ASME
SETTING THE STANDARD

Journal of Applied Mechanics

Published Bimonthly by ASME

VOLUME 73 • NUMBER 5 • SEPTEMBER 2006

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

Special Issue on Current Trends in Mechanics

FOREWORD

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 73, Number 5 SEPTEMBER 2006

TECHNICAL BRIEF

ADDITIONAL TECHNICAL PAPERS

DISCUSSION

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

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

Journal of Applied Mechanics

Foreword

This issue represents a compilation of papers contributed by former students, colleagues, and collaborators of Professor Wolfgang Knauss, Theodore von Karman Professor of Aeronatuics and Applied Mechanics at the California Institute of Technology on the occasion of his 70th birthday. Most of these papers were presented at the symposium "Current Trends in Mechanics," held on November 15–16, 2004 at the Graduate Aeronautical Laboratories of the California Institute of Technology (GALCIT), to commemorate this milestone.

Wolfgang Gustav Knauss was born in Mandel/Bad Kreuznach, Germany, on December 12, 1933. After graduating with honors from the Helmholtz Realgymnasium, Heidelberg, Germany in 1954, Wolfgang moved to Pasadena and made it his home over the last five decades, first as a student at the California Institute of Technology, getting his Ph.D. in Aeronautics in 1963, and since then as a member of the faculty at Caltech. Currently, he is an Emeritus Professor of Aeronautics and Applied Mechanics. In addition to his academic accomplishments, he has used his expertise in addressing practical problems, as a consultant to a number of aerospace companies such as Lockheed, Rocketdyne, Aerojet-General, Hercules, General Dynamics, and GE Space Division, as well as many producers and consumers of polymer products such as DuPont and Firestone.

Wolfgang has served on numerous advisory boards, most notably, the U.S. Highway Research Board's Committee on Strength and Deformation; the Executive Committee of the Adhesion Society; Chairman of the National Panel on Foreign Technology Assessment in Fracture Mechanics; Chairman of the U.S. National Committee on Theoretical and Applied Mechanics; and Chairman of the Subcommittee on Failure, Structural Integrity Committee of the Interagency Chemical Propulsion Group for the Joint Army, Navy, and Air Force Technical Advisory Committee. His service to the applied mechanics community includes initiation of a journal devoted to *Mechanics of Time-Dependent Materials*.

In recognition of his significant contributions to the literature and the community, Wolfgang has received many honors and awards for his work over the years, too numerous to list; here are some of the highlights. He was awarded a lectureship by the USSR Academy of Sciences in 1977, which enabled him to travel to different Soviet era academic and research institutions and get acquainted with Soviet research in time-dependent failure of polymeric materials. He was awarded the Alexander von Humboldt Senior U.S. Scientist fellowship in 1986–1987, during which tenure, he spent some time at the University of Kassel and the University of Karlsruhe in Germany. He received the Murray Medal from the Society for Experimental Mechanics (1995), the Kapitsa Medal of the Russian Academy of Sciences (1997), and the Koiter Medal from the American Society of Mechanical Engineers (2001). He was elected a foreign member of the Russian Academy of Natural Sciences in 1997. The impact of his research contributions can be gauged by his inclusion in the list of Highly Cited Researchers by ISI. His election to the National Academy of Engineering in 1998 for contributions to time-dependent fracture of polymers, at interfaces, and under dynamic loading was a fitting recognition of an illustrious career marked by high-quality, original work, and for significant service to the applied mechanics community.

Wolfgang is well known for his work in viscoelasticity, dynamic fracture, composites, and interface fracture. These interests have been bolstered in recent years by his emphasis on making measurements at ever smaller scales; the first seven papers in this issue fall in this category. Ioannis Chasiotis—Wolfgang's last Ph.D. student—describes work on fracture of free-standing MEMS films, and Wolfgang's colleagues Ares Rosakis and Ravichandran present work on the characterization of the geometrical imperfections and mechanical characteristics of thin film materials. Hongbing Lu describes work on modeling carbon nanotube/polyelectrolyte multilayer composites; this is followed by Tony Waas' innovative application of micro-Raman spectroscopy to the measurement of biaxial stresses in silicon.

Several papers were presented on the behavior of polymers, polymer composites, and nanocomposites. The paper by Igor Emri looks at multi-scale interactions in polymer structure formation. On the nanoscale, Cate Brinson provides an evaluation of interphases in nanocomposites and Ken Liechti examines self-assembled monolayers through experiments with the Interfacial Force Microscope.

The area of fracture mechanics is well represented in this issue by many contributions. Jimmy Liu, a long-time collaborator, and colleague Ravichandran describe the influence of confining pressure on crack growth in a highly filled elastomer. John Lambros and Phillippe Guebelle use cohesive zone modeling to examine quasi-static fracture in functionally graded polymers. Jay Walton provides an analysis of dynamic crack propagation in linearly viscoelastic materials. An application of fracture mechanics to the problem of truck tire durability is explored by Allan Zhong.

Response of different classes of materials are considered in four papers in this collection. Stelios Kyriakides looks into the crushing response of low density foams, Brian Moran develops a constitutive model with fiber matrix interactions for the human annulus fibrosus, and Dick Christensen presents an evaluation of different failure theories for fiber-reinforced composites. Sia Nemat Nasser describes an experimental evaluation of the dynamic collapse behavior of shape memory alloys.

Finally, we have an acknowledgement of GALCIT's historical involvement in structures with papers by former students Herzl Chai and Ravi-Chandar. Herzl's paper is about the crash worthiness of laterally constrained thin plate members under axial compression, while Ravi takes a look at the nonlinear dynamics of strings and rods, adding a biomechanical twist to the problem.

In addition to the papers presented in this volume, there were interesting oral presentations at the symposium from Kyung-Suk Kim, Albert Kobayashi, Gordon Williams, Phillipe Guebelle, Wei Tong, and Sandeep Sane.

K. Ravi-Chandar
Kenneth Liechti
Stelios Kyriakides
Guest Editors

Fracture Toughness and Subcritical Crack Growth in Polycrystalline Silicon

I. Chasiotis
Aerospace Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801
e-mail: chasioti@uiuc.edu

S. W. Cho
Materials Science and Engineering,
University of Virginia,
Charlottesville, VA 22904

K. Jonnalagadda
Aerospace Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801

The fracture behavior of polycrystalline silicon in the presence of atomically sharp cracks is important in the determination of the mechanical reliability of microelectromechanical system (MEMS) components. The mode-I critical stress intensity factor and crack tip displacements in the vicinity of atomically sharp edge cracks in polycrystalline silicon MEMS scale specimens were measured via an in situ atomic force microscopy/digital image correlation method. The effective (macroscopic) mode-I critical stress intensity factor for specimens from different fabrication runs was 1.00 ± 0.1 MPa$\sqrt{\text{m}}$*, where* 0.1 MPa$\sqrt{\text{m}}$ *is the standard deviation that was attributed to local cleavage anisotropy and grain boundary effects. The experimental near crack tip displacements were in good agreement with the linearly elastic fracture mechanics solution, which supports K dominance in polysilicon at the scale of a few microns. The mechanical characterization method implemented in this work allowed for direct experimental evidence of incremental (subcritical) crack growth in polycrystalline silicon that occurred with crack increments of* $1-2$ μm*. The variation in experimental effective critical stress intensity factors and the incremental crack growth in brittle polysilicon were attributed to local cleavage anisotropy in individual silicon grains where the crack tip resided and whose fracture characteristics controlled the overall fracture process resulting in different local and macroscopic stress intensity factors.* [DOI: 10.1115/1.2172268]

1 Introduction

The fracture of polycrystalline silicon (polysilicon) films in the presence of atomically sharp cracks is one of the least explored aspects of the mechanics of this material. Generally speaking, the extent that the material microstructure (grain orientation) influences the apparent fracture toughness (based on boundary measurements) depends on the local fracture toughness and the yield strength. Large values of fracture toughness and small values of yield strength would result in a large process zone that would decrease the importance of material anisotropy and inhomogeneity near the crack tip. In contrast, silicon has small fracture toughness and very high yield strength. Thus, it is expected that local material details will be reflected in the effective critical stress intensity factor, $K_{Ic,\text{PolySi}}$ [1,2]. The first efforts to measure an approximate $K_{Ic,\text{PolySi}}$ for polysilicon fabricated for microelectromechanical systems (MEMS) involved notched specimens with finite notch tip radii, on the order of one micron [3–6], resulting in $K_{Ic,\text{PolySi}}$ values in the range 1.4 ± 0.15 MPa$\sqrt{\text{m}}$ to 4.5 MPa$\sqrt{\text{m}}$. The degree of convergence of these finite notch radii measurements to the actual $K_{Ic,\text{PolySi}}$ is difficult to determine from theoretical/analytical considerations alone. A recent fracture mechanics work by Kahn et al. [7,8] showed that some of the previous estimates for $K_{Ic,\text{PolySi}}$ were considerably high. Their studies were conducted with the aid of on-chip electrostatically actuated microdevices and the effective fracture toughness was derived by a finite element (FE) analysis that required precise fabrication of interdigitated comb finger arrays.

