lean specimen), respectively. Comparing the results in the three figures, we observe that the load becomes lower for larger *n*. The result is expected since a larger *n* corresponds to a lower metal volume fraction, which results in a lower cohesive energy for the FGM.

Figure 10 shows the load versus crack extension curves for the same C(T) specimen studied in Fig. 9(*a*) with addition of the crack growth responses for homogeneous metal (Ti) and ceramic (TiB) specimens. These two additional configurations provide bounding solutions for the FGM responses. The load for pure Ti remains larger than those for the TiB/Ti FGM with various β_{met}. The loads during crack extension for the pure TiB, however, remain vanishingly small compared to the FGM.

5.4 Crack Growth in SE(B) Specimen. As a final numerical example, we consider an SE(B) specimen loaded by opening displacements applied uniformly through the thickness at the specimen center plane. A layered FGM version of the specimen has been recently tested ([25]). The first layer of the tested specimen consists of 15 percent Ti and 85 percent TiB, while the last layer (seventh layer) consists of 100 percent Ti. Crack initiation

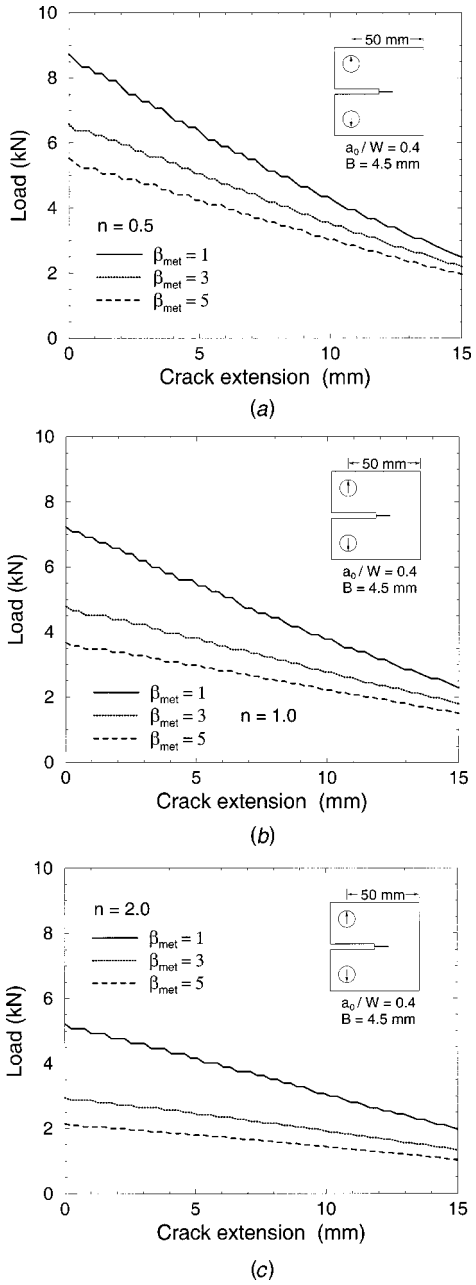


Fig. 9 Load-crack extension response for the C(T) Ti/TiB specimen with $a_0/W = 0.4$, $B = 4.5$ mm; (*a*) $n = 0.5$; (*b*) $n = 1.0$; (*c*) $n = 2.0$

occurred at a measured load of 920 Newtons (N). The experimental results show that load increases with crack extension during the initial growth and then decreases with further crack extension. The measured load corresponding to a crack growth of 5 mm is about 1200 N. Figure 11 shows the volume fraction of Ti in this TiB/Ti specimen. The dotted (stepped) line shows the property gradation in the experimentally tested specimen. A least-squares approximation yields the power exponent $n = 0.84$ in the metal volume fraction function of Eq. (31). Figure 12 shows the numerical results of the load versus crack extension responses for the SE(B) specimen with $\beta_{\text{met}} = 16$ and $n = 0.84$. For the β_{met} selected, the crack initiation load agrees quite closely with the experimentally measured value. Compared with the experimental observations after the crack initiation ([25]), the discrepancy in the trend of load versus crack extension response arises because the present analysis does not consider plasticity in the background

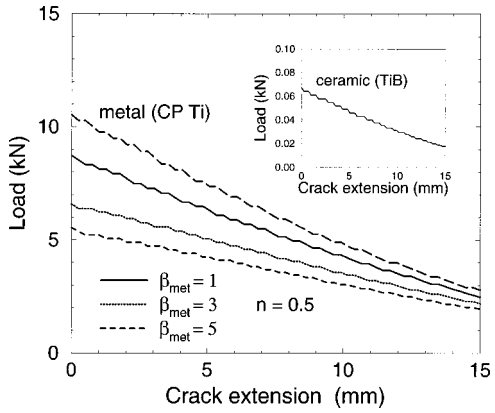


Fig. 10 Load-crack extension response for the C(T) Ti/TiB specimen with $a_0/W=0.4$, $B=4.5$ mm

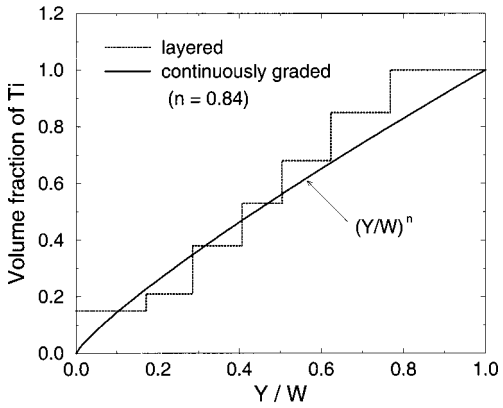


Fig. 11 Volume fraction of Ti in the TiB/Ti functional graded material (FGM)

material. When the plasticity effect is taken into account (work underway by the authors), we expect that the trend of the load versus crack extension will be more consistent with the experimental observations (the calibrated value of β_{met} may be larger than 16). Figure 12 also shows the numerical results of the load versus crack extension for plane-strain and plane-stress models. Though we have not found differences between the two-dimensional and three-dimensional responses, we expect that significant differences will develop with plasticity in the background material due to variations in crack front constraint and crack front tunneling.

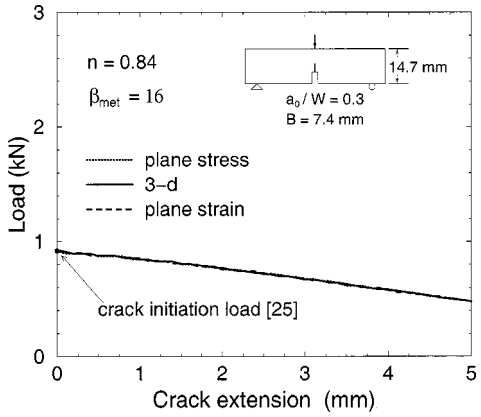


Fig. 12 Load-crack extension response for the SE(B) Ti/TiB specimen with $a_0/W=0.3$, $B=7.4$ mm, $n=0.84$

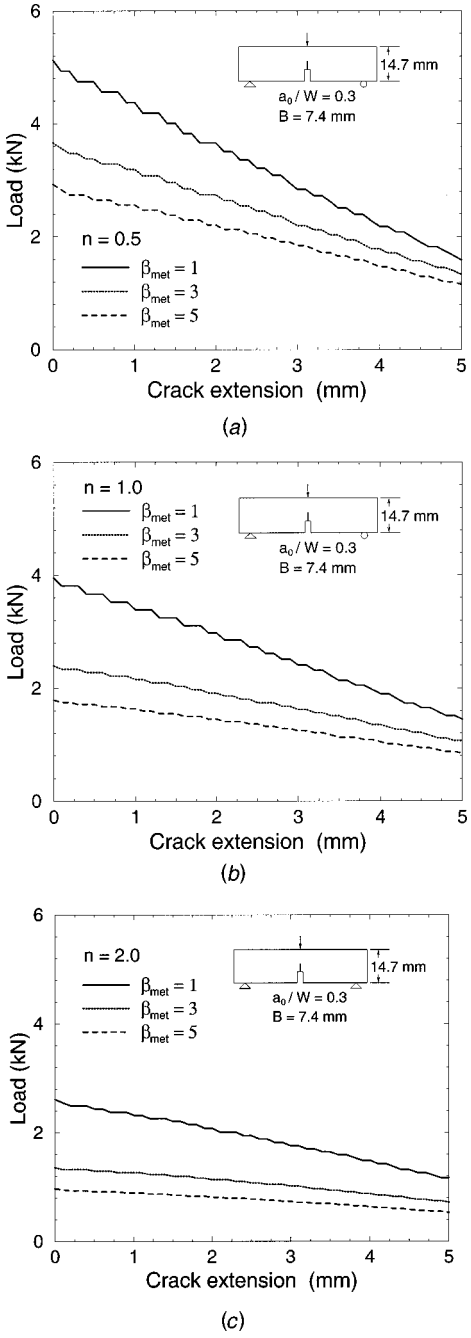


Fig. 13 Load-crack extension response for the SE(B) Ti/TiB specimen with $a_0/W=0.3$, $B=7.4$ mm; (*a*) $n=0.5$; (*b*) $n=1.0$; (*c*) $n=2.0$

Figure 13 shows the effect of β_{met} and n on the load versus crack extension responses for the SE(B) specimen. The power exponent n (shape index of the metal volume fraction) is 0.5 in Fig. 13(*a*), 1.0 in Fig. 13(*b*), and 2.0 in Fig. 13(*c*). Similar load versus crack extension behavior to that for the C(T) specimen is observed for the SE(B) specimen, i.e., for a given β_{met}, the load decreases steadily with crack extension; for a given crack extension, a larger β_{met} reduces the load; and finally, the load becomes lower for larger n. Because β_{met} has a pronounced effect on the load versus crack extension responses, we may expect to calibrate the values of β_{met} from experimental observations (see Section 2).

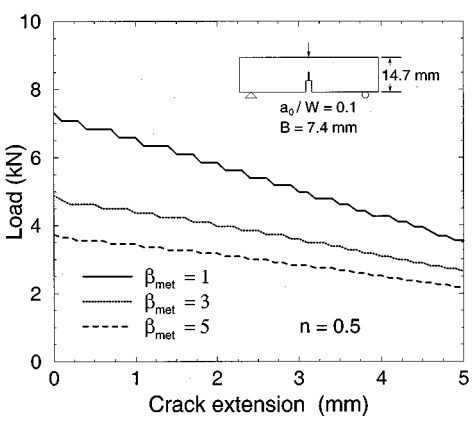


Fig. 14 Load-crack extension response for the SE(B) Ti/TiB specimen with $a_0/W=0.1$, $B=7.4$ mm; and $n=0.5$

Figure 14 shows the load versus crack extension curves for the SE(B) specimen with an initial nondimensional crack length $a_0/W=0.1$. Similar results to that shown in Fig. 13(*a*) can be observed. Therefore, without considering plasticity in the background material, the load decreases with crack extension due to the decreasing ligament for the laboratory crack size.

6 Concluding Remarks

This study presents a novel phenomenological cohesive fracture model for ceramic/metal functionally graded materials (FGMs) and the corresponding implementation in a three-dimensional finite element method framework. The model has six independent material parameters, i.e., the cohesive energy densities $(\Gamma^c_{\text{met}},\Gamma^c_{\text{cer}})$, the peak cohesive tractions of the metal and ceramic phases $(\sigma^c_{\text{met}},\sigma^c_{\text{cer}})$ and two cohesive gradation parameters $(\beta_{\text{met}},\beta_{\text{cer}})$ to represent approximately the transition between the fracture mechanisms of metal and ceramic phases, respectively. In contrast to existing models that consider only tensile mode fracture, the present model accommodates three-dimensional tensile and shear fracture modes although the numerical examples illustrate only mode I fracture behavior. Applications of the cohesive fracture model to the analysis of crack growth in both C(T) and SE(B) specimens of TiB/Ti FGM show that the load to cause crack extension in the FGM compares to that for a pure metal (Ti) specimen. In the present study, the load decreases steadily with subsequent crack extension, which contrasts with ductile fracture behavior of metals (the present analyses do not admit plasticity in the background material). The results obtained indicate that the cohesive gradation parameter for the metal has a pronounced effect on the load versus crack extension response. This suggests that the parameter may be reasonably calibrated by matching the predicted crack growth response with experimental observations. An extension of this work includes consideration of plasticity in the bulk FGM (background). In this case, we expect to detect strong three-dimensional effects of crack front tunneling and variations in crack front constraint. This investigation is currently being pursued by the authors.

Acknowledgments

This work was sponsored by a grant from the NASA Ames Research Center (NAG 2-1424) to the University of Illinois at Urbana-Champaign. Dr. Tina Panontin serves as the project Technical Monitor. We also acknowledge partial support of NSF under grant No. CMS-0115954 (Mechanics and Materials Program).

Appendix

In finite element analyses, the tangent modulus matrix $D_{ij}(i,j=1,2,3)$ for the cohesive models (18) defined below is needed

$$\dot{\sigma}_i = D_{ij}\dot{v}_j, \tag{32}$$

where $(\sigma_1,\sigma_2,\sigma_3)=(\sigma_{s1},\sigma_{s2},\sigma_n)$, $(v_1,v_2,v_3)=(v_{s1},v_{s2},v_n)$, and $D_{ij}=\partial\sigma_i/\partial v_j$. Here D_{ij} are the components of $\mathbf{D}_{\text{coh}}$ matrix in Eq. (25). The detailed expression for D_{ij} is given as follows. First note that

$$D_{ij}=\frac{\partial}{\partial v_j}\left(\frac{\partial\phi_{\text{fgm}}}{\partial\delta_{\text{eff}}}\frac{\partial\delta_{\text{eff}}}{\partial v_i}\right)=\sigma_{\text{eff}}\frac{\partial^2\delta_{\text{eff}}}{\partial v_i\partial v_j}+\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}\frac{\partial\delta_{\text{eff}}}{\partial v_i}\frac{\partial\delta_{\text{eff}}}{\partial v_j}. \tag{33}$$

It is clear from the above equation that $D_{ij}=D_{ji}$. Use of Eqs (12), (14), and (33) yields

$$\begin{aligned}
D_{11}&=\eta^2\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}+\frac{\eta^4v_1^2}{\delta_{\text{eff}}^2}\left(\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}-\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}\right),\\
D_{22}&=\eta^2\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}+\frac{\eta^4v_2^2}{\delta_{\text{eff}}^2}\left(\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}-\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}\right),\\
D_{33}&=\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}+\frac{v_3^2}{\delta_{\text{eff}}^2}\left(\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}-\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}\right),\\
D_{12}&=D_{21}=\eta^4\frac{v_1v_2}{\delta_{\text{eff}}^2}\left(\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}-\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}\right),\\
D_{13}&=D_{31}=\eta^2\frac{v_1v_3}{\delta_{\text{eff}}^2}\left(\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}-\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}\right),\\
D_{23}&=D_{32}=\eta^2\frac{v_2v_3}{\delta_{\text{eff}}^2}\left(\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}-\frac{\sigma_{\text{eff}}}{\delta_{\text{eff}}}\right),
\end{aligned} \tag{34}$$

where σ_{eff} is given by Eq. (18) and $\partial\sigma_{\text{eff}}/\partial\delta_{\text{eff}}$ is

$$\begin{aligned}
\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}&=\frac{V_{\text{met}}(\mathbf{x})}{V_{\text{met}}(\mathbf{x})+\beta_{\text{met}}[1-V_{\text{met}}(\mathbf{x})]}e\left(\frac{\sigma^c_{\text{met}}}{\delta^c_{\text{met}}}\right)\left(1-\frac{\delta_{\text{eff}}}{\delta^c_{\text{met}}}\right)\\
&\quad\times\exp\left(-\frac{\delta_{\text{eff}}}{\delta^c_{\text{met}}}\right)+\frac{1-V_{\text{met}}(\mathbf{x})}{1-V_{\text{met}}(\mathbf{x})+\beta_{\text{cer}}V_{\text{met}}(\mathbf{x})}\\
&\quad\times e\left(\frac{\sigma^c_{\text{cer}}}{\delta^c_{\text{cer}}}\right)\left(1-\frac{\delta_{\text{eff}}}{\delta^c_{\text{cer}}}\right)\exp\left(-\frac{\delta_{\text{eff}}}{\delta^c_{\text{cer}}}\right),
\end{aligned} \tag{35}$$

under loading conditions, and

$$\sigma_{\text{eff}}=\left(\frac{\sigma^{\max}_{\text{eff}}}{\delta^{\max}_{\text{eff}}}\right)\delta_{\text{eff}}, \tag{36}$$

$$\frac{\partial\sigma_{\text{eff}}}{\partial\delta_{\text{eff}}}=\frac{\sigma^{\max}_{\text{eff}}}{\delta^{\max}_{\text{eff}}} \tag{37}$$

for the unloading case.