The work in [8] benefited from an advancement in experimental fracture mechanics of MEMS that allowed for fabrication of fracture specimens with mathematically sharp cracks [9]. In the present work we generated atomically sharp cracks and measured the fracture toughness of polycrystalline silicon using uniaxial tension testing with externally loaded MEMS-scale specimens. This approach was combined with the atomic force microscopy (AFM)/digital image correlation (DIC) method [10–12] to obtain local deformation fields in the vicinity of mode-I loaded edge cracks. The following section describes the specimen geometry and fabrication of edge precracked specimens, the experimental procedures, and the methods used to analyze the experimental data. The third section presents measurements of the fracture toughness and a comparison of experimental crack tip displacements with the linear elastic fracture mechanics solution. Measurements of subcritical (incremental) crack growth under load control and the associated crack tip deformation fields are described and discussed. In addition, electron microscopy micrographs of the specimen fracture cross sections are presented. The last section summarizes this work and provides concluding remarks.

2 Experiments

Microscale fracture specimens were designed and fabricated following previous studies by the same group on the determination of the elastic constants and failure properties of MEMS materials [11,12]. The specimen design was conducted to benefit from the AFM/DIC local deformation measurement method so that local displacement fields could be obtained directly from the test samples.

2.1 Preparation of Fracture Specimens. Polycrystalline silicon specimens were manufactured at Cronos (Research Triangle Park, NC) (former Microelectronics Center of North Carolina (MCNC)) via multi-user MEMS processes (MUMPs)39 and MUMPs41. Each silicon die carried multiple specimens with gauge sections varying between $50-400$ μm in width and $500-1000$ μm in length. Two specimen thicknesses were employed: 1.5 μm and 2.0 μm, which represent different structural layers in the MUMPs fabrication process. All specimens were

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received by March 30, 2005; final manuscript received December 10, 2005. Review conducted by K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

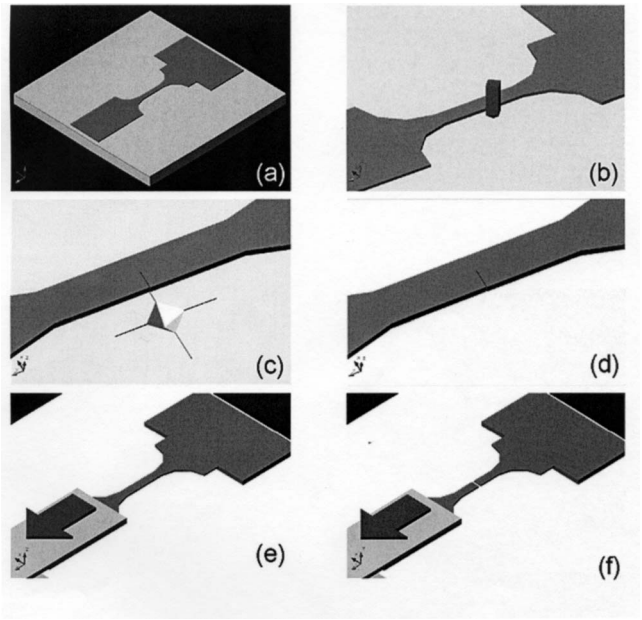


Fig. 1 Preparation and testing of MEMS-scale fracture specimens. (*a*) Specimen before indentation, (*b*) indenter tip brought in contact with the specimen substrate near the polysilicon specimen, (*c*) specimen with edge precrack after indentation, (*d*) freestanding specimen with edge precrack after substrate removal (release), (*e*) specimen loaded with a glass grip attached to the specimen paddle, and (*f*) fracture of specimen.

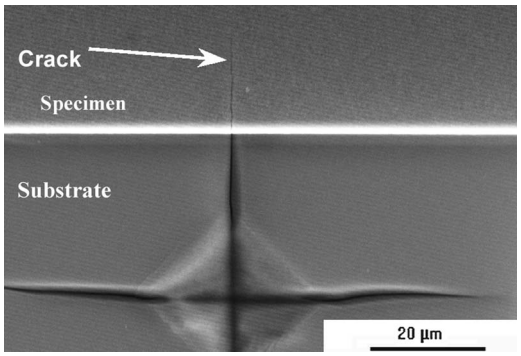


Fig. 2 Indentation created via a Vickers indenter on SiO_2 (substrate). The crack propagated (vertically) into the polycrystalline silicon specimen.

fabricated on a 2-μm thick SiO_2 substrate that served as a sacrificial layer on the Si wafer. The polysilicon test structures were flat after removing the underlying SiO_2 sacrificial layer, which attests the low average stress gradients in the film.

From the uniform tension MEMS-size specimens, fracture mechanics specimens with mathematically sharp cracks were fabricated with the aid of a Vickers microhardness indenter [13,14]. As opposed to existing methods for the determination of fracture toughness via indentation (for instance see [15]) the application of indentation in the present method was auxiliary for the creation of edge precracks. Figure 1 shows the procedure for generation of precracks in the microtension specimen in Fig. 1(*a*). First, indentations were created on the SiO_2 sacrificial layer by applying an indentation force in the range of 3–5 N and at a distance of 20–30 μm from the edge of the polysilicon specimen as seen in Fig. 1(*b*). One of the cracks emanating from the four corners of the indent propagated into the tensile specimen (that was chemically bonded to SiO_2) and generated a mathematically sharp edge crack (Figs. 1(*c*) and 2). The crack driving force generated by the indenter decreased away from the indentation mark resulting in crack arrest in the polysilicon specimen. The length of the edge precracks varied randomly between 2.5–26 μm for specimens with gauge sections 50–400 μm wide.

The process for generating cracks and removing the SiO_2 sacrificial layer under the specimens proved robust and repeatable and it could be directly applied to other thin film materials. The angle of crack propagation and the length of the crack were controlled by the indent orientation and its location with respect to the edge of the specimen. The length of the crack in the specimen was a function of the indentation load, the specimen thickness, and the distance from the edge of the specimen.

All precrack lengths were measured via AFM before removal of the sacrificial SiO_2 layer. Compared to optical and scanning electron microscopy (SEM), the AFM crack length measurements, although elaborate, were better as far as repeatability and accuracy in the calculation of K_{Ic} are of concern. The uncertainty in the measurement of crack lengths by an AFM was about 1–2 image pixels, i.e., 10 nm.

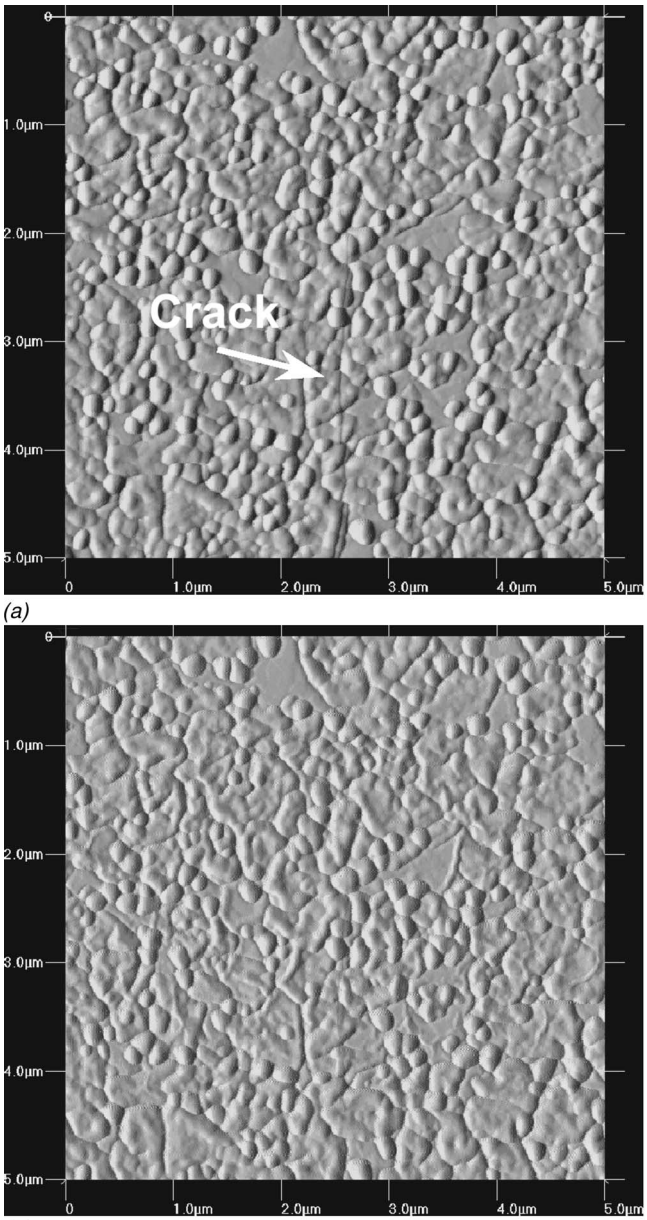
The specimens were released from the substrate by wet etching the SiO_2 sacrificial layer in 49 vol % HF to create *freestanding* fracture specimens with edge precracks (Fig. 1(*d*)). CO_2 supercritical drying was applied to prevent surface adhesion (stiction) between the specimens and the substrate. The crack opening displacement after HF release was zero and the precracks were nearly invisible even using high resolution AFM images [13]. In Fig. 3(*a*) the precrack was clearly visible in a 5×5 μm^2 AFM image, but it was barely visible in the AFM image (Fig. 3(*b*)) taken after removal of the sacrificial SiO_2 layer. This underlines the absence of local damage at the crack tip and provides confidence in using linear elasticity solutions to analyze the experimental data from the sharp cracks created for this work.