References

[1] Hirai, T., 1996, "Functionally Gradient Materials," *Materials Science and Technology: Processing of Ceramics, Part 2*, R. J. Brook eds., VCH Verlagsgesellschaft mbH, Weinheim, Germany, **17B**, pp. 292–341.
[2] Koizumi, M., 1993, "The Concept of FGMs," *Ceramic Transactions: Functionally Graded Materials*, J. B. Holt, M. Koizumi, T. Hirai, and Z. Munir, eds., American Ceramic Society, Westerville, OH, **34**, pp. 3–10.
[3] Suresh, S., and Mortensen, A., 1998, *Functionally Graded Materials*, The Institute of Materials, IOM Communications, London.
[4] Dugdale, D. S., 1960, "Yielding of Steel Sheets Containing Slits," J. Mech. Phys. Solids, **8**, pp. 100–104.
[5] Hillerborg, A., Modeer, M., and Petersson, P. E., 1976, "Analysis of Crack Formation and Crack Growth in Concrete by Means of Fracture Mechanics and Finite Elements," Cement Concrete Res, **6**, pp. 773–782.
[6] Mosalam, K. M., and Paulino, G. H., 1997, "Evolutionary Characteristic Length Method for Smeared Cracking Finite Element Models," Finite Elem. Anal. Design, **27**, pp. 99–108.

[7] Needleman, A., 1987, ''A Continuum Model for Void Nucleation by Inclusion Debonding,'' ASME J. Appl. Mech., **54**, pp. 525–531.
[8] Roy, Y. A., and Dodds, Jr., R. H., 2001, ''Simulation of Ductile Crack Growth in Thin Aluminum Panels Using 3-D Surface Cohesive Elements,'' Int. J. Fract., **110**, pp. 21–45.
[9] Finot, M., Shen, Y.-L., Needleman, A., and Suresh, S., 1994, ''Micromechanical Modeling of Reinforcement Fracture in Particle-Reinforced Metal-Matrix Composites,'' Metall. Trans. A, **25**, pp. 2403–2420.
[10] Jin, Z.-H., and Batra, R. C., 1996, ''Some Basic Fracture Mechanics Concepts in Functionally Graded Materials,'' J. Mech. Phys. Solids, **44**, pp. 1221–1235.
[11] Cai, H., and Bao, G., 1998, ''Crack Bridging in Functionally Graded Coatings,'' Int. J. Solids Struct., **35**, pp. 701–717.
[12] Williamson, R. L., Rabin, B. H., and Drake, J. T., 1993, ''Finite Element Analysis of Thermal Residual Stresses at Graded Ceramic-Metal Interfaces, Part I: Model Description and Geometrical Effects,'' J. Appl. Phys., **74**, pp. 1310–1320.
[13] Giannakopoulos, A. E., Suresh, S., Finot, M., and Olsson, M., 1995, ''Elastoplastic Analysis of Thermal Cycling: Layered Materials With Compositional Gradients,'' Acta Metall. Mater., **43**, pp. 1335–1354.
[14] Rose, J. H., Ferrante, J., and Smith, J. R., 1981, ''Universal Binding Energy Curves for Metals and Bimetallic Interfaces,'' Phys. Rev. Lett., **47**, pp. 675–678.
[15] Siegmund, T., and Needleman, A., 1997, ''A Numerical Study of Dynamic Crack Growth in Elastic-Viscoplastic Solids,'' Int. J. Solids Struct., **34**, pp. 769–788.
[16] Ortiz, M., and Pandolfi, A., 1999, ''Finite-Deformation Irreversible Cohesive Elements for Three-Dimensional Crack-Propagation Analysis,'' Int. J. Numer. Methods Eng., **44**, pp. 1267–1282.
[17] Tvergaard, V., and Hutchinson, J. W., 1992, ''The Relation Between Crack Growth Resistance and Fracture Process Parameters in Elastic-Plastic Solids,'' J. Mech. Phys. Solids, **40**, pp. 1377–1392.
[18] Guinea, G. V., Elices, M., and Planas, J., 1997, ''On the Initial Shape of the Softening Function of Cohesive Materials,'' Int. J. Fract., **87**, pp. 139–149.
[19] Camacho, G. T., and Ortiz, M., 1996, ''Computational Modelling of Impact Damage in Brittle Materials,'' Int. J. Solids Struct., **33**, pp. 2899–2938.
[20] 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.
[21] Cook, R. D., Malkus, D. S., and Plesha, M. E., 1989, *Concepts and Applications of Finite Element Analysis*, 3rd Ed., John Wiley and Sons, New York.
[22] Reiter, T., Dvorak, G. J., and Tvergaard, V., 1997, ''Micromechanical Models for Graded Composite Materials,'' J. Mech. Phys. Solids, **45**, pp. 1281–1302.
[23] Hill, R., 1965, ''A Self-Consistent Mechanics of Composite Materials,'' J. Mech. Phys. Solids, **13**, pp. 213–222.
[24] Nelson, G., and Ezis, A., 1996, ''Functionally Graded Material (FGM) Armor in the TiB/Ti system (U),'' *CERCOM Report*, Vista, CA.
[25] 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, **308–311**, pp. 837–842.
[26] Liang, W. W., 1999, ''Finite Element Analysis of Model I Crack Propagation in Layered Functionally Graded Materials,'' M.Sc. thesis, University of California, Davis, CA.
[27] Gullerud, A. S., Koppenhoefer, K. C., Roy, A., and Dodds, Jr., R. H., 2000, *WARP3D—Release 13.8 Manual*, Civil Engineering, Report No. UILU-ENG-95-2012, University of Illinois, Urbana, IL.

Reduction of Vibration Caused by Magnetic Force in a Switched Reluctance Motor by Topology Optimization

J. Yoo
Assistant Professor,
School of Mechanical Engineering,
Yonsei University,
Seoul 120-749, Korea
e-mail: yoojh@yonsei.ac.kr
Assoc. Mem. ASME

The topology optimization of a switched reluctance motor (SRM) excited by magnetic forces is an important issue to minimize the noise and vibration level. In this paper, the magnetic force is computed using the Maxwell stress method and the optimization problem is formulated to minimize the frequency response based on the homogenization design method (HDM). The developed method is applied to the stator of an SRM to minimize the deformation caused by the magnetic harmonic excitation. Numerical simulation shows that this method successfully decreases the vibration level of an SRM.
[DOI: 10.1115/1.1467093]

1 Introduction

Among many types of motors, a switched reluctance motor (SRM) has a very simple structure compared to the other types of motors because there is no wire winding around the rotor. Also, since no magnetization is necessary, the manufacturing cost can be significantly reduced ([1]). These make an SRM a strong contender in industrial applications.

However, the noise and vibration level of an SRM is higher than other competing motors. Vibration in an electric machine has a harmful effect on the efficiency of the machine, especially when the machine work must be accurate and quiet. Cameron and Lang [2] found that the dominant source of the noise and vibration is the radial deformation of a stator due to radial magnetic forces. They also found that the result is severe when the operating frequency coincides with the natural resonant frequency of the stator. The effect of the deformation of a rotor is small compared to that of a stator since a rotor is stiffer than a stator. Colby et al. [3] used structural finite element analysis to compute the eigenmodes of a four-phase SRM and presented heuristic arguments to find the operating conditions that excite the dominant eigenmodes. Wu and Pollock [4] derived a concept for the power electric controller to reduce the noise and vibration in a switched reluctance drive.

In low-frequency ranges, the flux flow between a stator and a rotor is the main cause of vibration of an SRM. The flux flow generates magnetic forces that perform as exciting forces for the vibration. On the other hand, in high-frequency ranges, the flux ripple of the air-gap dominates the vibration. In both cases, to analyze and solve the vibration problem in electric machines, the method to calculate the forces caused by magnetic flux must be chosen carefully according to the characteristic of the problem.

The methods of force calculation based on the finite element method can be classified as follows: the Maxwell stress method, the virtual work method, and the magnetizing current method. Ito et al. [5] compared these three methods and confirmed that the Maxwell stress method and the virtual work method are less sensitive to density of mesh discretization: however, the virtual work method is not good in large deformation cases. Hamler et al. [6] used all three of these methods to calculate the torque of a direct current (DC) motor and showed that the Maxwell stress method shows the best result with small number of elements. Removing or reducing the exciting forces that cause severe deformation can reduce the vibration of an SRM. However, we must consider that the magnetic forces work not only as exciting forces causing the noise and vibration but also as driving forces of an SRM.

In this paper, we consider the frequency response problem of an SRM excited by magnetic forces. The objective of the design is to minimize the vibration level of the motor. The finite element method together with the Maxwell stress method is used to calculate the magnetic forces, and the homogenization design method (HDM) is applied to obtain the optimal topology. First introduced by Bendsøe and Kikuchi [7] for structural topology design, this method has been successfully applied to structural optimization problems. Diaz and Kikuchi [8] extended the application of the HDM to the dynamic problem and determined the optimal shape of a structure to maximize a natural frequency. Ma et al. [9] developed a shape optimization method for the frequency response problem of a vibrating structure.

Based on these studies, we developed a process for minimizing vibration of a structure excited by magnetic forces and applied the process to obtain the optimal topology of a stator in an SRM.

2 Homogenization Design Method

In this section, homogenization theory in elasticity cases and the concept of the homogenization design method are explained. The homogenized material properties such as homogenized Young's modulus or density are computed using homogenization theory.

2.1 Homogenization Theory in Elasticity. The optimal shape of a design domain is determined by the optimal material distribution using a composite material composed of variable microstructures. Thus, a microstructure is introduced to solve the topology optimization problem. To determine whether a structural part is removed or not, we introduce a characteristic function such that

$$\chi_\Omega = \begin{cases} 1 & \text{in solid} \\ 0 & \text{in void} \end{cases}. \tag{1}$$

Figure 1 shows design domain composed of the nonhomogeneous composite material. $\mathbf{y}(y1,y2,y3)$ represents the micro-

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, Oct. 23, 2001. Associate Editor: A. K. Mal. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

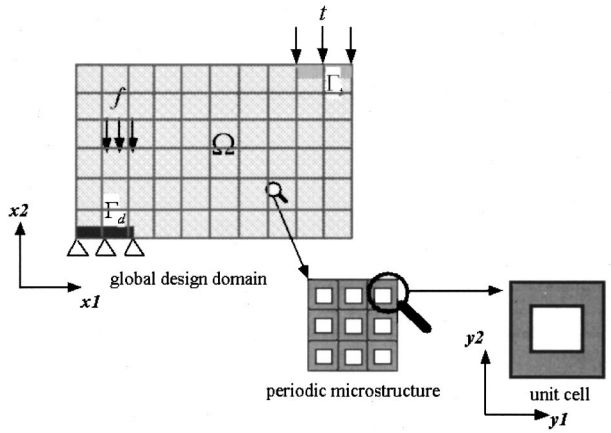


Fig. 1 Design domain compared of microstructures

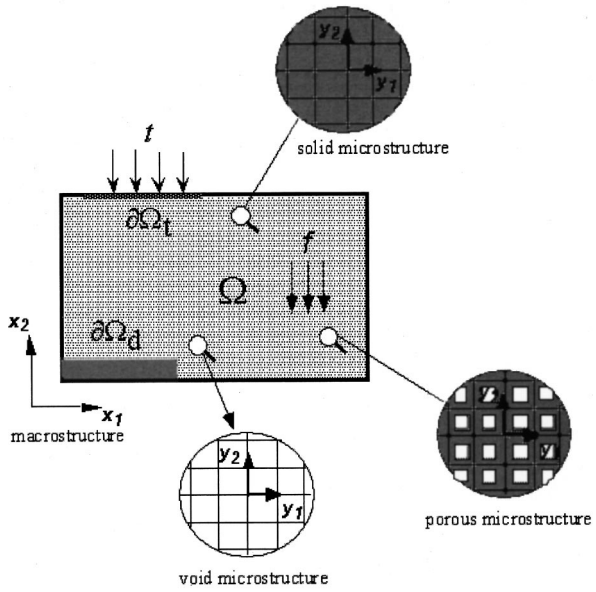


Fig. 2 Macro design domain and porous microstructures

scopic level coordinate while $\mathbf{x}(x1,x2,x3)$ represents the macroscopic level coordinate. Those two coordinates have the following relation:

$$\mathbf{y}=\frac{\mathbf{x}}{\varepsilon} \tag{2}$$

where ε is an asymptotic scale factor between the macroscopic and the microscopic structure. Domain Ω is an open connected domain of R^3 with smooth boundaries. The domain is composed of infinite number of microstructures, each of which has a hole as shown in the figure. Γ_t is the traction boundary condition and Γ_d is the displacement boundary condition of domain Ω.

Using the characteristic function defined in Eq. (1), the equilibrium of a structure composed of microstructures can be stated as a weak form:

$$\int_\Omega \chi E^0_{ijkl}\frac{\partial u_i^\varepsilon}{\partial x_j}\frac{\partial v_k}{\partial x_l}d\Omega+\int_\Omega \chi\rho_0\frac{\partial^2 u_i^\varepsilon}{\partial t^2}d\Omega$$
$$=\int_\Omega f_i^0 v_i d\Omega+\int_{\Gamma_t} t_i v_i d\Gamma \quad \text{for } \forall v_i\in V \tag{3}$$

where u_i^ε is the displacement at equilibrium, v_i is the virtual displacement, and t_i is the boundary traction. E^0_{ijkl}, ρ_0, and f_i^0 represent Young's modulus tensor, mass density of the solid portion, and the body force of the solid portion, respectively. V represents the space of kinematically admissible displacement fields defined as follows:

$$V=\{v\,|\,v\in H^1(\Omega), v=0 \text{ on } \Gamma_d\},\quad \Gamma_d:\text{displacement boundary}$$

where $H^1(\Omega)$ is the Sobolev space.

Since E^0_{ijkl}, ρ_0, and f_i^0 are not homogeneous in the design domain, they must be solved at macroscale and microscale levels. Using the homogenization theory, we can write the weak form as

$$\int_\Omega E^H_{ijkl}\frac{\partial u_i^0}{\partial x_j}\frac{\partial v_k}{\partial x_l}d\Omega+\int_\Omega \rho^H\frac{\partial^2 u_i^0}{\partial t^2}d\Omega=\int_\Omega f_i^H v_i d\Omega+\int_{\Gamma_t} t_i v_i d\Gamma \tag{4}$$

where u_i^0 represents the average displacement in the microscale domain. E^H_{ijkl}, ρ^H, and f_i^H are homogenized Young's modulus, mass density, and body force, respectively. E^H_{ijkl} is calculated using the characteristic displacement of the unit cell. ρ^H and f_i^H can be represented as $\rho^H=1/|Y|\int_Y \rho_0 dY$ and $f^H=1/|Y|\int_Y f_i^0 dY$, respectively, since they are proportional to the volume ratio of each unit cell Y. $|Y|$ represents the volume of Y ([7,9]).

2.2 Concept of the Homogenization Design Method. The design domain, Ω, is composed of a composite material with perforated microstructures as shown in Fig. 2. The design domain has given boundary conditions $\partial\Omega_d$ and $\partial\Omega_t$, which are the displacement boundary condition and traction boundary condition, respectively. The unit cell of the microstructure has a rectangular hole, as shown in Fig. 1. If the size of the hole in the unit cell is 0, the unit cell becomes a solid. On the other hand, if the size is the same as the unit cell size, the unit cell becomes a void. During the optimization process, the material is transferred from one part to another part in the design domain and finally the optimal material distribution is determined.

The optimization problem using the homogenization design method (HDM) is defined using the total potential energy or mean compliance of a design domain. As can be seen from Eq. (4), the weak form, which defines the energy, is formulated using the homogenized properties such as Young's modulus, density, and the body force. Thus, to obtain the homogenized properties of a given microstructure, we must solve the unit cell problem. The unit cell of three-dimensional microstructure is assumed to have a body hole of width 1-a, depth 1-b, and height 1-c. The size of body hole $\mathbf{D}=\{a,b,c\}$, and rotation angle $\Theta=\{\varphi,\theta,\psi\}$ are the design variables. Assuming $\mathbf{R}(\boldsymbol{\Theta})$ to be the rotational matrix based on rotational angle $\boldsymbol{\Theta}$, we can compute Young's modulus using the following equation:

$$\mathbf{E}=\mathbf{R}(\boldsymbol{\Theta})^T\mathbf{E}^H(\mathbf{D})\mathbf{R}(\boldsymbol{\Theta}). \tag{5}$$

During the optimization process, the design variables of each unit cell are changing continuously. Therefore, the homogenized values of Young's modulus must be obtained by continuously changing the size of the hole for all possibilities of material density. This is possible if homogenized Young's modulus can be expressed by a function of the design variables. However, homogenized Young's modulus is too complex to express as a single function. Therefore, we calculate the homogenized values only for some discrete sizes of the hole, and interpolate the other values by the Bezier function.

3 Magnetic Force Calculation

The force caused by magnetic harmonic flux of the air-gap can have a harmful effect on quiet electric machines if it works as an

exciting force of the machine vibration. Therefore, accurate computation of the force in the air-gap is necessary to solve the noise and vibration problem in electrical machinery.

The Maxwell stress method can be applied regardless of the material properties. Also, it can be used for linear material properties as well as nonlinear material properties while the integration path must be carefully selected for accurate. The integration path must not include any ferromagnetic material, and it must be positioned as far as possible from the ferromagnetic material. Also, the application of this method is restricted to the air-gap part. To calculate the force in the portion that includes a ferromagnetic material, Reyne et al. [10] suggested a modified stress tensor using an energy approach. In spite of these limitations, the Maxwell stress method is widely used to calculate the magnetic force since the results using this method are consistent with theory and experimental results.

Using this method, we define volume force density as

$$f_v = \nabla \cdot \sigma \tag{6}$$

where $\boldsymbol{\sigma}$ is the Maxwell stress tensor for three dimensional cases and it is defined as follows:

$$\sigma = \frac{1}{\mu_0}\begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{xy} & \sigma_{zz} \end{bmatrix} = \frac{1}{\mu_0}\begin{bmatrix} B_x^2 - \frac{1}{2}|B|^2 & B_xB_y & B_xB_z \\ B_yB_x & B_y^2 - \frac{1}{2}|B|^2 & B_yB_z \\ B_zB_x & B_zB_y & B_z^2 - \frac{1}{2}|B|^2 \end{bmatrix} \tag{7}$$

where B_x, B_y, and B_z represents the x, y, and z directional flux density value, respectively. $|B|^2$ is computed using those values as $B_x^2 + B_y^2 + B_z^2$. We can compute the magnetic force by integrating volume force density over an integration path as follows:

$$\mathbf{F} = \int_V \nabla \cdot \sigma dv. \tag{8}$$

The volume integration is transformed from volume integration to surface integration by the divergence theorem:

$$\mathbf{F} = \int_A \sigma \cdot d\mathbf{A}. \tag{9}$$

Although the Maxwell stress tensor can be defined in the x-y coordinate, usually it is determined in the normal and the tangential directions of a material to apply for circular shapes such as motors and generators. In the analysis of a switched reluctance motor (SRM), we compute the normal and the tangential directional forces based on Eqs. (7)–(9) and the relation of x-y and t-n coordinates.

4 Optimization Problem

In this section we formulate the topology optimization process for the frequency response problem. The objective of the optimization process is to minimize the response of the stator of a switched reluctance motor (SRM) excited by harmonic loading caused by magnetic flux. We obtain the objective function by the weak form of the structure based on the virtual displacement theory. For the optimization solver, the optimality criteria method is used to deal with the large number of design variables.

4.1 Formulation of the Optimization Problem. We consider a design domain Ω where the body force **f** and traction **t** are applied as shown in Fig. 1. Using the homogenized Young's modulus, mass density, body force, and ignoring the damping effect, the weak form for a simple dynamic case can be written as Eq. (4). Assuming that u is the displacement at equilibrium and v is the virtual displacement, we can define energy bilinear form $a(\mathbf{u},\mathbf{v})$ and load linear form $b(\mathbf{v})$ as follows:

$$a(\mathbf{u},\mathbf{v}) = \int_\Omega E_{ijkl}^H \frac{\partial u_i^0}{\partial x_j}\frac{\partial v_k}{\partial x_l} d\Omega + \int_\Omega \rho^H \frac{\partial^2 u_i^0}{\partial t^2} d\Omega \tag{10}$$

$$b(\mathbf{v}) = \int_\Omega f_i^H v_i d\Omega + \int_{\Gamma_t} t_i v_i d\Gamma. \tag{11}$$

Therefore, we can write the weak form as

$$a(\mathbf{u},\mathbf{v}) = b(\mathbf{v}) \quad \forall \mathbf{v} \in S_k \tag{12}$$

where S_k is the space of kinematically admissible displacement. We call the right term of Eq. (12) as the mean compliance. At equilibrium, Eq. (12) can be rewritten as

$$a(\mathbf{u},\mathbf{u}) = b(\mathbf{u}). \tag{13}$$

The energy bilinear form shown in Eq. (10) includes a time derivative term. For numerical analysis, the time discretization method or the quasi-static analysis can be used. If the exciting force is harmonic and the frequency is given, the weak form is simplified without considering the time difference. Assuming ω to be the given frequency for the exciting force, we can substitute the displacement u_i^0, body force f_i^h, and traction t_i as $U_i e^{j\omega t}$, $F_i e^{j\omega t}$, and $T_i e^{j\omega t}$, respectively. Therefore, we can write each term of Eq. (13) as follows:

$$a(\mathbf{u},\mathbf{u}) = \int_\Omega E_{ijkl}^H \frac{\partial U_i}{\partial x_j}\frac{\partial U_k}{\partial x_l} d\Omega - \omega^2 \int_\Omega \rho^H U_i \cdot U_i d\Omega \tag{14}$$

$$b(\mathbf{u}) = \int_\Omega F_i U_i d\Omega + \int_{\Gamma_t} T_i U_i d\Gamma. \tag{15}$$

The objective of the topology optimization can be defined as minimizing the mean compliance defined as Eq. (11) to minimize deformation. However, if the exciting frequency is large, the sensitivity of the objective function can be a negative value and it is hard to apply the optimality criteria method. Ma et al. [9] shifted the sensitivity value using shift parameter to make the value as a positive one. This method requires additional calculation of the shift parameter according to the sensitivity of the objective function and constraint functions. In this paper, we modified the objective function as the square of the mean compliance to satisfy the criteria:

$$\underset{\mathbf{v}\in S_k}{\text{minimize}} \quad b^2(\mathbf{v}) \tag{16}$$

where the square value of the mean compliance is used as the objective to make the objective as a positive value regardless of the exciting frequency.

In the optimization process, the design variables are updated based on the sensitivity of the objective function. The sensitivity can be computed based on the total potential energy and the relation of the total potential energy and the mean compliance. Considering Eq. (16), we modify the total potential energy T_f as

$$T_f(\mathbf{v}) = \tfrac{1}{2}a^2(\mathbf{v},\mathbf{v}) - b^2(\mathbf{v}). \tag{17}$$

Since Eq. (13) is satisfied at equilibrium, we can derive the following relation:

$$T_f(\mathbf{u}) = \tfrac{1}{2}a^2(\mathbf{u},\mathbf{u}) - b^2(\mathbf{u}) = -\tfrac{1}{2}b^2(\mathbf{u}). \tag{18}$$

By the relation of Eq. (16) and Eq. (18), we can say that minimizing the square of the mean compliance is the same as maximizing the value of T_f.

Assuming **X** to be the design variable, and using Eq. (14) and Eq. (15), we can derive the variation of T_f with respect to the design variable **X** and equilibrium displacement **u** as follows:

$$\delta T_f(\mathbf{u},\mathbf{X}) = 2\left(\int_\Omega E^H_{ijkl} \frac{\partial U_i}{\partial x_j} \frac{\partial U_k}{\partial x_l} d\Omega - \omega^2 \int_\Omega \rho^H U_i \; U_i d\Omega \right)$$
$$\left[\int_\Omega E^H_{ijkl} \frac{\partial U_i}{\partial x_j} \frac{\partial}{\partial x_l} \left(\delta U_k + \frac{\partial U_k}{\partial X_m} \delta X_m \right) d\Omega \, \delta X_m \right.$$
$$- \omega^2 \int_\Omega \rho^H U_i \left(\partial U_i + \frac{\partial U_i}{\partial X_m} \partial X_m \right) d\Omega \, \delta X_m$$
$$+ \frac{1}{2} \int_\Omega \frac{\partial E^H_{ijkl}}{\partial X_m} \frac{\partial U_i}{\partial x_j} \frac{\partial U_k}{\partial x_l} d\Omega \, \delta X_m - \frac{1}{2} \omega^2 \int_\Omega \frac{\partial \rho^H}{\partial X_m} U_i$$
$$\left. U_i d\Omega \, \delta X_m \right] - 2b(\mathbf{u}) b\left(\delta U_i + \frac{\partial U_i}{\partial X_m} \delta X_m \right). \tag{19}$$

The first and second terms in the bracket can be canceled out with the last term by setting $\mathbf{v} = \delta\mathbf{u} + (\partial\mathbf{u}/\partial X_m)\delta X_m$. Thus, we can define the sensitivity function to design variable $\mathbf{X}$ as

$$\frac{\partial T_f}{\partial X_m} = 2\left(\int_\Omega E^H_{ijkl} \frac{\partial U_i}{\partial x_j} \frac{\partial U_k}{\partial x_l} d\Omega - \omega^2 \int_\Omega \rho^H U_i \; U_i d\Omega \right)$$
$$\left(\frac{1}{2} \int_\Omega \frac{\partial E^H_{ijkl}}{\partial X_m} \frac{\partial U_i}{\partial x_j} \frac{\partial U_k}{\partial x_l} d\Omega - \frac{1}{2} \omega^2 \int_\Omega \frac{\partial \rho^H}{\partial X_m} U_i \; U_i d\Omega \right). \tag{20}$$

To satisfy the criteria of the optimality criteria method, this value must be positive. Since homogenized elastic-moduli are highly nonlinear according to the material volume ratio ([7]), satisfaction of this condition cannot be guaranteed. However, considering that an SRM rotates at high speed, the sensitivity value can be positive if the following conditions are satisfied:

$$\int_\Omega E^H_{ijkl} \frac{\partial U_i}{\partial x_j} \frac{\partial U_k}{\partial x_l} d\Omega \leqslant \omega^2 \int_\Omega \rho^H U_i \; U_i d\Omega$$

and

$$\int_\Omega \frac{\partial E^H_{ijkl}}{\partial X_m} \frac{\partial U_i}{\partial x_j} \frac{\partial U_k}{\partial x_l} d\Omega \leqslant \omega^2 \int_\Omega \frac{\partial \rho^H}{\partial X_m} U_i \; U_i d\Omega. \tag{21}$$

Since the ratio of left terms in Eq. (21) depends on the materials used, the application of the objective function defined in Eq. (16) is limited. In the given model of an SRM, ω must be large enough to satisfy the conditions in Eq. (21). Considering the metallic material properties such as Young's modulus and density of steel used for stator parts, we can compute that the conditions suggested in Eq. (21) are satisfied when ω is over than 50 Hz. ω is the exciting frequency of the magnetic harmonic force and it is consistent with the motor speed. 50 Hz is approximately 480 rpm and the SRM is usually operated at higher speed than 480 rpm. According to the Colby et al.'s work ([3]), the deformation modes of an SRM for severe vibration and noise are much higher than 1 kHz. Therefore, the objective function expressed as in Eq. (16) can be used for ordinary operating conditions.

Using the relation of the total potential energy and the mean compliance shown in Eq. (18), we can define the sensitivity of the mean compliance as

$$\frac{\partial b^2(\mathbf{u})}{\partial X_m} = \left(\int_\Omega E^H_{ijkl} \frac{\partial U_i}{\partial x_j} \frac{\partial U_k}{\partial x_l} d\Omega - \omega^2 \int_\Omega \rho^H U_i \; U_i d\Omega \right)$$
$$\left(- \int_\Omega \frac{\partial E^H_{ijkl}}{\partial X_m} \frac{\partial U_i}{\partial x_j} \frac{\partial U_k}{\partial x_l} d\Omega + \omega^2 \int_\Omega \frac{\partial \rho^H}{\partial X_m} U_i \; U_i d\Omega \right). \tag{22}$$

4.2 Optimality Criteria Method. In the design process, the design variables are updated at each of the iterations using the sensitivity values based on an updating rule. As in the homogenization design method (HDM) if the number of design variables is much larger than the number of constraints, the optimality criteria method is very efficient.

Generally, the optimization problem using the HDM can be defined as follows:

$$\underset{x,\Theta}{\text{minimize}} \;\; f \tag{23}$$

$$\text{subject to} \;\; h(\mathbf{X}) \leqslant 0 \tag{24}$$

$$x_i^l \leqslant x_i \leqslant x_i^u, \quad i = 1,2,\ldots,n_x \tag{25}$$

$$-\frac{\pi}{2} \leqslant \theta_i \leqslant \frac{\pi}{2}, \quad i = 1,2,\ldots,n_\theta \tag{26}$$

where f is the objective function and h is the constraint function defined as $\int_\Omega 1 d\Omega \leqslant V_0$ where V_0 is the pre-defined total volume ratio. $\mathbf{X} = \{x_i\}$ is the set of the sizing design variables which corresponds to the size of holes in the microstructure in a finite element model. $\Theta = \{\theta_i\}$ is the set of the orientation design variables that corresponds to the rotation of the microstructure in a finite element model. x_i^l and x_i^u are the lower and upper bounds of design variable x_i, respectively. n_x and n_θ are the number of the sizing design variable and the number of the orientation design variable, respectively.

Using the Lagrangian function and the Kuhn-Tucker conditions, the necessary condition for a stationary point can be written as

$$e_i = \frac{1}{\lambda} \left(-\frac{\partial f}{\partial x_i} \Big/ \frac{\partial h}{\partial x_i} \right) = 1 \quad \text{for} \;\; x_i^l < x_i < x_i^u \tag{27}$$

where e_i is the effectiveness of the ith design variable. If the structure is not in the optimal stage, the value of the effectiveness is not one. Using the effectiveness, the updating rule for the optimality criteria method is defined as follows [7]:

$$x_i^{k+1} = \begin{cases} x_i^l & \text{for} \;\; (e_i^k)^\eta x_i^k \leqslant x_i^l \\ (e_i^k)^\eta x_i^k & \text{for} \;\; x_i^l \leqslant (e_i^k)^\eta x_i^k \leqslant x_i^u \\ x_i^u & \text{for} \;\; x_i^u \leqslant (e_i^k)^\eta x_i^k \end{cases} \tag{28}$$

where x_i^k is the ith sizing design variable in the kth iteration, x_i^{k+1} is the ith updated sizing design variable, and η is a weighting.

The updating rule regarding orientation design variable θ_i is based on the work of Pedersen [11]. Angle θ_i must be aligned to the directions of principal stresses for materials of weak shear stiffness. Since the material used in the topology optimization usually has weak shear stiffness, Pedersen's work can be applied as an updating rule for the orientation design variables.

5 Application to a Switched Reluctance Motor (SRM)

The homogenization design method (HDM) was applied to a 6/4 (six stator poles–four rotor poles) type SRM to obtain the optimal shape of the stator for minimizing the frequency response. After obtaining the optimal shape, the material properties of the structure were changed according to the optimization result and ABAQUS was used to verify the effect of the optimal shape.

5.1 Finite Element Model and Topology Optimization. Since the rotor is rotating, the relative position of a stator to a rotor is changing continuously. We analyzed at three different positions to compare the results according to the stator-rotor position. Figure 3 displays a finite element model used to compute the magnetic force. The model is composed of hexahedral three-dimensional elements one layer deep. The air-gap portion between a rotor and a stator is composed of three layer finite elements and the integration path for applying the Maxwell stress method is defined along the middle layer. The stator-rotor position drawn in the figure is defined as position 1 while positions 2 and 3 are

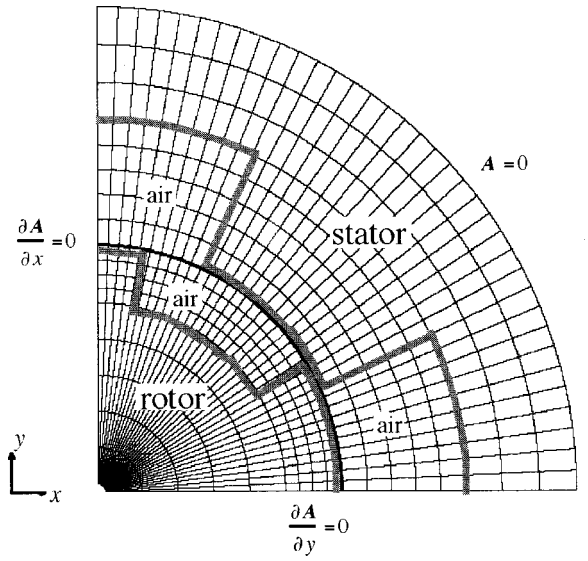


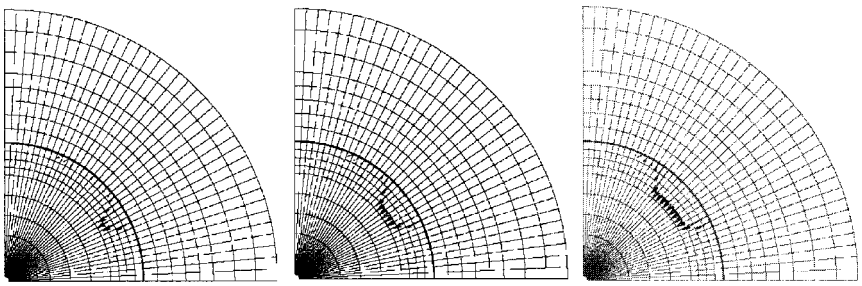
Fig. 3 Finite element model of a switched reluctance method (SRM) for magnetic force calculation: position 1

defined when the rotor is rotated counterclockwise from the position 1 by 10 deg and 20 deg, respectively. As can be seen from the figures, we use a quarter of the motor for modeling because of symmetry. Along the circular edge, the Dirichlet boundary condition is applied while the Neumman boundary conditions are applied along the symmetry line. Figure 4 shows the vector plot of magnetic forces computed by the Maxwell stress method based on the value of the flux density. As can be seen from Fig. 4, the magnetic forces are more widely distributed as the contact area of the rotor and the stator is larger. The magnetic forces shown in Fig. 4 are applied as exciting forces to obtain the optimal shapes for minimizing the frequency response at three different positions.