The AFM crack images (Figs. 3(*a*) and 3(*b*)) provided information about the fracture manner in polysilicon. The fracture induced by the indenter was predominantly transgranular and the cracks were arrested inside individual silicon grains or at triple junction points. The crack propagation path was slightly meandering following an overall straight line. The precracks were normal to the specimen edge with less than 0.7 deg deviation.

In terms of material quality, the surface roughness before and after HF release was found to be consistent without noticeable material degradation. Figures 3(*a*) and 3(*b*) show the topography of specimens before and after HF release where no surface pitting was observed. For such small specimens, the surface quality is very important: HF release can cause degradation of MUMPs polysilicon, especially when gold metallization is present [16,17]. Thus, for specimens from MUMPs41 no metallization was used and the exposure to HF was timed so that the data reported here are representative for polysilicon fabricated for MEMS.

The freestanding fracture specimens were tested under mode-I loading using a microtensile tester developed to carry out in situ AFM measurements of local deformation fields on MEMS-scale specimens [12]. The specimens were aligned on the test apparatus using a dual-axis microscope for both lateral and vertical tilt adjustment to ensure that the applied load was perpendicular to the crack surface and no out-of-plane bending moments were present. The large paddle of the specimen was mounted on a glass grip using an ultraviolet light (UV) curable adhesive (Fig. 1(*e*)) [18]. The specimens were loaded using a piezoelectric actuator (PZT) with 1 nm displacement resolution and ~0.5% linearity. The applied force at failure (Fig. 1(*f*)) was measured using a miniature load cell with better than 0.1 mN resolution. More details about this test system can be found in [12].

2.2 Calculation of Fracture Toughness. The edge precracked specimen geometry is easy to analyze and calculate the effective fracture toughness using readily available (LEFM) solu-



(a)

(b)

Fig. 3 (*a*) **Edge precrack in a tension specimen imaged by AFM before and** (*b*) **after removal of the sacrificial SiO_2 layer. All precracks were clearly visible in 5–10 μm AFM images recorded before HF release but were invisible after SiO_2 removal.**

tions for mode-I loading in fixed grip configuration [19]. The mode-I critical stress intensity factor K_{Ic} for the present configuration was determined from

$$K_{Ic} = Y\sigma_\infty\sqrt{\pi a} \tag{1}$$

where σ_∞ is the far field stress, a is the crack length, and Y is the shape factor for an edge crack in a finite width specimen given by

$$Y = 1.122 - 0.231\left(\frac{a}{w}\right) + 10.55\left(\frac{a}{w}\right)^2 - 21.71\left(\frac{a}{w}\right)^3 + 30.382\left(\frac{a}{w}\right)^4 + \text{HOT} \tag{2}$$

where w is the specimen width and HOT are higher order terms. The determination of Y is accurate within 0.5% with no dependence on specimen length l, when $l/2w > 0.6$ [19].

Using Eq. (1) the relative uncertainty in the determination of $K_{Ic,\text{PolySi}}$ was calculated. It was mostly proportional to the relative uncertainty in the determination of the stress at failure since the crack length had a smaller contribution to the relative uncertainty and it was measured with better than 0.5% accuracy. The minimum measurable force at failure was limited by the resolution of the load cell. The relative error in $K_{Ic,\text{PolySi}}$ due to the uncertainty in the determination of fracture stress was in the range of 0.25%–0.7% for specimens that failed at high and low far field force, respectively.

The deformation fields in the vicinity of cracks were obtained via the in situ AFM/DIC method with ~2 nm spatial resolution [11,12]. In this method, the undeformed and deformed surface topographies near the crack tip were obtained via an AFM at different levels of stress intensity. Then, the displacement field was resolved by DIC analysis. The linear elasticity displacement solution in the vicinity of an edge crack was compared with the experimentally determined displacements at different stress intensity factors. The isotropic elastic constants, i.e., Young's modulus and Poisson's ratio of polysilicon used in the linear elasticity solution were measured via the AFM/DIC method from 5 $\times 15$ μm^2 specimen domains [13,14].

Table 1 Fracture toughness of polysilicon for different MUMPs runs and layers

Material	MUMPs layer	Thickness (μm)	K_{Ic} (MPa$\sqrt{}$m)
Polysilicon MUMPs39	Poly1	1.5	1.06±0.01
Polysilicon MUMPs39	Poly2	2.0	1.05±0.20
Polysilicon MUMPs41	Poly2	2.0	1.00±0.10

3 Results and Discussion

Table 1 presents the average $K_{Ic,\text{PolySi}}$ for the two specimen thicknesses and the two fabrication runs calculated using Eqs. (1) and (2). The average $K_{Ic,\text{PolySi}}$ from all samples was 1.00±0.1 MPa$\sqrt{}$m which is in agreement with a previous report for polysilicon fabricated by another source [8]. The large standard deviation in $K_{Ic,\text{PolySi}}$ for 2.0-μm thick specimens from MUMPs39 is due to the small number of specimens (only 3) that were available for testing. On the other hand, 25 specimens from the MUMPs41 run were tested and while the minimum and maximum values were in the same range as those for the 2.0-μm thick specimens from MUMPs39, the standard deviation was half that. The values of fracture toughness for the 25 MUMPs41 specimens varied in the range of 0.84–1.24 MPa$\sqrt{}$m with multiple measurements recorded at both extremes (i.e., no outliers).

This large range of fracture toughness values is not due to experimental uncertainties. As mentioned before, the accuracy in determining $K_{Ic,\text{PolySi}}$ was better than 0.01 MPa$\sqrt{}$m. Similarly, fabrication dissimilarities are ruled out because, within the limits of the experimental data scatter, the fracture toughness between different fabrication runs and different structural layers of the MUMPs process (for the limited variations in thickness permitted by MUMPs) was approximately the same. Furthermore, a specimen size effect is also ruled out since an analysis of our data did not point to a dependence of fracture toughness on certain combinations of specimen lateral dimensions, thickness, and crack length.

The measured $K_{Ic,\text{PolySi}}$ was compared with the fracture toughness values of different single crystal silicon planes using the most widely referenced experimental data. Cleavage anisotropy is rather limited in Si: The low energy cleavage planes are {111} with K_{Ic}=0.83 MPa$\sqrt{}$m, followed by {110} with K_{Ic} =0.91 MPa$\sqrt{}$m, and {100} with K_{Ic}=0.95 MPa$\sqrt{}$m [20–22]. The average $K_{Ic,\text{PolySi}}$ measured here was 20% higher than the minimum K_{Ic} for single crystal silicon. On the other hand, the lowest values of $K_{Ic,\text{PolySi}}$ measured in this work were very close