Figures 5(*a*) and 5(*b*) display the design domain with boundary conditions and load conditions. The magnetic forces are applied at the nodes designated in the Fig. 5(*a*). The optimization was performed at 100 Hz exciting frequency by changing the load conditions corresponding to the three different positions with 60 percent volume constraint.

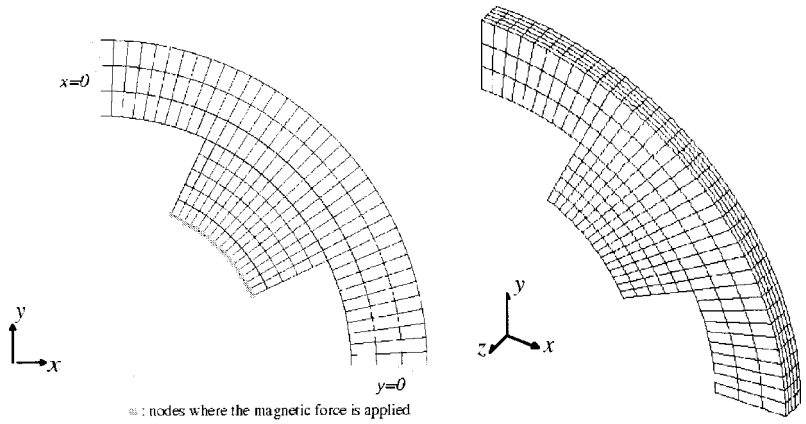
Figure 6 shows the optimal material distributions of the design domain at three different positions and the black parts represent high-density material. Figure 7 shows the three-dimensional optimal shapes composed of the selected elements whose density is larger than the 70 percent of the maximum density. As can be seen from these figures, more material is distributed around the stator pole as the contact area of the rotor and the stator becomes larger. Also, we can confirm that making the material density weak in the middle part of the stator pole is effective in reducing the weight without any harmful effect on the vibration level. Figure 8 shows the convergence history of the optimization process for the model representing position 1. We can verify that the objective function converges well until 100 iterations. This tendency is similar to the other two positions.

As shown in Fig. 7, the optimal topology is different according to the relative position between the stator and the rotor. It is almost impossible to make a stator that has 'changing' material distribution: however, we can designate the locations of the stator that keeps high-density material distribution regardless of the relative position and we can select such locations for the reinforcement in the design modification. If some relative positions such as



(*a*) position 1 (*b*) position 2 (*c*) position 3

Fig. 4 Vector plot of magnetic force



(*a*) 2-D (*b*) 3-D

Fig. 5 Design domain for the structural optimization of the stator

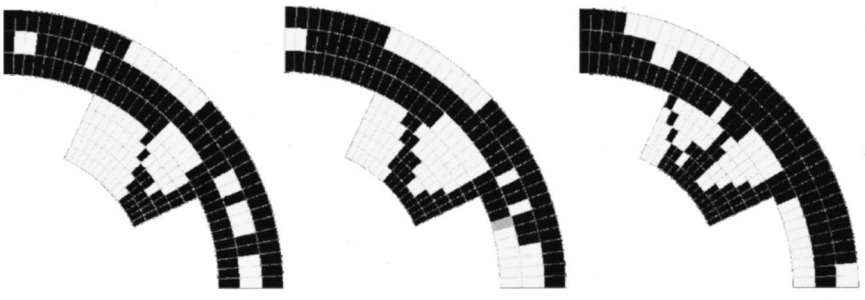


Fig. 6 Optimal shapes at three different positions

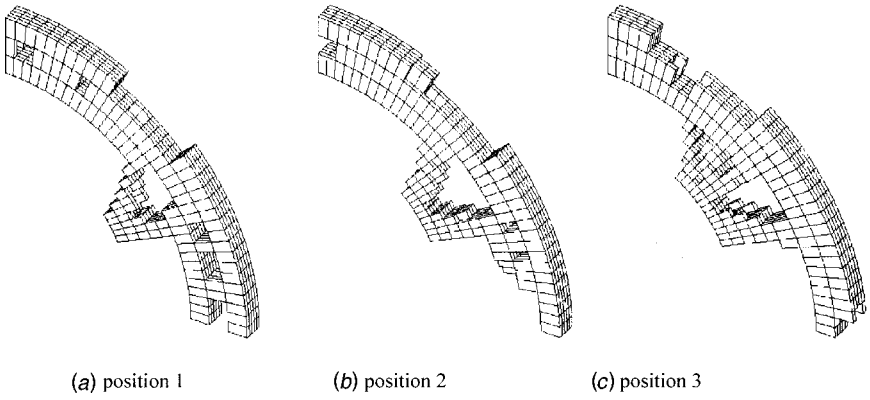


Fig. 7 Three-dimensional optimal shapes at three different positions

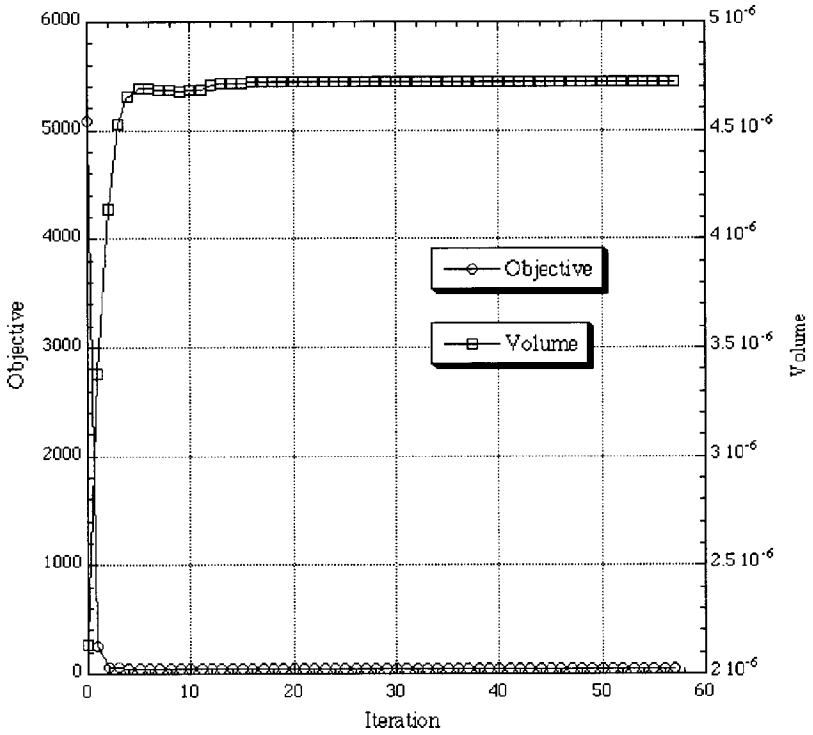


Fig. 8 Convergence history of the optimization for the position 1 with 100 Hz exciting frequency

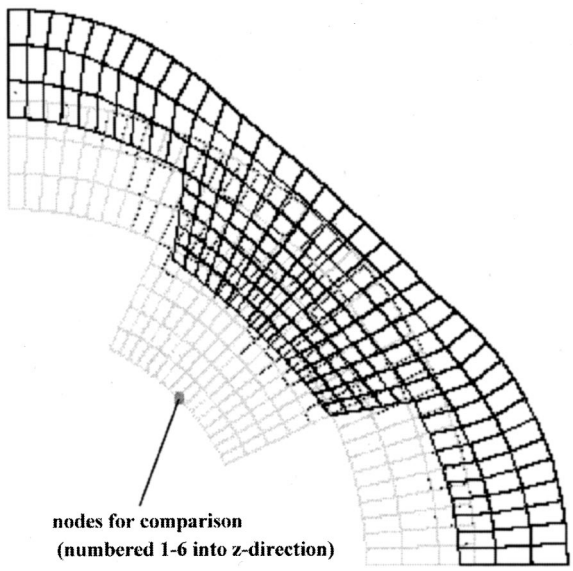


Fig. 9 Deformed shape by frequency response analysis at position 1

the contact-beginning position or the contact-ending position must be considered more seriously than others, the multiobjective function can be an effective alternative to obtain the optimal topology.

The magnetic force is also changed as the optimal material distribution of the stator is changed during the optimization process. In this study, the optimal topology shown in Fig. 7 is used to select the reinforcement part of the stator rather than removing the low-density part as in ordinary ''mass-minimization'' topology optimization. The final product of the stator is composed of two materials that have same ferromagnetic properties but different mechanical properties such as Young's modulus to minimize the frequency response as well as keeping the pattern of the magnetic force. Therefore, we can decouple the magnetic force calculation and the optimization process.

5.2 Verification by ABAQUS. We verified the optimal shapes using the commercial package ABAQUS. Considering that the maximum deformation occurs at the beginning of the transition period, position 1 is used for the verification. Figure 9 shows the deformed shape of the original stator by frequency response analysis using ABAQUS for position 1 at 100 Hz exciting frequency. As can be seen from the figure, the stator is deformed to the outward and the deformation at the stator pole edge, which is adjacent to the rotor pole, is very large. Therefore, we can verify the results by comparing the displacement at the nodes located at the edge as designated in Fig. 9.

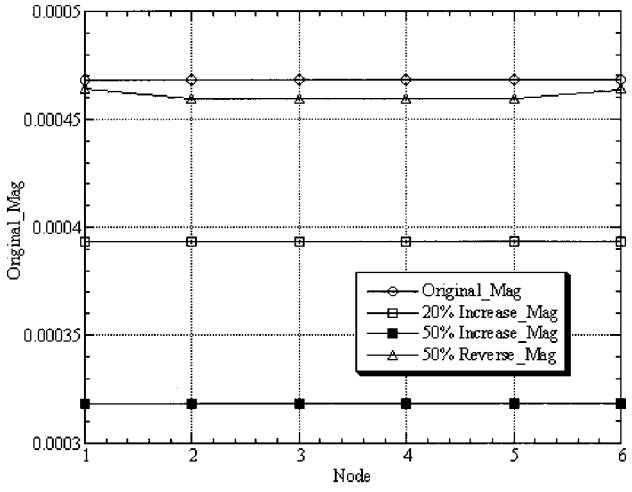


Fig. 10 Comparison of displacements at the selected nodes

As can be seen from Fig. 7, the optimal topology for minimizing vibration shows a hole at the stator pole. However, making a hole inside of a pole can distort magnetic flux and it can exert a bad influence upon the motor efficiency. As mentioned in Section 5.1, we focus the modification on changing the material property of the selected portion shown in Fig. 7(*a*) rather than removing the nonselected part to avoid the severe change in magnetic flux distribution. Therefore, we increased Young's modulus at the finite elements that are displayed in Fig. 7(*a*) while using the original material in the other part. For the first and second analyses, the value of Young's modulus of the elements was increased by 20 percent and 50 percent, respectively. Also, for a comparison, the value was increased by 50 percent at the elements that are not selected in Fig. 7(*a*) in the third analysis. During the analysis, the material density is assumed to be similar for the selected and nonselected parts. It can be realized by using the combination of iron and ferromagnetic alloy such as cast iron-nickel. Figure 10 shows the comparison of deformation magnitude at the selected nodes. The graph shows the root mean square value of *x* and *y*-directional displacements at the nodes. As can be seen from the graph, increasing Young's modulus is effective in reducing the deformation. However, increasing Young's modulus of the elements that are not selected by the optimization process is not effective in reducing the deformation. Therefore, we can confirm that not only increasing Young's modulus but also selecting the appropriate portion to increase Young's modulus is important in reducing the frequency response. By these results, we can confirm that the HDM is valid for the purpose of minimizing the frequency response.

6 Conclusion

In this paper, we have discussed on structural topology optimization using the homogenization design method (HDM) to minimize the frequency response of a switched reluctance motor (SRM). We have formulated the objective function as a square form of mean compliance to guarantee the positive value at ordinary exciting frequencies. The magnitude and direction of the exciting force were calculated using the Maxwell stress method. The developed optimization algorithm was applied to the stator of an SRM to minimize the frequency response caused by the magnetic forces.

The final design suggested is the optimal material distribution of two different materials in the stator to avoid the abrupt change of the magnetic flux distribution. It is also possible to obtain the optimal topology of a structure in magnetic fields to maximize the magnetic flux ([12]). This study can be extended as the multiobjective optimization considering both the magnetic flux and the frequency response to obtain the optimal topology composed of a single material.

References

[1] Lawrenson, P. J., Stephson, J. M., Blenkinsop, P. T., Corda, J., and Fulton, N. N., 1980, ''Variable Speed Switched Reluctance Motors,'' IEEE Proc. Part B, **133**(6), pp. 253–265.
[2] Cameron, D. E., and Lang, J. H., 1992, ''The Origin and Reduction of Acoustic Noise in Doubly Salient Variable Reluctance Motors,'' IEEE Trans. Ind. Appl., **28**(6), pp. 1250–1255.
[3] Colby, R. S., Mottier, F. M., and Miller, T. J. E., 1996, ''Vibration Mode and Acoustic Noise in a Four-Phased Switched Reluctance Motor,'' IEEE Trans. Ind. Appl., **32**(6), pp. 1357–1364.
[4] Wu, C., and Pollock, C., 1995, ''Analysis and Reduction of Vibration and Acoustic Noise in the Switched Reluctance Drive,'' IEEE Trans. Ind. Appl., **31**(1), pp. 91–98.
[5] Ito, M., Tajima, F., and Kanazawa, H., 1990, ''Evaluation of Force Calculation Methods,'' IEEE Trans. Magn., **26**(2), pp. 1035–1038.
[6] Hamler, A., Kreca, B., and Hribernik, B., 1992, ''Investigation of the Torque Calculation of a DC PM Motor,'' IEEE Trans. Magn., **28**(5), pp. 2271–2273.

[7] Bendsøe, M. P., and Kikuchi, N., 1988, ''Generating Optimal Topologies in Structural Design Using a Homogenization Method,'' Comput. Methods Appl. Mech. Eng., **71**, pp. 197–224.
[8] Diaz, A., and Kikuchi, N., 1992, ''Solution to Shape and Topology Eigenvalue Optimization Problems Using a Homogenization Method,'' Int. J. Numer. Methods Eng., **35**, pp. 1487–1502.
[9] Ma, Z. D., Kikuchi, N., and Hagiwara, I., 1993, ''Structural Topology and Shape Optimization for a Frequency Response Problem,'' Computational Mech., **13**, pp. 157–174.
[10] Reyner, G., Meunier, G., Imhoff, J. F., and Euxibie, E., 1988, ''Magnetic Forces and Mechanical Behavior of Ferromagnetic Materials. Presentation and Results on the Theoretical Experimental and Numerical Approaches,'' IEEE Trans. Magn., **24**(1), pp. 234–237.
[11] Pederson, P., 1989, ''On Optimal Orientation of Orthotropic Materials,'' Struct. Optim., **1**, pp. 101–106.
[12] Yoo, J., Kikuchi, N., and Volakis, J. L., 2000, ''Structural Optimization in Magnetic Fields Using the Homogenization Design Method,'' IEEE Trans. Magn., **36**(3), pp. 574–580.

Z.-G. Zhou
e-mail: zhouzhg@hope.hit.edu.cn

B. Wang

S.-Y. Du

Center for Composite Materials and
Electro-Optics Technology Center,
Harbin Institute of Technology,
P. O. Box 1247,
Harbin 150001, P. R. China

Investigation of Antiplane Shear Behavior of Two Collinear Permeable Cracks in a Piezoelectric Material by Using the Nonlocal Theory

[DOI: 10.1115/1.1445144]

1 Introduction

In the theoretical studies of crack problems, several different electric boundary conditions at the crack surfaces in piezoelectric materials have been proposed by numerous researchers ([1–4]). However, these solutions contain stress and electric displacement singularity. This is not reasonable according to the physical nature. To overcome the stress singularity in the classical elastic theory, Eringen [5] used the nonlocal theory to study the state of stress near the tip of a sharp line crack in an elastic plate subjected to antiplane shear. The solution did not contain any stress singularity. Recently, the same problems have been resolved in Zhou's papers ([6]) by using the Schmidt method.

In this paper, the behavior of two collinear symmetric cracks subjected to the antiplane shear loading in the piezoelectric materials is investigated by using the Schmidt method and the nonlocal theory for permeable crack surface conditions. The traditional concept of linear elastic fracture mechanics and the nonlocal theory are extended to include the piezoelectric effects. As expected, the solution in this paper does not contain the stress and electric displacement singularity at the crack tip.

2 Basic Equations of Nonlocal Piezoelectric Materials

As discussed in [7], for the antiplane shear problem, the basic equations of linear, nonlocal piezoelectric materials can be written as follows:

$$\frac{\partial \tau_{xz}}{\partial x}+\frac{\partial \tau_{yz}}{\partial y}=0 \tag{1}$$

$$\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}=0 \tag{2}$$

$$\tau_{kz}(X)=\int_V \alpha(|X'-X|)[c_{44}w_{,k}(X')+e_{15}\phi_{,k}(X')]dV(X'),\quad (k=x,y) \tag{3}$$

$$D_k(X)=\int_V \alpha(|X'-X|)[e_{15}w_{,k}(X')-\varepsilon_{11}\phi_{,k}(X')]dV(X'),\quad (k=x,y) \tag{4}$$

where the only difference from classical elastic theory is that in the stress and the electric displacement constitutive Eqs. (3)–(4), the stress $\tau_{zk}(X)$ and the electric displacement $D_k(X)$ at a point X depends on $w_{,k}(X)$ and $\phi_{,k}(X)$, at all points of the body. w and ϕ are the mechanical displacement and electric potential. $c_{44}, e_{15}, \varepsilon_{11}$ are the shear modulus, piezoelectric coefficient, and dielectric parameter, respectively. $\alpha(|X'-X|)$ is the influence function. As discussed in the papers ([5,6]), $\alpha(|X'-X|)$ can be assumed as follows:

$$\alpha(|X'-X|)=\frac{1}{\pi}(\beta/a)^2\exp[-(\beta/a)^2(X'-X)(X'-X)] \tag{5}$$

where β is a constant and a is the lattice parameter.

3 The Crack Model

Consider an infinite piezoelectric plane containing two collinear symmetric permeable cracks of length 1-b along the x-axis. $2b$ is the distance between two cracks. The boundary conditions of the present problem are

$$\tau_{yz}^{(1)}(x,0^+)=\tau_{yz}^{(2)}(x,0^-)=-\tau_0,\quad b\leqslant|x|\leqslant 1 \tag{6}$$

$$D_y^{(1)}(x,0^+)=D_y^{(2)}(x,0^-),\quad \phi^{(1)}(x,0^+)=\phi^{(2)}(x,0^-),\quad |x|\leqslant\infty \tag{7}$$

$$w^{(1)}(x,0^+)=w^{(2)}(x,0^-)=0,\quad 0<|x|<b, 1<|x| \tag{8}$$

$$w^{(k)}(x,y)=\phi^{(k)}(x,y)=0,\quad \text{for } (x^2+y^2)^{1/2}\to\infty,\quad (k=1,2). \tag{9}$$

Note that all quantities with superscript $k(k=1,2)$ refer to the upper half-plane and the lower half-plane.

As discussed in [7], the general solutions of Eqs. (1)–(2) satisfying (9) are, respectively,

$$w^{(1)}(x,y)=\frac{2}{\pi}\int_0^\infty A_1(s)e^{-sy}\cos(xs)ds,$$

$$\phi^{(1)}(x,y)-\frac{e_{15}}{\varepsilon_{11}}w^{(1)}(x,y)=\frac{2}{\pi}\int_0^\infty B_1(s)e^{-sy}\cos(xs)ds \tag{10}$$

$$w^{(2)}(x,y)=\frac{2}{\pi}\int_0^\infty A_2(s)e^{sy}\cos(xs)ds,$$

$$\phi^{(2)}(x,y)-\frac{e_{15}}{\varepsilon_{11}}w^{(2)}(x,y)=\frac{2}{\pi}\int_0^\infty B_2(s)e^{sy}\cos(xs)ds \tag{11}$$

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 17, 2001; final revision, Oct. 19, 2001. Associate Editor: A. K. Mal. Discussion on the paper should be addressed to the Editor, Prof. Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

where $A_1(s)$, $B_1(s)$, $A_2(s)$, $B_2(s)$ are to be determined from the boundary conditions.

For solving the problem, the gap functions of the crack surface displacements and the electric potentials are defined as follows:

$$f_w(x)=w^{(1)}(x,0^+)-w^{(2)}(x,0^-) \tag{12}$$

$$f_\phi(x)=\phi^{(1)}(x,0^+)-\phi^{(2)}(x,0^-). \tag{13}$$

Substituting Eqs. (10)–(11) into Eqs. (3)–(4), (12)–(13), applying the Fourier transform and the boundary conditions (6)–(8), it can be obtained as

$$\frac{1}{\pi}\int_0^\infty s\bar{f}_w(s)\mathrm{erfc}(\varepsilon s)\cos(sx)ds=\frac{\tau_0}{c_{44}},\quad b\leqslant|x|\leqslant 1 \tag{14}$$

$$\frac{1}{\pi}\int_0^\infty \bar{f}_w(s)\cos(sx)ds=0,\quad 0<|x|<b,\quad 1<|x|<\infty \tag{15}$$

and $\bar{f}_\phi(s)=0$, $f_\phi(x)=0$, for all s and x. $\varepsilon=a/2\beta$, $\mathrm{erfc}(z)=1-\Phi(z)$, $\Phi(z)=2/\sqrt{\pi}\int_0^z \exp(-t^2)dt$.

4 Solution of the Triple Integral Equations

As discussed in [6], the Schmidt method ([8]) can be used to solve the triple-integral Eqs. (14)–(15). The gap functions of the crack surface displacement can be represented by the following series:

$$f_w(x)=\sum_{n=0}^\infty a_n P_n^{(1/2,1/2)}\left(\frac{x-\frac{1+b}{2}}{\frac{1-b}{2}}\right)\times\left(1-\frac{\left(x-\frac{1+b}{2}\right)^2}{\left(\frac{1-b}{2}\right)^2}\right)^{1/2},$$

$$\text{for } b\leqslant x\leqslant 1,\ y=0 \tag{16}$$

$$f_w(x)=0,\quad \text{for } 0<x<b,\ 1<x,\ y=0 \tag{17}$$

where a_n is unknown coefficients to be determined and $P_n^{(1/2,1/2)}\times(x)$ is a Jacobi polynomial. The Fourier transformation of Eq. (16) is

$$\bar{f}_w(s)=\sum_{n=0}^\infty a_n Q_n G_n(s)\frac{1}{s}J_{n+1}\left(s\frac{1-b}{2}\right) \tag{18}$$

$$Q_n=2\sqrt{\pi}\frac{\Gamma\left(n+1+\frac{1}{2}\right)}{n!},$$

$$G_n(s)=\begin{cases}(-1)^{n/2}\cos\left(s\dfrac{1+b}{2}\right), & n=0,2,4,6,\ldots\\ (-1)^{(n+1)/2}\sin\left(s\dfrac{1+b}{2}\right), & n=1,3,5,7,\ldots\end{cases} \tag{19}$$

where $\Gamma(x)$ and $J_n(x)$ are the Gamma and Bessel functions, respectively. By substituting Eq. (18) into Eqs. (14)–(15), respectively, Eq. (15) can be automatically satisfied. Then the remaining Eq. (14) reduces to the form

$$\sum_{n=0}^\infty a_n Q_n\int_0^\infty \mathrm{erfc}(\varepsilon s)G_n(s)J_{n+1}\left(s\frac{1-b}{2}\right)\cos(sx)ds=\frac{\pi}{c_{44}}\tau_0. \tag{20}$$

Equations (20) can now be solved for the coefficients a_n by the Schmidt method ([8]).

5 Numerical Calculations and Discussion

τ_{yz} and D_y along the crack line can be expressed as

$$\tau_{yz}=\tau_{yz}^{(1)}(x,0)=-\frac{c_{44}}{\pi}\sum_{n=0}^\infty a_n Q_n\int_0^\infty \mathrm{erfc}(\varepsilon s)G_n(s)J_{n+1}\left(s\frac{1-b}{2}\right)\cos(xs)ds \tag{21}$$

$$D_y=D_y^{(1)}(x,0)=-\frac{e_{15}}{\pi}\sum_{n=0}^\infty a_n Q_n\int_0^\infty \mathrm{erfc}(\varepsilon s)G_n(s)J_{n+1}\left(s\frac{1-b}{2}\right)\cos(xs)ds=\frac{e_{15}}{c_{44}}\tau_{yz}^{(1)}(x,0). \tag{22}$$

So long as $\varepsilon\neq 0$, the semi-infinite integration and the series in Eqs. (20) is convergent for any variable x. Equations (21) and (22) give finite stress and electric displacement all along $y=0$, so there are no stress and electric displacement singularity at the crack tips. The results are plotted in Figs. 1 and 2. From the results, the dimensionless stress field is found to be independent of the material parameters. They just depend on the length of the crack and the lattice parameter. However, the electric displacement field is found to depend on the stress loads, the shear modulus, the length of the crack, the lattice parameter and piezoelectric coefficient except the dielectric parameter ε_{11}. Contrary to the impermeable crack surface condition solution, it is found that the electric displacement for the permeable crack surface conditions is much smaller than the results for the impermeable crack surface conditions.

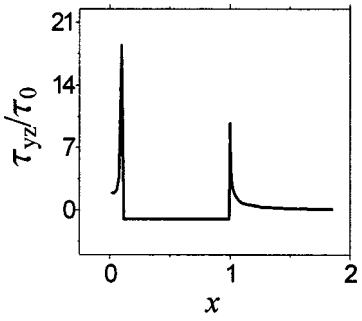


Fig. 1 The stress along the crack line versus x for $b=0.1$, $a/2\beta=0.0005$ (PZT-5H)

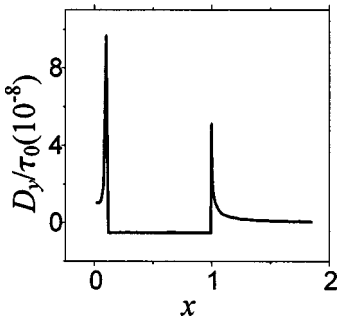


Fig. 2 The electric displacement along the crack line versus x for $b=0.1$, $a/2\beta=0.0005$ (PZT-5H)

References

[1] Deeg, W. E. F., 1980, The Analysis of Dislocation, Crack and Inclusion Problems in Piezoelectric Solids, Ph.D. thesis, Stanford University, Stanford, CA.

[2] Pak, Y. E., 1992, ''Linear Electro-Elastic Fracture Mechanics of Piezoelectric Materials,'' Int. J. Fract., **54**, pp. 79–100.

[3] Suo, Z., Kuo, C.-M., Barnett, D. M., and Willis, J. R., 1992, ''Fracture Mechanics for Piezoelectric Ceramics,'' J. Mech. Phys. Solids, **40**, pp. 739–765.

[4] Zhang, T. Y., and Hack, J. E., 1992, ''Mode-III Cracks in Piezoelectric Materials,'' J. Appl. Phys., **71**, pp. 5865–5870.

[5] Eringen, A. C., 1979, ''Linear Crack Subject to Antiplane Shear,'' Eng. Fract. Mech., **12**, pp. 211–219.

[6] Zhou, Z. G., Han, J. C., and Du, S. Y., 1999, ''Investigation of a Crack Subjected to Antiplane Shear by Using the Nonlocal Theory,'' Int. J. Solids Struct., **36**, No. 26, pp. 3891–3901.

[7] Zhou, Z. G., and Wang, B., 2001, ''Investigation of Antiplane Shear Behavior of Two Collinear Cracks in the Piezoelectric Materials by Using the Nonlocal Theory,'' Int. J. Solids Struct., submitted for publication.

[8] Morse, P. M., and Feshbach, H., 1958, *Methods of Theoretical Physics*, Vol. 1, McGraw-Hill, New York.

Journal of Applied Mechanics

Brief Notes

A Brief Note is a short paper that presents a specific solution of technical interest in mechanics but which does not necessarily contain new general methods or results. A Brief Note should not exceed 1500 words *or equivalent* (a typical one-column figure or table is equivalent to 250 words; a one line equation to 30 words). Brief Notes will be subject to the usual review procedures prior to publication. After approval such Notes will be published as soon as possible. The Notes should be submitted to the 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.

Propagation of a Shear Direction Acoustic Wave in Piezoelectric Coupled Cylinders

Q. Wang
Mem. ASME

J. Jin

S. T. Quek

Department of Civil Engineering, National University of Singapore, Singapore 119260

1 Introduction

Acoustic wave polarized in the horizontal shear direction (SH wave) have attracted much attention in the topic of smart materials research because of the following reason ([1–3]). The particle displacement associated with SH wave has almost zero component normal to the free surface. The absence of the surface normal component avoids fast dispersion of energy into surrounding medium when the subject is immersed in liquids ([4]). The properties of SH wave traveling in the piezoelectric coupled cylindrical structure are however quite different from that in plate or beam structures ([5,6]). In the case of cylindrical geometry, SH wave is polarized along the axis of the cylinder with zero particle displacement component normal to the cylindrical surface. The present paper is concerned with SH wave circulating around a long metallic cylinder covered with a piezoelectric layer. Dispersion relations are obtained, which provides potential application of this structure in the design of nondestructive evaluation of cylindrical structures submerged in water.

2 Formulation of mechanical model

The model under consideration involves a long metallic cylinder of radius a covered by a piezoelectric material of uniform thickness h (see Fig. 1). The equations of motion can be written in cylindrical coordinate system as

$$\frac{\partial T_{rz}}{\partial r}+\frac{1}{r}\frac{\partial T_{\theta z}}{\partial \theta}+\frac{1}{r}T_{rz}=\rho\ddot{u}_z, \tag{1}$$

where ρ is the mass density and dot denotes time derivative.

For an SH wave $u_r=u_\theta=0$ and u_z is independent of z. Hence, the nonzero strain components are

$$S_{rz}=\frac{\partial u_z}{\partial r} \text{ and } S_{\theta z}=\frac{1}{r}\frac{\partial u_z}{\partial \theta}. \tag{2}$$

The pertinent constitutive relations for the piezoelectric layer can be reduced to

$$T_{\theta z}=c_{44}S_{\theta z}-e_{15}E_\theta \tag{3a}$$

$$T_{rz}=c_{44}S_{rz}-e_{15}E_r \tag{3b}$$

$$D_r=e_{15}S_{rz}+\epsilon_{11}E_r \tag{3c}$$

$$D_\theta=e_{15}S_{\theta z}+\epsilon_{11}E_\theta \tag{3d}$$

where c_{44}, e_{15}, and ϵ_{11} are the elastic, piezoelectric, and dielectric constants of the piezoelectric layer, respectively, T denotes stress, E the electric field, and D the electric displacements.

By employing the quasi-static approximation, the relationship between electric field E and electric potential ϕ can be reduced to

$$E=-\nabla\phi \tag{4}$$

and the electric displacement can be derived as

$$\nabla\cdot D=0. \tag{5}$$

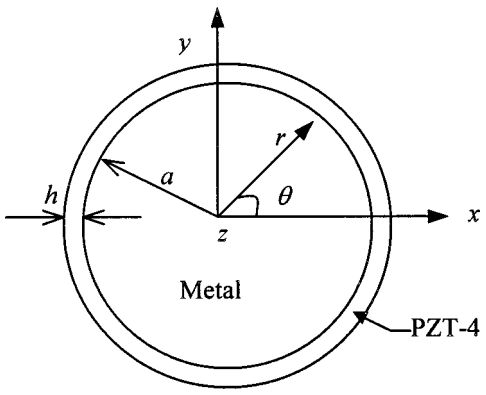


Fig. 1 Geometry of the layered long cylinder

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Feb. 15, 2001; final revision, Aug. 29, 2001. Associate Editor: A. K. Mal.