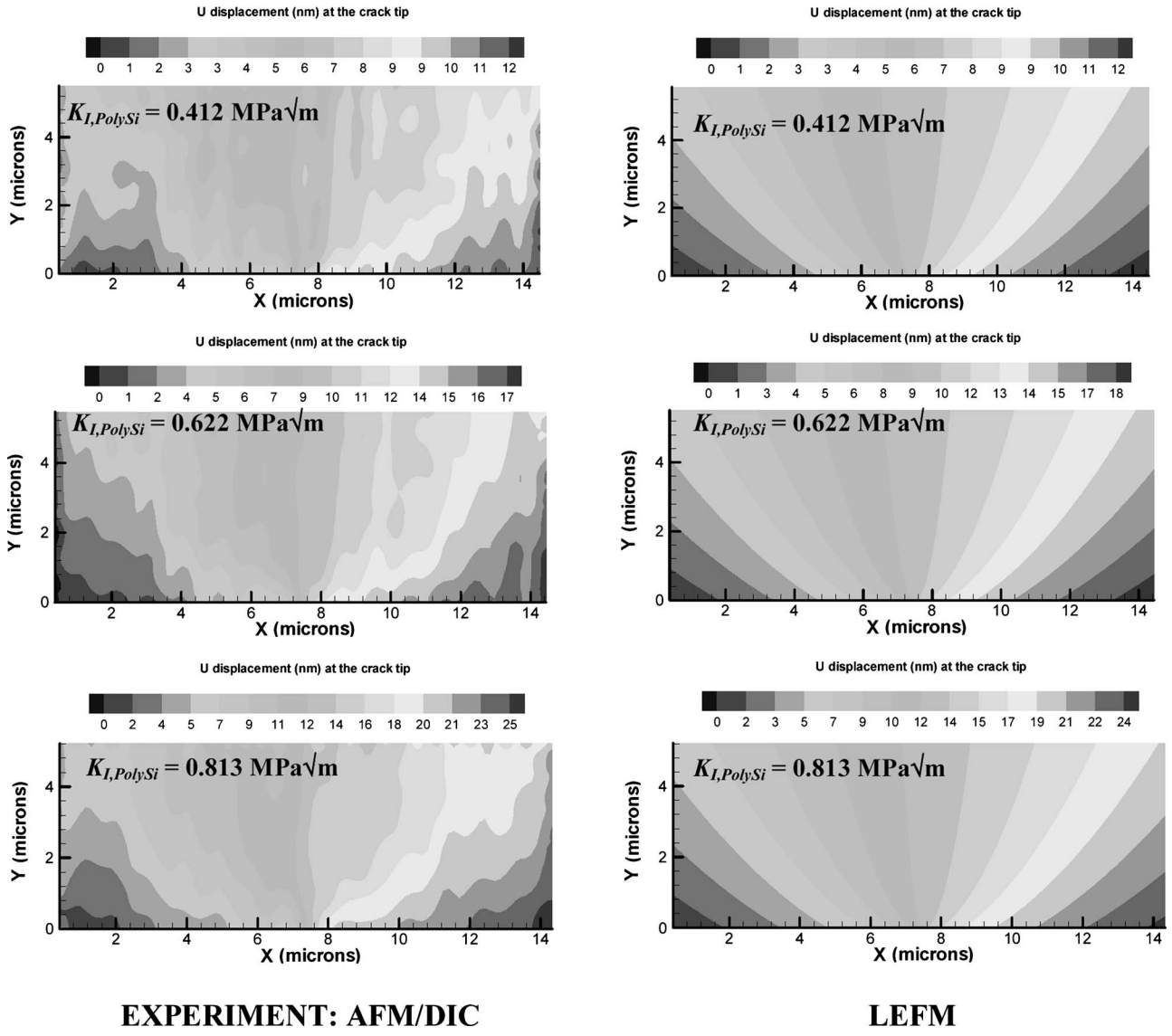


Fig. 4 U-displacement fields in the vicinity of a crack tip for increasing $K_{I,\text{polySi}}$ as obtained by AFM/DIC and LEFM, respectively. The crack tip is located at the root of the contours. The left-to-right order of gray levels in the contour plots is the same as in the contour legends.

(0.84 MPa$\sqrt{\text{m}}$) to the fracture toughness of the favorable cleavage plane (111) of monolithic silicon (0.82–0.83 MPa$\sqrt{\text{m}}$.) The authors in [20] also reported a value for the fracture toughness of bulk polysilicon. Their value for large grained polycrystalline silicon was 0.75±0.06 MPa$\sqrt{\text{m}}$, which is outside the bounds of the results reported in the present work. However, the material tested in [20] contained impurities and large, unaccounted in the derivation of K_{Ic}, residual tensile stresses.

The large distribution of $K_{Ic,\text{PolySi}}$ values reflects the difference between the local and apparent fracture toughness that is attributed to polycrystallinity and local cleavage anisotropy. This has been computationally demonstrated both in the context of macroscopic [1,23] and microscopic [2] brittle material structures. Abdel-Tawab and Rodin [1] defined the ratio of grain diameter d to the radius of process zone r_p as the grain brittleness parameter (GB) and used it as an indication of the influence of polycrystallinity on the effective fracture toughness:

$$\text{GB} = \frac{d}{r_p} \tag{3}$$

The magnitude of GB is a measure of the contribution of the material details of the individual grain where the crack tip resides to the effective fracture toughness. Given $K_{Ic,\text{Si}}$=0.9 MPa$\sqrt{\text{m}}$ [20], $\sigma_{y,\text{Si}} \approx 7.5$ GPa [24,25], and d_{PolySi}=0.3 μm and assuming small scale yielding and plane strain, the extent of the plastic zone can be estimated by $r_p = 1/3\pi(K_{Ic}/\sigma_y)^2$ [27]. Thus, for silicon at room temperature $r_p \approx 2$ nm (5 nm for plane stress), which is considerably smaller than a polysilicon grain. This calculation agrees with transmission electron microscopy (TEM) measurements of strains in the vicinity of cracks in single crystal silicon [26]. In the latter work the measured strain fields agreed very well with the LEFM solution at distances as small as 10 nm away from the crack tip.

Using Eq. (3) we compute GB≈60–150 (considering plane stress and plane strain as extremes in calculating the process zone [27]) which implies that the details of a single grain will determine to a large extent the macroscopic fracture toughness. According to [1], GB≈60–150 places polysilicon in the category of materials for which local crystal anisotropy dictates the effective fracture behavior and the size effect associated with polycrystallinity is insignificant because the grain is very large compared to r_p.

For strong elastic crystal anisotropy, on the other hand, the difference between the local and the effective fracture toughness

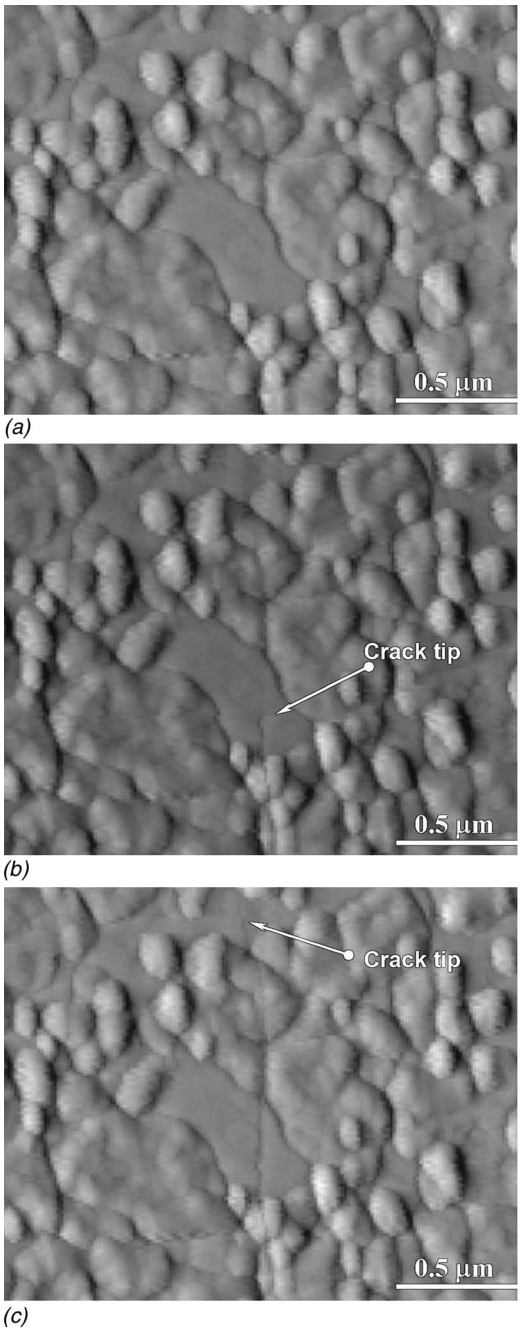


Fig. 5 Subcritical crack growth in a polysilicon specimen. Note the steps at the crack tip in (*b*) and (*c*).

can be controlled by a significant number of grains ranging in the hundreds that provide crack tip "shielding" [1]. In that event the fracture toughness should be described in statistical terms to account for the distribution of grain orientations near the crack tip.

Since the process zone is limited to within one grain,[1] there is K dominance within the grain where the crack tip resides. Polycrystalline anisotropy prohibits the use of LEFM with isotropic material properties inside that grain. Furthermore, in the vicinity of the grain where the crack tip is located, an ensemble of grains is required to establish isotropic effective behavior and permit the use of LEFM. The authors in [2] showed that a set of 1000 columnar polysilicon grains constitutes a representative material element and thus, an equivalent number of grains affect the local elastic deformations around the crack tip. Moreover, they estimated that the distribution of fracture toughness values for polycrystalline silicon should be rather narrow and the local fracture toughness should be on average only ~10% smaller than the effective fracture toughness; a result that agrees with our experimental results.

From this discussion it becomes obvious that the application of LEFM to polysilicon depends on the degree of elastic anisotropy and grain size; for small grain sizes (like those involved in this work) a small number of grains outside the one where the crack tip resides can be used to provide a statistically homogeneous material domain and establish K dominance near the crack tip. This is supported experimentally by comparison of the LEFM solution with the local deformation measurements described in the next section.

It should be also mentioned that variation in the local state of stress in individual grains may also contribute to a distribution of fracture toughness values. Although the specimens tested in this work were freestanding (i.e., the average residual stress was relaxed) and the stress gradient was negligible (the test structures remained flat after HF release), local residual stresses within each grain are still possible but must be considerable to support the large range of measured values for $K_{Ic,\mathrm{PolySi}}$. A thorough study to assess the importance of local residual stresses would involve orientation imaging microscopy assisted by subsequent object oriented finite element analysis to determine the local residual stresses near the crack tip [28].

3.1 Comparison With Analytical Solutions. For some fracture specimens AFM images were acquired at various load levels before fracture. The AFM/DIC measurements were challenging as the entire displacement field near the crack tip was only 25 nm, i.e., less than two data pixels of deformation across the entire AFM field of view that was comprised of 1024 pixels. The subpixel resolution of DIC proved sufficient to resolve the small nonuniform displacements seen in Fig. 4. The figure shows the experimental displacement fields in the vicinity of a crack and the LEFM solution for the same apparent stress intensity factors. Despite the scatter in the experimental data, mostly for small $K_{I,\mathrm{PolySi}}$, the experimental displacement fields were found to be in good agreement with the LEFM solution. The distortion in the experimental contours was caused by the very small local displacements that were resolved from the AFM images. DIC was applied to a square grid of data points, which, for nonuniform subpixel displacements, could induce local distortions that depend on the size of the correlation square. These distortions were reduced significantly in the contour plot with the largest deformations ($K_{I,\mathrm{PolySi}}=0.813$ MPa$\sqrt{\mathrm{m}}$). For local, nonuniform, displacements of larger amplitude this distortion is minimal as was demonstrated for notches with finite stress concentration factors [12,29].

This good comparison of experimental data with the LEFM solution using the isotropic properties of polysilicon further supports the previous argument that a small number of grains is required to establish the isotropic material behavior, which, in turn, permits the assumption of K dominance near the crack tip. This is further corroborated by AFM/DIC displacement field data [11,14] where polysilicon was shown to behave as isotropic at the scale of a few microns. In order to accurately describe the local deformation field in the vicinity of the crack tip (especially for large grain sizes or significant elastic anisotropy) the anisotropic elasticity constants must be used for each grain and the orientation of the grains surrounding the crack tip must be determined a priori.

3.2 Effect of Crystal Anisotropy on Subcritical Crack Growth. Since the fracture process in polysilicon is mainly controlled by the grain where the crack tip resides, fracture initiation

[1] A crack tip very close to a grain boundary or at a triple junction point would complicate this argument; these precracks were responsible for the upper bound of $K_{Ic,\mathrm{PolySi}}$ values (<1.24 MPa$\sqrt{\mathrm{m}}$) reported here.

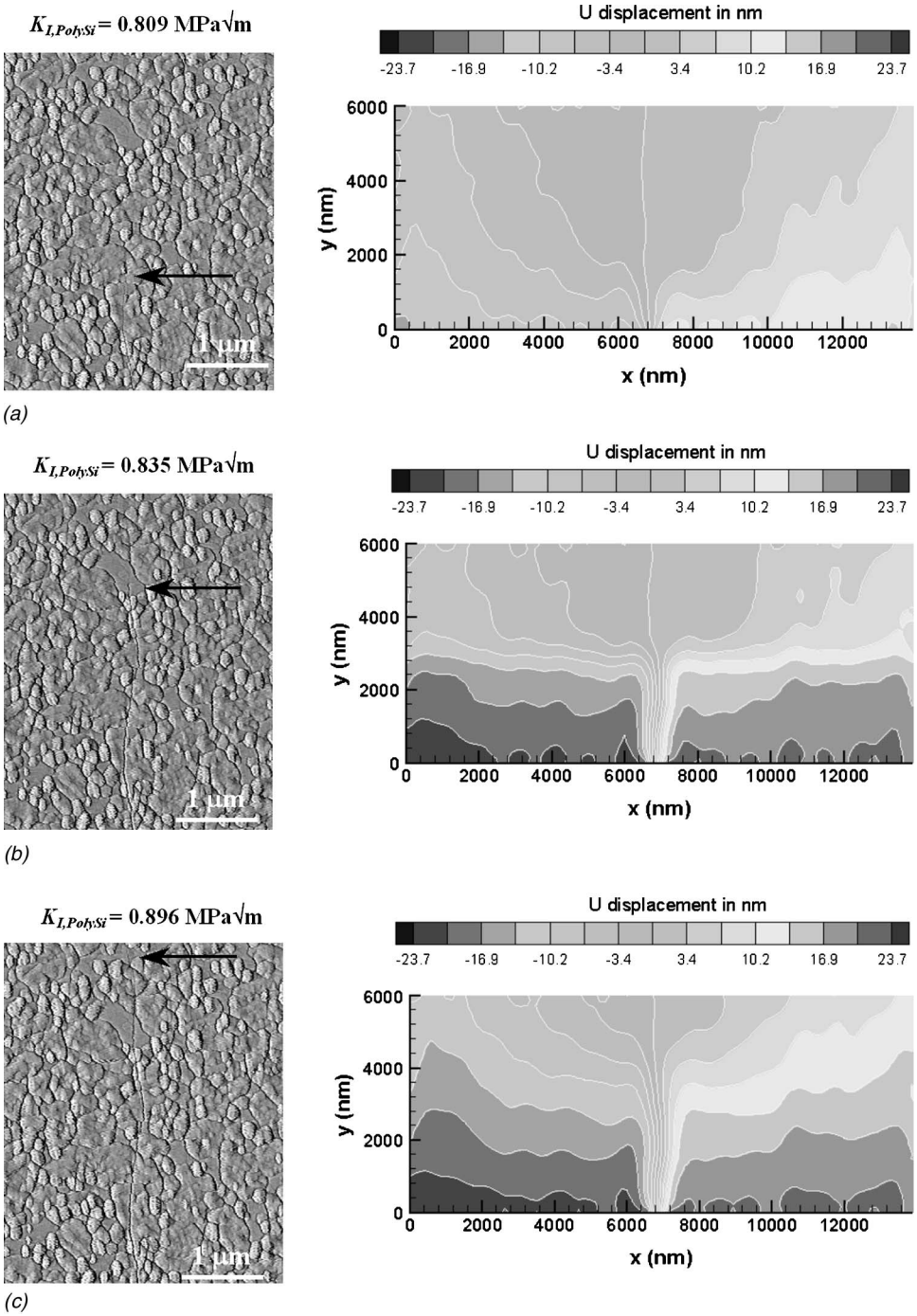


Fig. 6 AFM images showing subcritical (incremental) crack growth in polysilicon and the corresponding deformation fields. (*a*) Initial crack tip location at $K_{I,PolySi}$=0.809 MPa$\sqrt{m}$, (*b*) crack at $K_{I,PolySi}$=0.835 MPa$\sqrt{m}$ and (*c*) $K_{I,PolySi}$=0.896 MPa$\sqrt{m}$. The applied far field stress is normal to the path of crack propagation. After (*c*) the crack maintained its length until $K_{I,PolySi}$=1.063 MPa$\sqrt{m}$ when it grew catastrophically. The arrows point to the crack tip at each stress intensity level. Because of the fine crack line the AFM images were enhanced for publication purposes via an edge finder filter. The deformation field contour plots span wider fields of view compared to the AFM images. The left-to-right order of gray levels in the contour plots is the same as in the contour legends.

may occur at a crack tip located at a low energy cleavage plane while the grains ahead of the crack tip are oriented so that the local fracture toughness is larger than the local stress intensity factor of the extended crack. In that event a crack can initiate fracture and grow but subsequently arrest inside a neighboring grain where the local fracture toughness exceeds the local stress intensity factor.

This mode of incremental subcritical crack growth was directly evidenced via spatially resolved crack growth measurements using the AFM/DIC method. Figure 5 shows the details of a growing crack while traversing a large grain. The crack initially resided inside a grain that, for resolution purposes, is outside the field of view in Fig. 5(*a*) but will be pointed out later in Fig. 6(*a*). An increment in the effective stress intensity factor resulted in sub-

critical crack growth and arrest inside the large grain at the center of Fig. 5(*b*). Crack arrest occurred with a step at the tip of the crack but crack growth continued at the original crack plane in the next growth step. The presence of the large grain in the crack path (with not as favorable orientation of a low energy cleavage plane as the grain where the crack tip originally resided) was instrumental in crack arrest. Further increase in the effective stress intensity factor resulted in additional crack growth and arrest, as seen in Fig. 5(*c*). After the third step of crack growth a significant increase in stress intensity to $K_{I,\mathrm{PolySi}}=1.063\ \mathrm{MPa}\sqrt{\mathrm{m}}$ was required for catastrophic fracture initiation.

Figure 6 shows further details of the crack growth process supported by contour plots of local displacements near the crack tip. Incremental crack growth and arrest in individual polysilicon grains was recorded after $K_{I,\mathrm{PolySi}}=0.809\ \mathrm{MPa}\sqrt{\mathrm{m}}$ (Fig. 6(*a*)) and $K_{I,\mathrm{PolySi}}=0.835\ \mathrm{MPa}\sqrt{\mathrm{m}}$ (Fig. 6(*b*)). These values of $K_{I,\mathrm{PolySi}}$ were close to those for Si {111} cleavage planes ($0.83\ \mathrm{MPa}\sqrt{\mathrm{m}}$). After two increments of crack growth until $K_{I,\mathrm{PolySi}}=0.896\ \mathrm{MPa}\sqrt{\mathrm{m}}$ (Fig. 6(*c*)) catastrophic failure did not occur until $K_{I,\mathrm{PolySi}}=1.063\ \mathrm{MPa}\sqrt{\mathrm{m}}$, which was among the frequently recorded values of $K_{Ic,\mathrm{PolySi}}$. The crack tip displacement fields, Fig. 6, computed via DIC also point to the location of the crack tip at each crack growth increment.

This mode of crack growth is purely brittle and is not supported by plasticity because, at room temperature, crack initiation in silicon is not associated with dislocation activity [30] and as calculated before $r_p<5$ nm. Instead, it is attributed to the statistical variation of cleavage energy along the crack path, which may result in crack arrest depending on the value of $K_{Ic,\mathrm{local}}$. Crack initiation began at $K_{Ic,\mathrm{local}}$ with subsequent crack growth in neighboring grains. Crystal planes with cleavage energy higher than that of the plane at which fracture was initiated were encountered after $1-2\ \mu$m of crack extensions. At those grains, the crack arrested because $K_{I,\mathrm{local}}<K_{Ic,\mathrm{local}}$. Such a sudden arrest may occur without deceleration at values of stress intensity lower than the stress intensity factor that causes failure at quasistatic loading [31]. Although, from a material/device reliability viewpoint this incremental crack growth is not a desirable behavior, it provides support to enhanced macroscopic $K_{Ic,\mathrm{PolySi}}$ compared to that of single crystal silicon. In comparison to former studies on controlled subcritical crack growth in silicon and polysilicon, the phenomenon reported here is stochastic in nature and does not allow for stable crack growth.