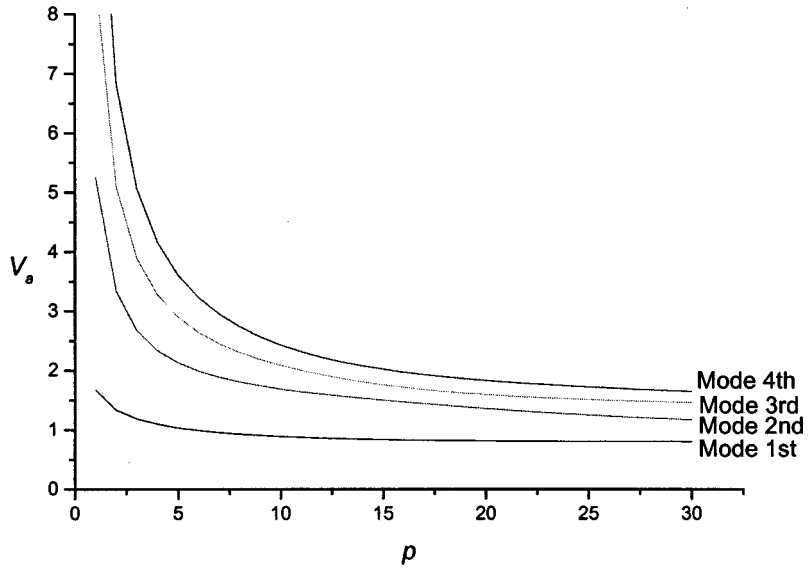


Fig. 2 Dispersion curves for the first and fourth modes

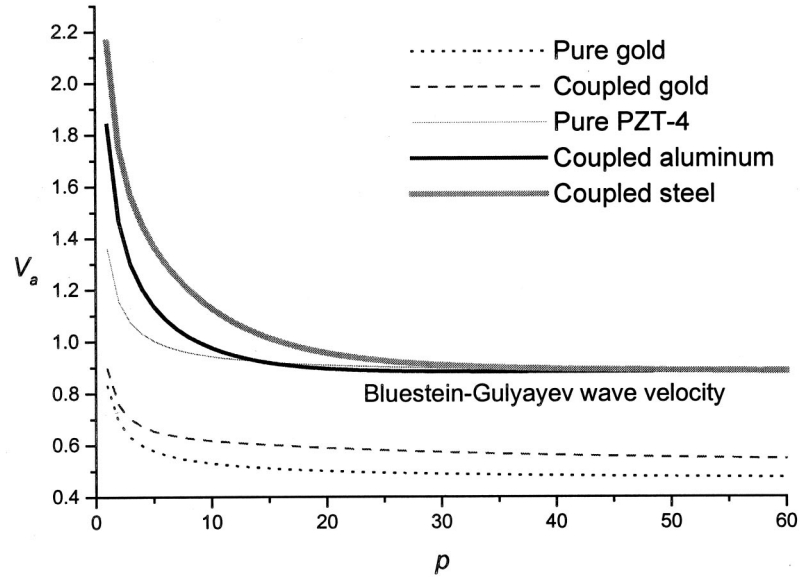


Fig. 3 Dispersion curves of cylinder with different core material

Substituting (2)–(4) into (1) and (5) yield the electromechanical coupled governing equations,

$$c_{44}\nabla^2 u_z + e_{15}\nabla^2\phi = \rho\ddot{u}_z \tag{6}$$

$$e_{15}\nabla^2 u_z - \epsilon_{11}\nabla^2\phi = 0, \tag{7}$$

which can be simplified as

$$\overline{c_{44}}\nabla^2 u_z = \rho\ddot{u}_z \tag{8}$$

where $\overline{c_{44}} = c_{44} + e_{15}^2/\epsilon_{11}$ and

$$\nabla^2\varphi = 0 \tag{9}$$

where $\varphi = \phi - (e_{15}/\epsilon_{11})u_z$.

The stress can be written as

$$T_{rz} = \overline{c_{44}}\frac{\partial_z}{\partial r} + e_{15}\frac{\partial\varphi}{\varphi r}. \tag{10}$$

The corresponding equations for the metallic cylinder core are

$$\mu'\nabla^2 u_z' = \rho'\ddot{u}_z' \tag{11}$$

$$T_{rz}' = \mu'\frac{\partial u_z'}{\partial r} \tag{12}$$

Table 1 Material properties

	Aluminum	Steel	Gold	PZT-4
Mass density ($\times 10^3$ kg/m^3)	$\rho'=2.8$	$\rho'=7.8$	$\rho'=19$	$\rho=7.5$
Shear modulus ($\times 10^{10}$ N/m^2)	$\mu'=2.7$	$\mu'=7.6$	$\mu'=2.6$	$c_{44}=2.6$
e_{15} (k/m^2)	-	-	-	12.7
ϵ_{11} ($\times 10^{-9}$ ϕ/m)	-	-	-	6.45
c_2 ($\times 10^3$ m/s)	3.11	3.12	1.17	2.60

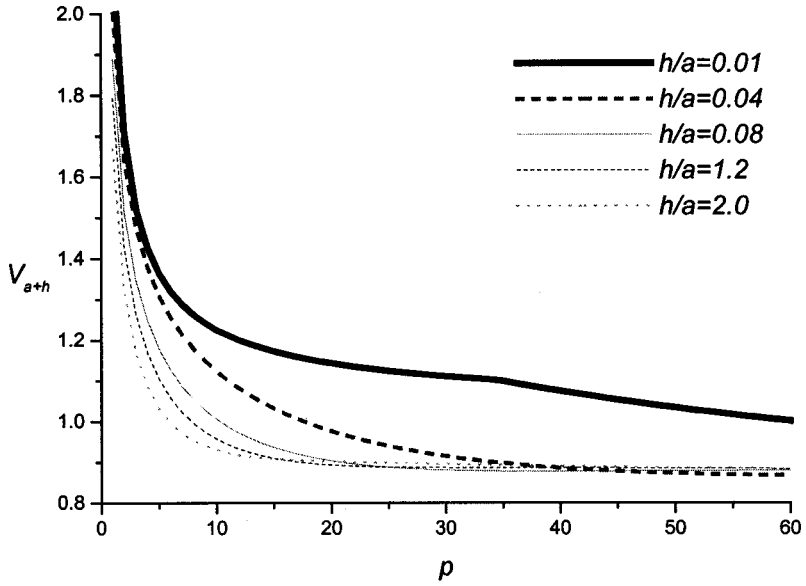


Fig. 4 Dispersion curves of an aluminum cylinder with a piezoelectric layer of different thickness

where μ' is the shear modulus and prime denotes variables of the core.

It is assumed that the poling direction of the PZT-4 layer coincides with the z-axis and the outside surface of the piezoelectric layer is electrically shorted. Hence, the electrical potential throughout both sides of the piezoelectric layer is zero. The outside surface of the layer is stress free and the continuity conditions should be satisfied at the interface. The boundary conditions can thus be written as

$$\phi=0, \quad \text{at } r=a+h \tag{13a}$$

$$T_{rz}=0, \quad \text{at } r=a+h \tag{13b}$$

$$\phi=0, \quad \text{at } r=a \tag{14a}$$

$$u_z=u_z', \quad \text{at } r=a \tag{14b}$$

$$T_{rz}=T_{rz}', \quad \text{at } r=a. \tag{14c}$$

3 Dispersion Relations

The harmonic wave solution for both the cylinder core and the piezoelectric layer can be represented as

$$u_z'(r,\theta,t)=U_z'(r)e^{i(p\theta-\omega t)}, \quad 0\leqslant r\leqslant a \tag{15}$$

$$u_z(r,\theta,t)=U_z(r)e^{i(p\theta-\omega t)}, \quad a\leqslant r\leqslant a+h \tag{16}$$

$$\varphi_z(r,\theta,t)=\psi_z(r)e^{i(p\theta-\omega t)}, \quad a\leqslant r\leqslant a+h \tag{17}$$

where P is the angular wave number and ω is the angular frequency. Substituting (15)–(17) into the governing Eq. (8), (9), and (11) gives

$$\frac{d^2U_z'}{dr^2}+\frac{1}{r}\frac{dU_z'}{dr}+\left(q_l^2-\frac{p^2}{r^2}\right)U_z'=0 \tag{18}$$

$$\frac{d^2U_z}{dr^2}+\frac{1}{r}\frac{dU_z}{dr}+\left(q_c^2-\frac{p^2}{r^2}\right)U_z=0 \tag{19}$$

$$\frac{d^2\psi_z}{dr^2}+\frac{1}{r}\frac{d\psi_z}{dr}-\frac{pl^2}{r^2}\psi_z=0 \tag{20}$$

where $q_1^2=\rho'\omega^2/\mu$, and $q_c^2=\rho\omega^2/\overline{c_{44}}$.

Solutions for (18)–(20) are

$$u_z'=AJ_p(q_lr)e^{i(p\theta-\omega t)}, \qquad 0\leqslant r\leqslant a \tag{21}$$

$$u_z=[BJ_p(q_cr)+CY_p(q_cr)]e^{i(p\theta-\omega t)}, \qquad a\leqslant r\leqslant a+h \tag{22}$$

$$\varphi=(Dr^p+Er^{-p})e^{i(p\theta-\omega t)}, \qquad a\leqslant r\leqslant a+h \tag{23}$$

where $J_p(z)$ and $Y_p(z)$ are Bessel function of the first and second kind. Utilizing boundary conditions of (13)–(14) yield a system of five equations as follows:

$$\begin{bmatrix} 0 & \frac{e_{15}}{\epsilon_{11}}J_p(\beta V_a p) & \frac{e_{15}}{\epsilon_{11}}Y_p(\beta V_a p) & a^p\beta^p & a^{-p}\beta^{-p} \\ 0 & \overline{c_{44}}\frac{\partial J_p(\beta V_a p)}{\partial r} & \overline{c_{44}}\frac{\partial Y_p(\beta V_a p)}{\partial r} & e_{15}pa^{p-1}\beta^{p-1} & -e_{15}pa^{-p-1}\beta^{-p-1} \\ 0 & \frac{e_{15}}{\epsilon_{11}}J_p(V_a p) & \frac{e_{15}}{\epsilon_{11}}Y_p(V_a p) & a^p & a^{-p} \\ J_p(\alpha V_a p) & -J_p(V_a p) & -Y_p(V_a p) & 0 & 0 \\ \mu'\frac{\partial J_p(\alpha V_a p)}{\partial r} & -\overline{c_{44}}\frac{\partial J_p(V_a p)}{\partial r} & -\overline{c_{44}}\frac{\partial Y_p(V_a p)}{\partial r} & -e_{15}pa^{p-1} & e_{15}pa^{-p-1} \end{bmatrix} \begin{Bmatrix} A \\ B \\ C \\ D \\ E \end{Bmatrix} = 0 \tag{24}$$

where $\alpha=(p'\overline{c_{44}}/\rho\mu')^{1/2}$, $\beta=(1+h/a)$. The phase velocity at $r=a$ is given by $v_a=\omega a/p$, which can be nondimensionalized as

$$V_a=v_a(\overline{c_{44}}/\rho)^{-1/2}. \tag{25}$$

The dispersion equation is obtained by setting the determinant of coefficients of (24) to be zero.

4 Results and Discussions

The case of an aluminum cylindrical core of radius 0.1 m wrapped with a PZT-4 piezoelectric layer of thickness 0.01 m (i.e., $h/a=0.1$ is considered. Using (24), the dispersion curves V_a as a function of p for the first four mode shapes are computed and shown in Fig. 2. It can be seen that as the wave number increases, the phase velocity of each mode decreases. When the wave number is large, the curves approach asymptotically to the Bluestein-Gulyayev wave velocity ([6]). This is consistent in view of the fact that when wave number is large relative to the curvature of the cylindrical surface, the cylindrical problem can be reduced to a plane problem.

4.1 Dispersion Curves of Cylinder With Different Metal Core. To study the significance of the coupling effect due to different metallic core, three materials, namely, aluminum, steel, and gold are considered. The dispersion curves for the first mode are plotted in Fig. 3. For comparison purpose, the dispersion curves for a pure PZT-4 cylinder and a pure gold cylinder are also plotted.

It is noted that the phase velocity of both the coupled steel cylinder and the coupled aluminum cylinder approaches the Bluestein-Gulyayev wave velocity, since the softer material is the piezoelectric layer. While the asymptotic velocity of the coupled gold cylinder is relatively lower, since gold is slightly softer and denser.

For low wave number, it is noted that the phase velocities of the coupled cylinders follow closely that of the respective rotational wave propagation velocity c_2 of core materials, where $c_2'=\sqrt{\mu'/\rho'}$ for metallic material and $c_2=\sqrt{\overline{c_{44}}/\rho}$ for piezoelectric material. The latter values are tabulated in Table 1 and it can be seen that $c_2^s>c_2^a>c_2^p>c_2^g$, where c_2^s, c_2^a, c_2^p, c_2^g, are rotational wave velocities of steel, aluminum, PZT-4, and gold, respectively. Due to the relatively lower c_2 of the piezoelectric material, the coupling effect contributed by the piezoelectric layer lowers the phase velocity of the coupled aluminum cylinder or that of the coupled steel cylinder. In contrast, since $c_2^p>c_2^g$, the coupling effect of the piezoelectric layer induces a higher phase velocity for the coupled gold cylinder with the dispersion curve of the pure piezoelectric cylinder as the upper bound.

4.2 Dispersion Curves of Different Thickness of Piezoelectric layer. From (24), the radius a can be removed from the dispersion equation of the controlling geometrical parameter is the thickness ratio, h/a. The nondimensional phase velocity at $r=a+h$, $V_{a+h}=V_a(a+h)/a$, is used to study the effect of the thickness of the piezoelectric layer by varying h/a. The dispersion curves (based on V_{a+h}) of the coupled aluminum cylinder with piezoelectric layer of different thicknesses are plotted in Fig. 4.

It is noted that as the thickness of piezoelectric layer increases, the phase velocity decreases. This can be explained by the fact that c_2 of piezoelectric layer is slower than that of aluminum core. As the thickness of the layer increases, the coupling effect increases resulting in a lower phase velocity. It can be observed that the Bluestein-Gulyayev wave velocity is reached at lower wave number as the thickness increases. However, the lower bound is limited by the dispersion curve of a pure PZT-4 cylinder given in Fig. 3.

References

[1] Curtis, R. G., and Redwood, M., 1973, "Transverse Surface Waves on a Piezoelectric Material Carrying a Metal Layer of Finite Thickness," J. Appl. Phys. **44**, pp. 2002–2007.

[2] Lee, C. K., and Moon, F. C., 1989, "Laminated Piezopolymer Plates for Torsion and Bending Sensors and Actuators," J. Acoust. Soc. Am. **85**, pp. 2432–2439.

[3] Sun, C. T., and Zhang, X. D., 1995, "Use of Thickness Shear Mode in Adaptive Sandwich Structures," Smart Mater. Struct. **4**, pp. 202–206.

[4] Boris, D. Z., and Shrinivas, G. J., 1999, "Propagation of QSH (Quasi Shear Horizontal) Acoustic Waves in Piezoelectric Plates," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, pp. 1298–1302.

[5] Sun, C. T. and Cheng, N. C., 1974, "Piezoelectric Waves on a Layered Cylinder," J. Appl. Phys. **45**, pp. 4288–4294.

[6] Chen, C. L., 1973, "On Electroacoustic Waves Guided by a Cylindrical Piezoelectric Interface," J. Appl. Phys. **44**, p. 3841–3847.

A Generalized Load-Penetration Relation for Sharp Indenters and the Indentation Size Effect

Z. Y. Li[1]
School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47907-1287

S. Chandrasekar
Professor, Mem. ASME

Schools of Industrial Engineering and Materials Engineering, 1287 GRIS, Purdue University, West Lafayette, IN 47907-1287

H. T. Yang
Professor and Chancellor, Department of Mechanical and Environmental Engineering, University of California, Santa Barbara, CA 93106. Mem. ASME

A dimensional analysis has been made of elastic-plastic indentation of an anisotropic solid, and of a solid showing pressure-sensitive yield behavior. It is found that, $P\propto\delta^2$, for indentation with sharp, self-similar indenters, where P is the load applied by the indenter and δ is the corresponding distance of penetration of the indenter into the solid. This extends and generalizes a similar result obtained for isotropic solids showing conventional plastic behavior. When a strain-gradient plasticity is incorporated into the material model, then it is found that P is no longer proportional to δ^2. Implications of the results for the indentation size-effect and for the determination of stress-strain curves from indentation are discussed. [DOI: 10.1115/1.1458557]

Introduction

Measurement of hardness involves loading a hard sphere, cone, or pyramid indenter into the surface of a material to produce a well-developed plastic zone and measuring the resulting indentation size. A hardness value is obtained by dividing the applied load (P) by the projected area (or surface area) of the indentation.

[1]Presently at MMC Technology Corporation, San Jose, CA.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Jan. 25, 2001; final revision, Oct. 5, 2001. Associate Editor: A. K. Mal.

Because of the relative simplicity of hardness measurement and the semi-empirical relation that exists between hardness and the yield strength of a material, indentation hardness has found widespread use as a technique for characterizing the plastic properties of solids ([1,2]).

A recent trend in indentation hardness testing is the use of depth-sensing indentation methods. In depth-sensing indentation, a hardness value is derived from a measurement of the penetration of the indenter into the material rather than from a measurement of the indentation area. An important example of this approach is nano-indentation. The use of depth-sensing indentation for hardness measurement requires a characterization of the relation(s) that exists between applied load (P), distance of penetration (δ) of indenter into the material and characteristic dimensions of the indent ([3]). The determination of this relation(s) should also facilitate efforts aimed at determining the stress-strain curve of a solid from a measurement of the load-penetration curve during indentation ([4,5]).

It has been shown recently, using dimensional analysis, that $P \propto \delta^2$ for elastic-plastic indentation of an isotropic solid by sharp (self-similar) indenters ([6]). Here, we show that a similar result, $P \propto \delta^2$, holds also for the more general case of elastic-plastic indentation of anisotropic solids, and of solids which show pressure-sensitive yield behavior. When a strain-gradient plasticity model is incorporated into the analysis, it is found that $P \propto \delta^2$ is no longer valid. Implications of these results for the "indentation size effect," depth sensing indentation and determination of stress-strain curves from indentation are discussed.

Anisotropic Solid

Consider the elastic-plastic indentation of an anisotropic half-space by a rigid, sharp conical indenter having a semi-apical angle, θ. The generalized Hooke's law for elastic deformation is

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \tag{1}$$

where C_{ijkl} is the stiffness tensor of the fourth order. We use the following yield criterion proposed by Hill [7], to describe the onset of plastic flow in an anisotropic solid:

$$\begin{aligned} 2f(\sigma_{ij}) \equiv A(\sigma_y-\sigma_z)^2 + B(\sigma_z-\sigma_x)^2 + C(\sigma_x-\sigma_y)^2 + 2D\tau_{yz}^2 \\ + 2E\tau_{zx}^2 + 2F\tau_{xy}^2 = 1 \end{aligned} \tag{2}$$

where A,B,C,D,E,F are material parameters characteristic of the current state of anisotropy. Linear terms are not included in (2) since it is assumed that the material shows no Bauschinger effect.