3.3 Fractographic Studies of Fracture Cross Sections. Crack propagation was transgranular both during generation of precracks and catastrophic fracture of the freestanding samples. Although for particular polysilicon films fracture is intergranular supported by significant microcracking [17], in the present case no microcracking was evidenced for the resolution of an AFM and a SEM. During incremental crack growth in $1-2\ \mu$m steps the crack arrested with formation of a small pyramidal step on the crack plane as pointed by the white arrows in the SEM micrograph in Fig. 7. Step-like crack propagation was also evidenced in AFM images (Fig. 6(*b*)) in analogy to crack propagation in crystalline brittle materials [32]. This is due to silicon's strong preference for {111} cleavage planes [33,34]; when a crack is initiated on a {110} plane it will still shift to a {111} plane [20,21].

The path of crack propagation was relatively straight until crack branching occurred. If a specimen was loaded at an angle other than that normal to the crack plane then the crack would not propagate in a straight manner. Furthermore, if the precrack formed a kink then the initial portion of crack propagation would have been at an oblique angle with respect to the direction of the applied far-field force, although far from the kink the crack plane would still become normal to the applied far field stress. Any data from tests that followed the last two cases were omitted from the calculation of the average $K_{Ic,\mathrm{PolySi}}$.

The indentation precrack was planar in the vicinity of the crack tip and at a distance of about $2\ \mu$m. Figure 7 (and similarly 8(*a*)) attest the planarity of the crack near its tip. The crack surfaces were rough at the point of entry in the polysilicon specimen but they became smooth near the crack tip. Note that the SEM images of the crack surface after complete fracture in Figs. 7 and 8 are taken at a small angle with respect to the normal in the specimen cross section.

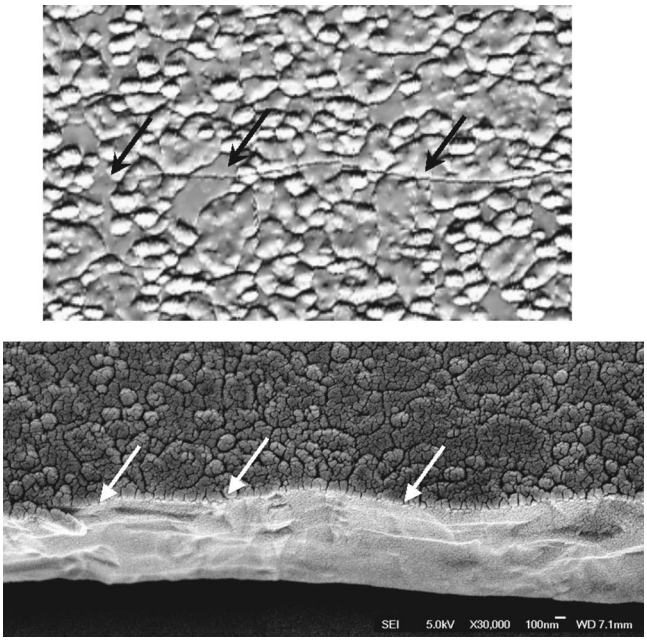


Fig. 7 Crack location during subcritical growth as seen in AFM (top image) and SEM (bottom image) micrographs that show the top specimen surface and the crack surface, respectively. The arrows point to the location of the crack at different macroscopic stress intensities according to Fig. 6.

A specimen cross section containing familiar microstructural aspects of brittle fracture is shown in Figs. 8(*a*)–8(*d*). The left end of each SEM image continues at the right end of the subsequent micrograph. The sequence of crack propagation regions, mirror area, mist, and hackle, was not clearly distinguished. The initial precracked region was characterized by smooth surfaces (Fig. 8(*a*)) that rapidly became rough as the crack propagated catastrophically (Fig. 8(*b*)), similar to other reports on brittle materials [31]. Only mirror surfaces were developed in the region of incremental crack growth, but soon after initiation of catastrophic crack growth a hackle region was evident. This rapid transition from mirror to hackle, dominated by lateral microscopic crack branching, is attributed to the small grain size, cleavage anisotropy, and the strong preference for crack propagation on specific crystal planes. The local microstructure mostly affected the local characteristics of the failure process inside individual grains but hackle patterns developed at scales considerably larger than the material microstructure (Figs. 8(*b*)–8(*d*)).

The crack plane roughness increased considerably with increasing K_I (Fig. 8(*c*)) until large crack surface undulations began (Fig. 8(*d*)). The latter represent attempts of crack branching and were indeed followed by crack branching that occurred only for wide specimens (>150-μm wide with initial crack lengths of $10-25\ \mu$m.) The location of crack branching occurred for a relatively consistent value of $\sigma_\infty\cdot r^{1/2}=2.1\pm0.4\ \mathrm{MPa}\sqrt{\mathrm{m}}$ (calculated from ten specimens) that did not show any particular trend with respect to specimen width or critical crack length. In the last relationship, r is the crack length before branching. The value of $\sigma_\infty\cdot r^{1/2}$ is similar to that for soda lime [35,36], while the branching patterns resembled those reported before for glass slides [32].

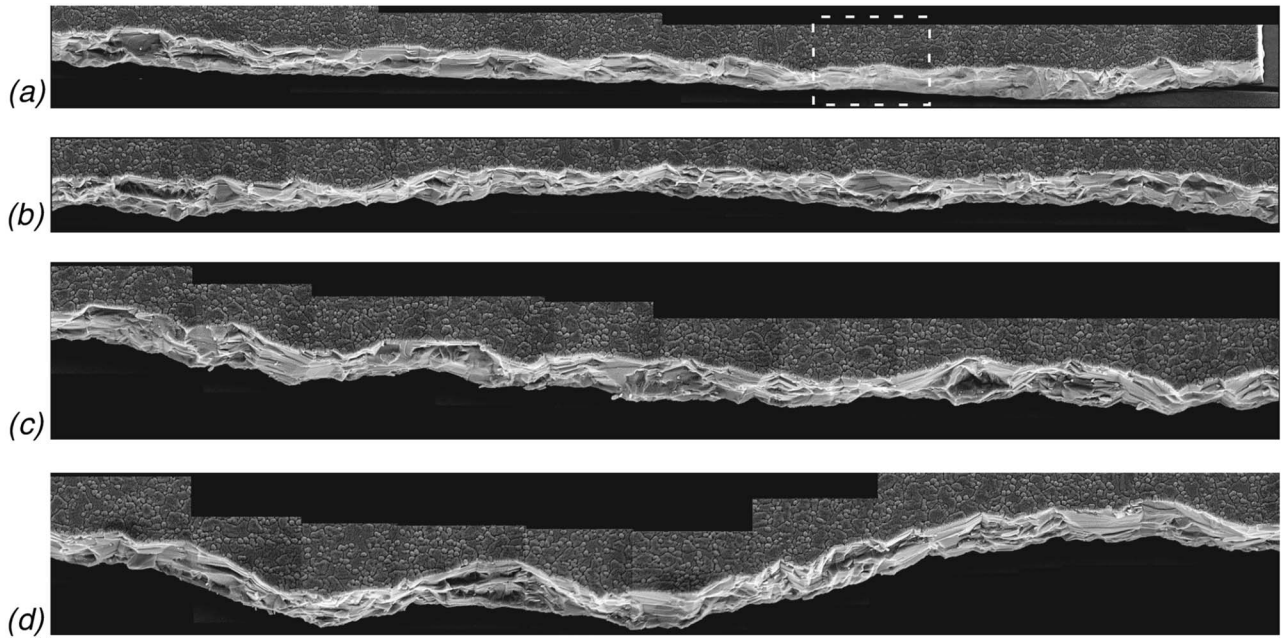


Fig. 8 SEM images of the fracture cross section of a polysilicon specimen. The section marked by dashed lines in (*a*) indicates the segment of incremental crack growth. After crack initiation hackle lines are apparent in (*b*) and (*c*), and the surface roughness increased dramatically resulting in crack branching in (*d*). Each cross section is comprised of a large number of high-resolution SEM images. The specimen cross sections are tilted with respect to the viewing angle. The specimen thickness is 2 μm.

4 Conclusions

The mode-I fracture behavior of polycrystalline silicon fabricated for MEMS by the MUMPs process was investigated. The values of the effective $K_{Ic,\mathrm{PolySi}}$ agreed with prior reports for nominally the same material, with similar microstructural characteristics (columnar structure and similar grain size), that was fabricated by a different facility [8]. Although agreed in average values, the experimental uncertainty in fracture toughness values reported in [7,8] varied between 0.04 MPa$\sqrt{\mathrm{m}}$ and 0.5 MPa$\sqrt{\mathrm{m}}$. The approach followed in this study resulted in experimental uncertainty on the order of 0.01 MPa$\sqrt{\mathrm{m}}$ owed to the use of an easy-to-calibrate load-measurement device. This small experimental uncertainty pointed to more precise bounds for $K_{Ic,\mathrm{PolySi}}$ that varied in the range of 0.84 MPa$\sqrt{\mathrm{m}}$ and 1.24 MPa$\sqrt{\mathrm{m}}$. This distribution of $K_{Ic,\mathrm{PolySi}}$ values was due to local cleavage anisotropy and relative orientation of the grains containing the crack tip with respect to the far field load, as well as due to grain boundaries and triple junction points near the crack tip.

Furthermore, the displacement fields in the vicinity of zero radius edge cracks were obtained for the first time at the scale of MEMS and were in good agreement with those predicted by LEFM. The small grain size and rather limited elastic anisotropy allowed for K dominance at the scale of a few microns. It is thus appropriate to state that since the fracture properties of polysilicon are not size dependent at the scale of a few microns (as opposed to its strength [37]), macroscopically applicable fracture mechanics models are still valid at the scale of most MEMS. The AFM-based experimental method implemented in this study supported the direct observation of subcritical (incremental) crack growth for $K_{I,\mathrm{PolySi}} < K_{Ic,\mathrm{PolySi}}$, which was attributed to variations in grain orientations along the crack path.

Although the average $K_{Ic,\mathrm{PolySi}}$ values differed only by 25% from those reported for bulk [20] there is still the need to test thin film specimens of the size of MEMS components. The thin film microstructure (grain size, and relatively columnar grain geometry) is different from bulk polysilicon (large equiaxed grains) and the large number of grains in the cross section of macroscopic specimens statistically minimize the influence of the details of individual grains on the fracture process. In thin polysilicon films, the crack tip plane is often located in a single columnar grain rather than spanning a large number of grains across the specimen thickness. As a consequence, polycrystalline anisotropy dominates in thin polysilicon films resulting in a range of effective stress intensity factors and subcritical crack growth, which have not been reported for bulk polysilicon.

Acknowledgment

The authors gratefully acknowledge the support by the Air Force Office of Scientific Research (AFOSR) through Grant No. F49620-03-1-0080 with Dr. B. L. Lee as monitor.

References

[1] Abdel-Tawab, K., and Rodin, G. J., 1998, "Fracture Size Effects And Polycrystalline Inhomogeneity," Int. J. Fract. **93**, pp. 247–259.
[2] Ballarini, R., Mullen, R. L., and Heuer, A. H., 1999, "The Effects of Heterogeneity and Anisotropy on the Size Effect in Cracked Polycrystalline Films," Int. J. Fract. **95**, pp. 19–39.
[3] Ballarini, R., Mullen, R. L., Yin, Y., Kahn, H., Stemmer, S., and Heuer, A. H., 1997, The Fracture Toughness of Polysilicon Microdevices: A First Report," J. Mater. Res. **12**(4), pp. 915–922.
[4] Kahn, H., Ballarini, R., Mullen, R. L., and Heuer, A. H., 1999, Electrostatically Actuated Failure of Microfabricated Polysilicon Fracture Mechanics Specimens," Proc. R. Soc. London, Ser. A **455**, pp. 3807–3823.
[5] Sharpe, W. N., Yuan, B., and Edwards, R. L., 1997, "Fracture Tests of Polysilicon Film," Mater. Res. Soc. Symp. Proc. **505**, pp. 51–56.
[6] Tsuchiya, T., Sakata, J., and Taga, Y., 1998, "Tensile Strength and Fracture Toughness of Surface Micromachined Polycrystalline Silicon Thin Films Prepared Under Various Conditions," Mater. Res. Soc. Symp. Proc. **505**, pp. 285–290.
[7] Kahn, H., Tayebi, N., Ballarini, R., Mullen, R. L., and Heuer, A. H., 2000, "Fracture Toughness of Polysilicon MEMS Devices," Sens. Actuators, A **82**, pp. 274–280.
[8] Kahn, H., Ballarini, R., and Heuer, A. H., 2001, "On the Fracture Toughness of Polysilicon MEMS Structures," Mater. Res. Soc. Symp. Proc. **657**, pp. 13–18.
[9] Keller, C., 1998, *Microfabricated High Aspect Ratio Silicon Flexures*, MEMS Precision Instruments, El Cerrito, CA, pp. 187–190.
[10] Chasiotis, I., and Knauss, W. G., 1998, "Investigation of Thin Film Mechanical Properties by Probe Microscopy," Proc. SPIE **3512**, pp. 66–75.
[11] Chasiotis, I., and Knauss, W. G., 2002, "A New Microtensile Tester for the Study of MEMS Materials with the aid of Atomic Force Microscopy," Exp. Mech. **42**(1), pp. 51–57.
[12] Cho, S. W., Chasiotis, I., Friedman, T. A., and Sullivan, J., 2005, "Direct

Measurements of Young's Modulus, Poisson's Ratio and Failure Properties of ta-C MEMS," J. Micromech. Microeng. **15**(4), pp. 728–735.
[13] Chasiotis, I., Cho, S. W., Jonnalagadda, K., and McCarty, A., 2004, "Fracture Toughness of Polycrystalline Silicon and Tetrahedral Amorphous Diamond-like Carbon (ta-C) MEMS," *Proceedings of the Society for Experimental Mechanics*, X International Congress, Costa Mesa, CA, June 7–10, Society of Experimental Mechanics, Bethel, CT, pp. 37–45.
[14] Chasiotis, I., 2004, "Mechanics of Thin Films and Microdevices," IEEE Trans. Device Mater. Reliab. **4**(2), pp. 176–188.
[15] Lawn, B., and Wilshaw, R., 1975, "Indentation Fracture: Principles and Applications," J. Mater. Sci. **10**, pp. 1049–1081.
[16] Chasiotis, I., and Knauss, W. G., 2003, "The Mechanical Strength of Polysilicon Films: 1. The Influence of Fabrication Governed Surface Conditions," J. Mech. Phys. Solids **51**, pp. 1533–1550.
[17] Chasiotis, I., and Knauss, W. G., 2003, "Experimentation at the Micron- and Submicron Scale," *Comprehensive Structural Integrity* (Vol. 8, Interfacial and Nanoscale Failure, W. Gerberich and W. Yang, eds.) Elsevier, New York, pp. 58, 66.
[18] Chasiotis, I., and Knauss, W. G., 2000, "Microtensile Tests with the Aid of Probe Microscopy for the Study of MEMS Materials," Proc. SPIE **4175**, pp. 96–103.
[19] Tada, H., Paris, P. C., and Irwin, G. R., 2000, *The Stress Analysis of Cracks Handbook*, 3rd ed., ASME Press, New York, pp. 52–53.
[20] Chen, C. P., and Leipold, M. H., 1980, "Fracture Toughness of Silicon," Am. Ceram. Soc. Bull. **59**, pp. 469–472.
[21] Yasutake, K., Iwata, M., Yoshii, K., Umeno, M., and Kawabe, H., 1986, "Crack Healing and Fracture Strength of Silicon Crystals," J. Mater. Sci. **21**, pp. 2185–2192.
[22] Ericson, F., Johansson, S., and Schweitz, J. A., 1988, "Hardness and Fracture Toughness of Semiconducting Materials Studied by Indentation and Erosion Techniques," Mater. Sci. Eng., A **105**, pp. 131–141.
[23] Abdel-Tawab, K., and Rodin, G. J., 1993, "On the Relevance of Linear Elastic Fracture Mechanics to Ice," Int. J. Fract. **62**, pp. 171–181.
[24] Umeno, Y., Kushima, A., Kitamura, T., Gumbsch, P., and Li, J., 2005, "Ab Initio Study of the Surface Properties and Ideal Strength of (100) Silicon Thin Films," Phys. Rev. B **72**, pp. 165431-1–165431-7.
[25] Komai, K., Minoshima, K., and Inoue, S., 1998, "Fracture and Fatigue Behavior of Single Crystal Silicon Microelements and Nanoscopic AFM Damage Evaluation," Microsyst. Technol. **5**, pp. 30–37.
[26] Xing, Y. M., Dai, F. L., and Yang, W., 2003, "Nano-moiré Method and Nanoscopic Crack Tip Deformation," *Comprehensive Structural Integrity* (Vol. 8, Interfacial and Nanoscale Failure, W. Gerberich and W. Yang, eds.), Elsevier, New York, p. 401.
[27] Kanninen, M. F., and Popelar, C. H., 1985, "Advanced Fracture Mechanics," Oxford University Press, New York, 172–180.
[28] Fang, Y., Ravi-Chandar, K., and White, K. W., 2002, "Influence of Surface Residual Stress State on Crack Path Evolution in Polycrystalline Alumina," J. Am. Ceram. Soc. **85**(7), pp. 1783–1787.
[29] Cho, S., Cárdenas-García, J. F., and Chasiotis, I., 2005, "Nano-Displacement Based Microscopic Hole Method for Determination of Thin Film Properties," Sens. Actuators, A **120**, pp. 163–171.
[30] Sumino, K., 1999, "Deformation Behavior of Silicon," Metall. Mater. Trans. A, **30A**, pp. 1465–1479.
[31] Ravi-Chandar, K., and Knauss, W. G., 1984, "An Experimental Investigation into dynamic Fracture: II Microstructural Aspects," Int. J. Fract. **26**, pp. 65–80.
[32] Lawn, B., 1994, *Fracture of Brittle Solids*, 2nd ed., Cambridge University Press, Cambridge, England, pp. 94, 198.
[33] Pérez, R., and Gumbsch, P., 2000, "Directional Anisotropy in the Cleavage Fracture of Silicon," Phys. Rev. Lett. **84**(23), pp. 5347–5350.
[34] Pérez, R., and Gumbsch, P., 2000, "An ab initio Study of the Cleavage Anisotropy in Silicon," Acta Mater. Metall. **48**(18–19), pp. 4517–4530.
[35] Mecholsky, J. J., 1994, "Quantitative Fractographic Analysis of Fracture Origins in Glass," *Fractography of Glass*, R. C. Bradt, and R. E. Tressler, eds., Plenum Press, NY, pp. 45.
[36] Mecholsky, J. J., Rice, R. W., and Freiman, S. W., 1974, "Predictions of Fracture Energy and Flaw Size in Glasses from Mirror Size Measurements," J. Am. Ceram. Soc. **57**(10), pp. 440–443.
[37] Chasiotis, I., and Knauss, W. G., 2003, "The Mechanical Strength of Polysilicon Films: 2. Size Effects Associated with Elliptical and Circular Perforations," J. Mech. Phys. Solids **51**, pp. 1551–1572.