The Levy-Mises formulation for the stress-strain increment relations during plastic flow is

$$\begin{aligned} d\varepsilon_x &= d\lambda[C(\sigma_x-\sigma_y) + B(\sigma_x-\sigma_z)] \\ d\varepsilon_y &= d\lambda[A(\sigma_y-\sigma_z) + C(\sigma_y-\sigma_x)] \\ d\varepsilon_z &= d\lambda[B(\sigma_z-\sigma_x) + A(\sigma_z-\sigma_y)] \\ d\gamma_{yz} &= d\lambda D\tau_{yz} \\ d\gamma_{zx} &= d\lambda E\tau_{zx} \\ d\gamma_{xy} &= d\lambda F\tau_{xy} \end{aligned} \tag{3}$$

where $d\lambda$ is a material parameter that can be obtained from uniaxial stress-strain data.

The indentation load P is completely determined by the constants, C_{ijkl}, the parameters, A,B,C,D,E,F and $d\lambda$, the penetration distance, δ, of the indenter into the half-space, the coefficient of friction, μ, between indenter and solid, and the cone semi-apical angle (θ). Thus, we have

$$P = f(C_{ijkl}, A,B,C,D,E,F,d\lambda,\delta,\mu,\theta). \tag{4}$$

Dimensional formulas for each of the above parameters may be written in terms of the independent dimensions of C_{1111} and δ as

Quantity	Dimensional Formula
P	$[C_{1111}]^1[\delta]^2$
C_{ijkl}	$[C_{1111}]^1[\delta]^0$
A,B,C,D,E,F	$[C_{1111}]^{-2}[\delta]^0$
$d\lambda$	$[C_{1111}]^1[\delta]^0$
δ	$[C_{1111}]^0[\delta]^1$
μ	$[C_{1111}]^0[\delta]^0$
θ	$[C_{1111}]^0[\delta]^0$.

Using the Π theorem of dimensional analysis ([8]) with C_{1111} and δ as the independent dimensions we get

$$P = C_{1111}\delta^2\Pi\left(\frac{C_{ijkl}}{C_{1111}}, \frac{A}{C_{1111}^{-2}}, \frac{B}{C_{1111}^{-2}}, \frac{C}{C_{1111}^{-2}}, \frac{D}{C_{1111}^{-2}}, \frac{E}{C_{1111}^{-2}}, \frac{F}{C_{1111}^{-2}}, \frac{d\lambda}{C_{1111}}, \mu, \theta\right) \tag{5}$$

where $\Pi(\cdot)$ is a function of the dimensionless parameters enclosed in brackets. Thus, $P \propto \delta^2$, for cone indentation of an anisotropic half-space.

If we set

$$D = E = F = 3A = 3B = 3C,$$

$$C_{1111} = C_{2222} = C_{3333} = (1-\nu)E/(1+\nu)(1-2\nu),$$

$$C_{1122} = C_{1133} = C_{2233} = \nu E/(1+\nu)(1-2\nu),$$

$$C_{2323} = C_{1313} = C_{1212} = E/2(1+\nu)$$

and the other $C_{ijlk} = 0$ in Eq. (5), then an isotropic material results with E and ν being its Young's modulus and Poisson's ratio, respectively. This is the case studied by Cheng and Cheng [6].

Solid With Pressure-Sensitive Yield Behavior

Plasticity in some materials such as glasses and ceramics is often influenced by hydrostatic pressure. For such materials, the Drucker-Prager [9] yield criterion often provides a good description of the onset of yield. The Drucker-Prager yield surface is given by

$$F(J_1, J_2') = \sqrt{J_2'} - \beta J_1 - \kappa = f_\xi(J_1, J_2') - \kappa = 0 \tag{6}$$

where $f_\xi(J_1, J_2') = \sqrt{J_2'} - \beta J_1$, $J_1 = \sigma_{ii}$, $J_2' = 1/2\, S_{ij}S_{ij}$ and $S_{ij} = \sigma_{ij} - p\,\delta_{ij}$, p is the hydrostatic component of the stress state, and β and κ are material constants with β describing the pressure sensitivity.

Using the associated flow rule with the yield function set equal to the plastic potential, the plastic strain increments are

$$d\varepsilon_{ij}^p = d\lambda\,\frac{\partial f_\xi}{\partial \sigma_{ij}} \tag{7}$$

where $d\lambda$ is the aforementioned material parameter. If we con-

sider indentation of this pressure-sensitive material by a cone, then the load applied by the indenter can be expressed as

$$P = f(E, \nu, \beta, K, d\lambda, \delta, \theta) \tag{8}$$

where E and ν are, respectively, the Young's modulus and Poisson's ratio of the material. By applying the Π theorem as before, with E and δ as the independent dimensions, we get

$$P = E\delta^2 \Pi\left(\beta, \frac{K}{E}, \nu, d\lambda, \theta\right) \tag{9}$$

where $\Pi(\cdot)$ is a function of the dimensionless parameters enclosed in brackets. Thus, $P \propto \delta^2$ is seen to hold also for this case.

The analysis leading to Eqs. (5) and (8) carries over with little modification for other self-similar indenters, e.g., wedge, Vickers, and Knoop.

These results imply that $P \propto \delta^2$ should hold for a large class of materials. Furthermore, it is seen that the statements, $P \propto \delta^2$, and the hardness being independent of load for self-similar indenters, are equivalent. This equivalence poses a challenge to the derivation of the stress-strain curve of a material from the P versus δ relation obtained with a sharp, self-similar indenter.

Discussion

There is a wealth of evidence available from hardness tests with pyramid indenters which shows the hardness to be essentially independent of load when the size of the indent is sufficiently large ([10]). In fine-grained homogeneous materials, this observation appears to hold true for indentation widths in excess of about ten micrometers. Numerical simulations of elastic-plastic indentation have also shown the hardness to be independent of load (or $P \propto \delta^2$) for solids that obey conventional plasticity. Giannakopoulos and Larson [11] have analyzed the indentation of materials showing a pressure-sensitive yield behavior and noted that $P \propto \delta^2$ appears to be a good fit for both Vickers and Berkovitz indentations. These results are consistent with the findings of our dimensional analysis.

But how can $P \propto \delta^2$ be reconciled with observations of an indentation ''size effect''? The size effect refers to a dependence of hardness on indentation size that has been observed in experiments ([10]). Several explanations have been proposed to explain this effect including an error in hardness estimate due to imprecise knowledge of indenter tip geometry, the presence of oxide or deformed layers on surfaces, and a lack of knowledge of the exact load bearing area of an indentation due to material pile-up and sinking in. Recently, a strain gradient plasticity model has been proposed ([12]) to explain the indentation size effect. In this model, the size effect arises due to the combined effect of two factors: a dependence of the flow stress (σ) on a characteristic length scale for the material ($\hat{l}$); and the presence of strain gradients in the indentation field.

It is easy to incorporate the strain gradient plasticity model into our dimensional analysis by assuming the effective flow stress (σ) of the material in the presence of a strain gradient to be given by ([13])

$$\left(\frac{\sigma}{\sigma_0}\right)^2 = 1 + \hat{l}\chi \tag{10}$$

where σ_0 is the flow stress in the absence of a gradient and χ is the effective strain gradient; and assuming the material undergoes power-law hardening with

$$\sigma_0 = K\varepsilon^n \tag{11}$$

where n is a work hardening exponent, and K is a material constant. If we consider the indentation of this material by a cone, then the load applied by the indenter may be written as

$$P = f(E, \nu, K, n, \hat{l}, \delta, \theta). \tag{12}$$

Using the Π-theorem as before with E and δ as independent dimensions, we get

$$P = E\delta^2 \Pi\left(\nu, \frac{K}{E}, n, \hat{l}/\delta, \theta\right). \tag{13}$$

Equation (13) shows that P is no longer proportional to δ^2 but is dependent on an additional parameter ($\hat{l}/\delta$) which involves the indenter penetration depth (δ).

Thus, the hardness of the material is no longer independent of indentation size, but is dependent among other factors on ($\hat{l}/\delta$), thereby, giving rise to an indentation size effect. It should be possible to establish the form of the dependence of P on $\hat{l}/\delta$ by numerical simulation and/or indentation experiments.

Nix and Gao [12] indicate that $\hat{l}$ has values of around 10 μm for annealed copper. Since it is observed experimentally that hardness becomes independent of indent size (or load), for sufficiently large indents, Eq. (13) would imply that the parameter ($\hat{l}/\delta$) appears in the function $\Pi(\cdot)$ such that P/δ^2 approaches a constant value asymptotically as $\hat{l}/\delta \to 0$.

Conclusion

It has been shown using dimensional analysis that the indenter load (P) is proportional to the square of the indenter penetration distance (δ) for elastic-plastic indentation of solids with sharp self-similar indenters. This relation is found to apply for solids showing a wide range of yield behavior. The relation, however, is seen to break down when strain-gradient plasticity is incorporated into the material model.

Acknowledgment

We would like to acknowledge the National Science Foundation for grant DMI-9800920 that supported this work.

References

[1] Tabor, D., 1951, *The Hardness of Metals*, Clarendon Press, Oxford, UK.

[2] Hill, R., Storakers, B., and Zdunek, A. B., 1989, ''A Theoretical Study of the Brinell Hardness Test,'' Proc. R. Soc. London, Ser. A, **423**, p. 301.

[3] Oliver, W. C., and Pharr, G. M., 1992, ''An Improved Technique for Determining Hardness and Elastic Modulus Using Load and Displacement Sensing Indentation Experiments,'' J. Mater. Res., **7**, p. 1564.

[4] Myers, S. M., Knapp, J. A., Follstaedt, D. M., and Dugger, M. T., 1998, ''Mechanical Properties of Nickel Ion-Implanted With Titanium and Carbon and Their Relation to Microstructure,'' J. Appl. Phys., **83**, p. 1256.

[5] Cheng, Y.-T., and Cheng, C.-M., 1999, ''Can Stress-Strain Relationships be Obtained From Indentation Curves Using Conical and Pyramidal Indenters,'' J. Mater. Res., **14**, p. 3493.

[6] Cheng, Y.-T., and Cheng, C.-M., 1998, ''Scaling Approach to Conical Indentation in Elastic-Plastic Solids With Strain Hardening,'' J. Appl. Phys., **84**, p. 1284.

[7] Hill, R., 1950, *The Mathematical Theory of Plasticity*, Oxford University Press, Oxford, UK.

[8] Bridgman, P. W., 1931, *Dimensional Analysis*, Yale University Press, New Haven, CT.

[9] Drucker, D. C., and Prager, W., 1952, ''Soil Mechanics and Plastic Analysis or Limit Design,'' Q. Appl. Math., **10**, p. 157.

[10] Ma, Q., and Clarke, D. R., 1995, ''Size Dependent Hardness of Silver Single Crystals,'' J. Mater. Res., **10**, No. 4, p. 853.

[11] Giannakopoulos, A. E., and Larsson, P. L., 1997, ''Analysis of Pyramid Indentation of Pressure-Sensitive Hard Metals and Ceramics,'' Mech. Mater., **25**, p. 1.

[12] Nix, W. D., and Gao, H., 1998, ''Indentation Effects in Crystalline Materials: A Law for Strain Gradient Plasticity,'' J. Mech. Phys. Solids, **46**, p. 411.

Resonance and the Aging Spring

T. J. Lardner
Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, MA 01003-5202. e-mail: lardner@ecs.umass.edu. Fellow ASME

The steady-state response of the classic mass-spring-dashpot model when the spring stiffness decays exponentially with time (an aging spring) and the system is excited by a forcing function whose frequency is equal to the natural frequency of the system with the constant initial stiffness is investigated. The steady-state response is obtained in terms of the damping and decay constants of the system and exhibits an oscillation about a nonzero value.
[DOI: 10.1115/1.1458559]

1 Introduction

The classical mass-spring-dashpot model used to describe vibration characteristics of simple mechanical systems under time-dependent loading is well known. When the forcing function is harmonic with a frequency equal to the natural frequency of the undamped system, the system is said to be in resonance and the steady-state response is harmonic with an amplitude dependent on the amount of damping in the system. The amplitude of the response of the system at other values of the forcing frequency is less than that at resonance and often is of less concern.

The purpose of this brief note is to determine the steady-state response of the mass-spring-dashpot model when the value of the spring stiffness decays exponentially with time from a constant initial value to zero and the system is excited by a forcing function with a frequency equal to the natural frequency of the system with the initial stiffness. The spring in this case is called an *aging spring* ([1,2]). This model has a simple interpretation of a physical system becoming less stiff in time while excited by a loading that would lead to resonance if the stiffness remained constant. The problem is of interest because of the importance of understanding resonant systems and because the results for the steady-state response are somewhat unexpected.

2 Formulation

The equation of motion for the displacement $u(\tau)$ of a mass in a mass-spring-dashpot system is

$$m\ddot{u}+c\dot{u}+k(\tau)u=p(\tau) \tag{1}$$

where $m,c,k(\tau)$ are the mass, dashpot constant, spring stiffness that is a function of the physical time τ, and $p(\tau)$ is the forcing function. We are interested in the case of resonance in which the forcing function is a sine function with a frequency equal to the natural frequency of the the system with an undamped constant spring stiffness. The spring stiffness is assumed to decay in time as (an *aging spring* ([1,2])),

$$k(\tau)=k_0\exp[-\hat{\epsilon}\tau] \tag{2}$$

where k_0 is the spring stiffness at time zero and $\hat{\epsilon}$ is the time decay constant. Exponential decay is selected because of the convenience in the analysis ([1,2]) and the expectation that the stiffness of a physical system will asymptote to small values in time. In the absence of decay, $\hat{\epsilon}=0$, the natural frequency of the undamped system is

$$\omega=\sqrt{k_0/m}. \tag{3}$$

For resonance loading the equation of motion becomes

$$m\ddot{u}+c\dot{u}+k_0\exp[-\hat{\epsilon}\tau]=p_0\sin\omega\tau \tag{4}$$

with initial conditions prescribed for $u(0)$ and $\dot{u}(0)$.

The amount of damping in the system is defined by the damping ratio $J=c/2\sqrt{k_0m}$ which is the amount of damping relative to critical damping in a system in which $\hat{\epsilon}=0$. It is convenient to nondimensionalize Eq. (4) by writing

$$y=\frac{u}{u_0},\quad J=\frac{c}{2\sqrt{k_0m}},\quad \omega=\sqrt{k_0/m}$$
$$t=\omega\tau,\quad u_0=p_0/k_0,\quad \epsilon=\hat{\epsilon}/\omega \tag{5}$$

where t is the nondimensional time. Equation (4) becomes

$$\ddot{y}+2J\dot{y}+e^{-\epsilon t}y=\sin t \tag{6}$$

with initial conditions

$$y(0)=y_0;\quad \dot{y}(0)=v_0 \tag{7}$$

where $(\dot{})$ is the derivative with respect to the nondimensional time t. Usually in the study of resonance the initial conditions are equal to zero; see, e.g., [3]. In the case of no decay, $\epsilon=0$, Eq. (6) reduces to the classical case

$$\ddot{y}+2J\dot{y}+y=\sin t \tag{8}$$

for which the solution satisfying the initial conditions is

$$y(t)=e^{-Jt}\left\{\left(y_0+\frac{1}{2J}\right)\cos\omega_Dt+\frac{\left(v_0+\frac{1}{2}+y_0J\right)}{\omega_D}\sin\omega_Dt\right\}$$
$$-\frac{1}{2J}\cos t \tag{9}$$

where $\omega_D=\sqrt{1-J^2}$. The steady-state portion of the solution, that is, the solution as $t\to\infty$, has an amplitude of oscillatory motion of $(1/2J)$ which for lightly damped systems, $J\ll1$, can be large. The steady-state behavior is also independent of the initial conditions. The steady-state solution in Eq. (9) follows directly from Eq. (8) where the damping force term balances the forcing function and the inertia and spring force term cancel one another.

In the case $\epsilon\neq0$ in Eq. (6), we note that as $t\to\infty$, the steady-state solution must satisfy

$$\ddot{y}+2J\dot{y}=\sin t \tag{10}$$

the solution to which is

$$y(t)=C_1-\frac{1}{1+4J^2}[\sin t+2J\cos t],\quad \text{as}\quad t\to\infty \tag{11}$$

This result is surprising because the amplitude of the steady-state oscillatory motion is now approximately unity having decreased from $(1/2J)$ when $\epsilon=0$. Further, the solution oscillates in general about a nonzero value of y given by C_1 that will depend upon the initial conditions and the values of J and ϵ.

We now wish to obtain the exact solution of Eq. (6) using the method of variation of parameters; see e.g., [4], from which the steady-state solution of the form in Eq. (11) can be extracted exhibiting the value of C_1.

3 Exact Solution

The complete solution to Eq. (6) in terms of the two solutions to the homogeneous equation $u_1(t)$, $u_2(t)$ can be written in the form

$$y(t)=u_1(t)F_1(t)+c_1u_1(t)+u_2(t)F_2(t)+c_2u_2(t) \tag{12}$$

where

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 30, 2001; final revision, September 5, 2001. Associate Editor: N. C. Perkins.

$$F_1(t) = -\int \frac{\sin t u_2(t)}{W(t)}\,dt \tag{13}$$

$$F_2(t) = \int \frac{\sin t u_1(t)}{W(t)}\,dt \tag{14}$$

$$W(t) = \begin{vmatrix} u_1 & u_2 \\ \dot{u}_1 & \dot{u}_2 \end{vmatrix}, \tag{15}$$

Wronskian and c_1 and c_2 are the constants of integration. Upon applying the initial conditions, Eq. (7), we find that the solution can be written in the form

$$\begin{aligned} y(t) = {} & u_1(t)F_1(t) - u_1(t)F_1(0) + u_2(t)F_2(t) - u_2(t)F_2(0) \\ & + [u_1(t)\{y_0\dot{u}_2(0) - v_0 u_2(0)\} - u_2(t)\{y_0\dot{u}_1(0) \\ & - v_0 u_1(0)\}]\frac{1}{W(0)}. \end{aligned} \tag{16}$$

The solutions to the homogeneous equation

$$\ddot{y} + 2J\dot{y} + e^{-\epsilon t}y = 0 \tag{17}$$

follow upon making the substitutions ([1])

$$y(t) = e^{-Jt}v(t), \quad s = s_0 e^{-t/s_0}, \quad s_0 = (2/\epsilon) \tag{18}$$

to obtain a Bessel equation

$$s^2 v'' + s v' + v(s^2 - \lambda^2) = 0 \tag{19}$$

where $\lambda = (2J/\epsilon)$ and primes indicate differentiation with respect to s. The solutions to Eq. (19) are Bessel functions $J_\lambda(s)$, $J_{-\lambda}(s)$ with $\lambda \neq n$, an integer. The solutions to Eq. (17) can be written in the form

$$u_1(t) = e^{-Jt}J_\lambda(s_0 e^{-t/s_0}); \quad u_2(t) = e^{-Jt}J_{-\lambda}(s_0 e^{-t/s_0}) \tag{20}$$

and $F_1(t)$ and $F_2(t)$ follow from Eqs. (13–15), (20),

$$F_1(t) = -\frac{\pi s_0}{2\sin\lambda\pi}\int e^{Jt}\sin t J_{-\lambda}(s)\,dt \tag{21}$$

$$F_2(t) = \frac{\pi s_0}{2\sin\lambda\pi}\int e^{Jt}\sin t J_\lambda(s)\,dt. \tag{22}$$

To obtain the solution as $t\to\infty$, from Eq. (16), we need to determine the expansions of $u_1(t)$, $u_2(t)$, $F_1(t)$, and $F_2(t)$ as $t\to\infty$ and the values of $F_1(0)$ and $F_2(0)$. It follows from the expansions of the Bessel functions as $t\to\infty$, $s\to 0$, that

$$u_1(t)\to e^{-2Jt}\left(\frac{s_0}{2}\right)^\lambda \frac{1}{\lambda!}; \quad u_2(t)\to\left(\frac{2}{s_0}\right)^\lambda \Big/ (-\lambda)! \tag{23}$$

In order to find $F_1(t)$ and $F_2(t)$ as $t\to\infty$, we expand the Bessel functions about $s=0$, and integrate to find

$$F_1(t) = -\frac{\pi s_0}{2\sin\lambda\pi}\left(\frac{2}{s_0}\right)^\lambda \frac{1}{(-\lambda)!}\left[\frac{e^{2Jt}(2J\sin t - \cos t)}{1+4J^2}\right] \tag{24}$$

$$F_2(t) = -\frac{\pi s_0}{2\sin\lambda\pi}\left(\frac{s_0}{2}\right)^\lambda \frac{1}{\lambda!}\cos t. \tag{25}$$

The combination $u_1(t)F_1(t) + u_2(t)F_2(t)$ in Eq. (16) as $t\to\infty$ becomes

$$\frac{-(\sin t + 2J\cos t)}{1+4J^2} \tag{26}$$

which agrees with Eq. (11). The remaining terms in Eq. (16) can be evaluated, e.g., $F_2(0)$ can be found from Eq. (22) by expanding the Bessel function $J_\lambda(s)$ in series form, integrating term by term, and setting $t=0$ to find

$$F_2(0) = -\frac{\pi}{\sin\lambda\pi}\left(\frac{s_0}{2}\right)^{\lambda+1}\sum_0^\infty \frac{(-1)^k}{k!(k+\lambda)!}\left(\frac{s_0}{2}\right)^{2k}\frac{1}{1+(2k/s_0)^2}. \tag{27}$$

It follows that the steady-state solution is

$$\begin{aligned} y(t) = {} & \epsilon^{\lambda-1}\Gamma(\lambda)\sum_0^\infty \frac{(-1)^k}{k!(k+\lambda)!}\left(\frac{s_0}{2}\right)^{2k+\lambda}\frac{1}{1+(2k/s_0)^2} \\ & - \epsilon^{\lambda-1}\Gamma(\lambda)\{y_0 J_{\lambda-1}(s_0) - v_0 J_\lambda(s_0)\} - \frac{\sin t + 2J\cos t}{1+4J^2} \end{aligned} \tag{28}$$

The value of y about which the oscillations of the steady-state occur is given by the first two terms in Eq. (28). We see therefore that the steady-state resonance response in the presence of an aging spring is an oscillatory motion of approximately unit amplitude about a constant value that depends on J and ϵ and the values of the initial conditions. This result is one that might not have been anticipated on physical grounds. For example, if the initial conditions are zero with $J=0.1$ and $\epsilon=0.3$, the steady-state value is $=-0.282$. As an additional result, we note that in the case of a zero forcing function with the system excited only by the initial conditions, the steady-state response is a constant occurring after a finite number of oscillations.

4 Concluding Remarks

The interest in the nature of the solution was motivated by the work in [1,2] in which the finite number of zeros in the solution for the case of no damping and no external loading was discussed. A related case is

$$k(\tau) = k_1 + (k_0 - k_1)\exp[-\hat{\epsilon}\tau]$$

in which the spring stiffness at long times approaches a nonzero value k_1. In this case the solution involves Bessel functions of imaginary order. However, the behavior of the steady-state motion is oscillatory about $y=0$ with an amplitude determined by the amount of damping relative to critical damping at the value of k_1.

Acknowledgments

The comments of the two anonymous reviewers are very much appreciated and were insightful for carrying the solution forward from zero initial conditions to the general case treated above.

References

[1] Chen, H., and Wu, T. T., 1970, "An Aging Spring," Stud. Appl. Math., **XLIX**, No. 2, pp. 183–185.

[2] Wan, F. Y. M., 1976, "Education Notes, Applications of Differential Equations," *Applied Mathematics Note*, Vol. 2, Canadian Applied Math Soc., pp. 27–31.

[3] Chopra, A. K., 2001, *Dynamics of Structures*, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ, Section 3.2.2.

[4] Hildebrand, F. B., 1976, *Advanced Calculus for Applications*, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ.

An Exact Solution for Response Spectral Density of a Single-Degree-of-Freedom System Under Both Parametric and Additive White Noise Excitations

M. F. Dimentberg
Mechanical Engineering Department, Worcester Polytechnic Institute, Worcester, MA 01609-2280.
e-mail: diment@wpi.edu

Y. K. Lin
Center for Applied Stochastics Research, Florida Atlantic University, Boca Raton, FL 33431-0991.
e-mail: linyk@fau.edu

Consider the following equation of motion

$$\ddot{x}+2\alpha\dot{x}[1+\eta(t)]+\Omega^2x[1+\xi(t)]=\zeta(t) \tag{1}$$

where $\xi(t)$, $\eta(t)$, $\zeta(t)$ are Gaussian white noises (in the Stratonovich sense) with spectral densities $K_{\xi\xi}$, $K_{\eta\eta}$, $K_{\zeta\zeta}$, respectively. The two parametric white noises, $\eta(t)$ and $\xi(t)$, are correlated, with cross-spectral density $K_{\xi\eta}$. The objective is to obtain the spectral density (PSD) of the response $x(t)$, which exists if $x(t)$ is stationary in the wide sense. This is the case, provided $x(t)$ has constant finite mean and mean square values, and its PSD so obtained is positive everywhere.

Rewrite Eq. (1) in the space state form

$$\dot{x}_1=x_2,\quad \dot{x}_2=-2\alpha x_2-\Omega^2x_1-2\alpha x_2\eta(t)-\Omega^2x_1\xi(t)+\zeta(t) \tag{2}$$

and take the ensemble averages $E[x_i]=m_i$ to obtain

$$\dot{m}_1=m_2,\quad \dot{m}_2=-2\gamma m_2-\Lambda^2m_1,$$
$$\gamma=\alpha(1-\alpha K_{\eta\eta}),\quad \Lambda^2=\Omega^2(1-\alpha K_{\xi\eta}) \tag{3}$$

in which the so-called Wong-Zakai correction ([1]) has been incorporated to evaluate the following expectations of the products of state variables and excitations ([1–3]):

$$E[-x_2\eta(t)]=(K_{\eta\eta}/2)E[(-2\alpha x_2)(d/dx_2)(-2\alpha x_2)]=2\alpha^2K_{\eta\eta}m_2$$

$$E[-x_1\xi(t)]=(K_{\xi\eta}/2)E[(-\Omega^2x_1)(d/dx_1)(-2\alpha x_2)]=\alpha\Omega^2K_{\xi\eta}m_1.$$

The parameters γ and Λ, as defined in Eq. (3), may be regarded as the effective damping factor and effective natural frequency, respectively, provided that γ is positive and Λ is real (negative γ or imaginary Λ, or both, implies system instability in the mean, which is not considered herein).

To obtain the PSD of $x(t)$ at the state of stationarity in the wide sense, Eq. (1) is supplemented with a "measuring filter" ([2]), governed by the following second-order differential equation:

$$\ddot{z}+2\beta\dot{z}+\omega^2z=x(t). \tag{4}$$

Expressing the mean square value of $z(t)$ as the integral of the product of the PSD $\Phi_{xx}(\omega)$ of $x(t)$ and the squared magnitude of the transfer function of filter (4), one can show that the following relation holds:

$$E[z^2]=\lim_{\beta\to0}[\pi\Phi_{xx}(\omega)/2\beta\omega^2]. \tag{5}$$

Thus, the response PSD of system (1) can be obtained when the mean square response of the measuring filter (4) is found.

To solve this problem in the framework of the stochastic differential equation (SDE) calculus, Eq. (4) with z and $\dot{z}$ renamed as x_3 and x_4, respectively, is first rewritten in the state space form as follows:

$$\dot{x}_3=x_4,\quad \dot{x}_4=-2\beta x_4-\omega^2x_3+x_1. \tag{6}$$

Combining Eqs. (2) and (6), and denoting $u_{ij}=x_ix_j$; $\dot{u}_{ij}=\dot{x}_ix_j+x_i\dot{x}_j$ $(i,j=1-4)$ one obtains the following SDE set for u_{ij}:

$$\dot{u}_{11}=2u_{12},$$
$$\dot{u}_{22}=-4\alpha u_{22}[1+\eta(t)]-2\Omega^2u_{12}[1+\xi(t)]+\sqrt{u_{22}}\zeta(t),$$
$$\dot{u}_{12}=u_{22}-\Omega^2u_{11}[1+\xi(t)]-2\alpha u_{12}[1+\eta(t)]+\sqrt{u_{11}}\zeta(t),$$
$$\dot{u}_{13}=u_{23}+u_{14},\quad \dot{u}_{14}=u_{24}-\beta u_{14}-\omega^2u_{13}+u_{11},$$
$$\dot{u}_{23}=u_{24}-2\alpha u_{23}[1+\eta(t)]-\Omega^2u_{13}[1+\xi(t)]+\sqrt{u_{33}}\zeta(t), \tag{7}$$
$$\dot{u}_{24}=-2(\alpha+\beta)u_{24}-2\alpha u_{24}\eta(t)-\omega^2u_{23}-\Omega^2u_{14}[1+\xi(t)]+u_{12}+\sqrt{u_{44}}\zeta(t),$$
$$\dot{u}_{33}=2u_{34},\quad \dot{u}_{44}=-4\beta u_{44}-2\omega^2u_{34}+2u_{14},$$
$$\dot{u}_{34}=u_{44}-2\beta u_{34}-\omega^2u_{33}+u_{13}.$$

The expectation operator is now applied to the ten SDEs (7) to obtain a set of deterministic differential equations for the second-order response moments $D_{ij}=E[u_{ij}]=E[x_ix_j]$ $(i,j=1-4)$. The general formula for Wong-Zakai corrections [1–3] is used once again to evaluate the expectations of products of the state variables and excitations. At the state of stationarity, $E[\dot{u}_j]=0$, these equations reduce to algebraic equations. Furthermore, it is evident from the first and the eighth of these equations, that $D_{12}=0$ and $D_{34}=0$. The remaining eight equations are given by

$$-4\alpha(1-2\alpha K_{\eta\eta})D_{22}+\Omega^4K_{\xi\xi}D_{11}+K_{\zeta\zeta}=0,$$
$$D_{22}=\Lambda^2D_{11},\quad D_{23}+D_{14}=0,$$
$$D_{24}-2\beta D_{14}-\omega^2D_{13}+D_{11}=0,$$
$$D_{24}-2\gamma D_{23}-\Lambda^2D_{13}=0, \tag{8}$$
$$2(\gamma+\beta)D_{24}=-\omega^2D_{23}-\Lambda^2D_{14},\quad 4\beta D_{44}=2D_{14},$$
$$D_{44}-\omega^2D_{33}+D_{13}=0$$

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, June 20, 2001; final revision, Sept. 25, 2001. Associate Editor: N. C. Perkins.

which can be split into several subsets, and solved analytically. In particular, the mean square displacement D_{11} can be obtained from the first and second equations in (8) as

$$D_{11}=E[x^2]=K_{\zeta\zeta}/[4\alpha\Omega^2(1-2\alpha K_{\eta\eta})(1-\alpha K_{\xi\eta})-\Omega^4 K_{\xi\xi}]. \tag{9}$$

The denominator on the right-hand side of (9) is assumed to be positive, which is, in fact, the condition for mean square stability of system (1), and is a necessary and sufficient condition for the existence of the stationary state response in the wide sense.

From the seventh and the eighth equations in (8), we obtain

$$2\beta\omega^2 E[z^2]=D_{14}+2\beta D_{13}. \tag{10}$$

Note that $E[z^2]=D_{33}$. Using relation (5), and imposing then the condition $\beta\to 0$ in the remaining equations yield

$$\pi\Phi_{xx}(\omega)=2\gamma\Lambda^2 E[x^2]/[(\omega^2-\Lambda^2)^2+4\gamma^2\omega^2]. \tag{11}$$

Thus, the response PSD is seen to be of the same basic shape as that without the parametric excitations but with reduced bandwidth and a shifted peak position, according to Eq. (3). The former of these effects is due to random variations of damping, the latter due to the cross correlation of these variations, which affects the stiffness of the system. The validity of the obtained PSD is restricted under the imposed inequalities for system stability in the mean and in the mean square. The condition for mean-square stability is, of course, stricter than that for the mean. It indicates, in particular, that the apparent bandwidth of the system never goes down to zero (contrary to the deterministic systems). For example, in a simple case $K_{\xi\xi}=0$ its minimal value is just $\gamma_{\min}=\alpha/2$ since the system becomes unstable in mean square when the value of $\alpha K_{\eta\eta}$ exceeds 1/2. In such a case, the response PSD does not exist, as can be seen in Eq. (9). The bound $K_{\xi\eta}^2/K_{\xi\xi}K_{\eta\eta}\leqslant 1$ must hold also for realistic modeling of the random excitations. It is of interest to note that Eq. (11) for the response PSD coincides with Eq. (70) in paper [4], which was derived using an approximate averaging procedure, under the conditions of small α as well as small excitation spectral densities. In contrast, the derivation presented herein is exact, without restrictions beyond those for stochastic stability of the system. Of course, this exact solution is valid only if the excitations are modeled as white noises, so that the SDE calculus is applicable.

References

[1] Wong, E., and Zakai, M., 1965, "On the Relation Between Ordinary and Stochastic Equations," Int. J. Eng. Sci., **3**, pp. 213–229.

[2] Dimentberg, M. F., 1988, *Statistical Dynamics of Nonlinear and Time-Varying Systems*, Research Studies Press, Taunton, UK.

[3] Lin, Y. K., and Cai, G. Q., 1995, *Probabilistic Structural Dynamics*, McGraw-Hill, New York.

[4] Krenk, S., Lin, Y. K., and Rudinger, F., 2002, "Effective System Properties and Spectral Density in Random Vibration With Parametric Excitation," ASME J. Appl. Mech., submitted for publication.