Michal A. Brown
Department of Materials Science,
California Institute of Technology,
M/C 205-45,
Pasadena, CA 91125
e-mail: mabrown@Caltech.edu

Tae-Soon Park
Oraxion Diagnostics,
3077 Skyway Court,
Fremont, CA 94539

Ares Rosakis
Graduate Aeronautical Laboratories,
California Institute of Technology,
Pasadena, CA 91125

Ersan Ustundag
Department of Materials Science and Engineering,
Iowa State University,
2220 Hoover Hall,
Ames, IA 50011

Young Huang
Department of Mechanical Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801

Nobumichi Tamura

Bryan Valek

Advanced Light Source,
Lawrence Berkeley National Laboratory,
1 Cyclotron Road,
Berkeley, CA 94720

A Comparison of X-Ray Microdiffraction and Coherent Gradient Sensing in Measuring Discontinuous Curvatures in Thin Film: Substrate Systems

The coherent gradient sensor (CGS) is a shearing interferometer which has been proposed for the rapid, full-field measurement of deformation states (slopes and curvatures) in thin film-wafer substrate systems, and for the subsequent inference of stresses in the thin films. This approach needs to be verified using a more well-established but time-consuming grain orientation and stress measurement tool, X-ray microdiffraction (XRD). Both CGS and XRD are used to measure the deformation state of the same W film/Si wafer at room temperature. CGS provides a global, wafer-level measurement of slopes while XRD provides a local micromeasurement of lattice rotations. An extreme case of a circular Si wafer with a circular W film island in its center is used because of the presence of discontinuous system curvatures across the wafer. The results are also compared with a theoretical model based on elastic plate analysis of the axisymmetric bi-material film-substrate system. Slope and curvature measurements by XRD and by CGS compare very well with each other and with theory. The favorable comparison demonstrates that wafer-level CGS metrology provides a quick and accurate alternative to other measurements. It also demonstrates the accuracy of plate theory in modeling thin film-substrate systems, even in the presence of curvature discontinuities
[DOI: 10.1115/1.2150500]

Introduction and Motivation

As the semiconductor industry develops ever smaller dimensions of thin metal film interconnections and more complex multilayered (film stack) structures, the mechanical properties and stresses of thin films used for these structures become major factors in controlling the reliability of integrated circuits (ICs). However, due to the small size of features of modern ICs, the only way to measure their stresses directly is using X-ray microdiffraction (XRD) since spot sizes are now becoming available at comparable dimensions to such features (e.g., 1 μm and below). At the same time, it is impractical (very time-consuming and costly) to routinely perform many X-ray studies across large areas on a wafer. Therefore, it is desirable to develop another technique that can yield quick and full-field information regarding the stress state in film structures deposited on wafers, ideally to be used as a quality control tool. A recent optical method developed by Rosakis and co-workers at Caltech has been proposed for this purpose [1–3].

Called coherent gradient sensing (CGS), this technique employs an expanded laser beam to measure all components of the curvature tensor field in thin film-substrate systems. It does so in full-field and produces curvature component maps across the surface of the entire wafer. CGS is insensitive to rigid-body motions, such as those induced by vibration, and is quite suitable for in situ studies. As such, it is ideal for quick, full-field measurements of macro stresses across a wafer surface and can easily be used in quality control. It is further desirable, however, to attain stress state information at the level of individual features in a microelectronic circuit. In this context, Rosakis at Caltech, Suresh at MIT and their co-workers [4–7] have developed sophisticated micromechanical models that relate the data obtained from CGS (macro stresses) to local (micro) stresses in circuits.

In this article, we first describe the CGS and XRD techniques, and then begin the process of validating the CGS methodology.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received March 18, 2005; final manuscript received October 14, 2005. Review conducted by K. M. Liechti. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

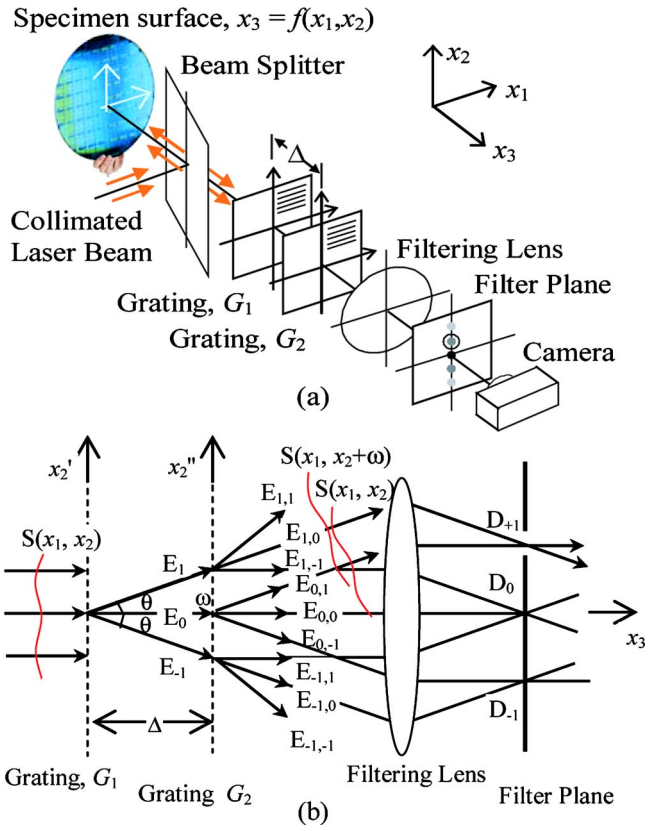


Fig. 1 Schematic of the CGS setup in reflection mode (*a*) and its working principle (*b*)

The first step, described here, is to compare curvature measurements made with both techniques on specimens with relatively simple features. In the present experiment, a sample consisting of a circular W thin film island deposited on a much thicker circular Si substrate is used. This radially symmetric geometry is particularly interesting because the discontinuous film thickness causes severe discontinuities in system curvatures across the island perimeter, which must be captured by both techniques for accurate measurement and comparison. Since the sample is circular, the bare Si portion has its own curvature which also must be measured. This sample is an excellent test specimen due to its radial geometry, large curvatures, and severe curvature discontinuities that develop because of the partial coverage of the Si substrate by the circular island. The radial symmetry of the system is also amenable to analytical treatment. Indeed an analysis of the relation between film stress and system deformations (slopes, curvatures) is possible based on plate analysis of the thin film-substrate system [8].

Coherent Gradient Sensing (CGS)

The coherent gradient sensing (CGS) method is a self-referencing interferometric technique that produces fringe patterns of surface slope by laterally shearing an initially planar wave front which has been reflected from a specimen (wafer). Figure 1(*a*) shows a schematic of the CGS setup in reflection [1–3]. A coherent, collimated laser beam (300 mm or less in diameter) is directed to a reflecting wafer surface via a beam splitter. In general, the wafer is nonplanar and its surface distortion can be described by the equation $x_3=f(x_1,x_2)$. The beam reflected from the wafer is also distorted by the nonplanar shape of the wafer. The resulting wave form is described by a two-dimensional surface in space whose equation is given by $x_3=S(x_1,x_2)$, where $S(x_1,x_2)=2f(x_1,x_2)$. This distorted wave front is again passed through the beam splitter and is then incident upon a pair of identical high-density gratings, G_1 and G_2, separated by a distance Δ. The gratings act to optically "shear" or "differentiate" the incident wave front to produce a series of diffracted beams. These beams are separated using a filtering lens to form distinct diffraction spots on a filter plane. An aperture placed in this plane serves to isolate the diffraction order of interest, which is then imaged onto the photographic film plane. For present purposes, either of the ±1 diffraction orders is of interest, as will be clear in the following discussion.

Figure 1(*b*) is a two-dimensional schematic illustrating the principles of the CGS method. The figure shows the distorted optical wave front, $S(x_1,x_2)$, incident on the two gratings in which the lines are taken to be oriented along x_1. At the first grating, G_1, the incident wave front is diffracted into several wave fronts, $E_0,E_1,E_{-1},E_2,E_{-2}$, etc., of which only the first three are drawn in Fig. 1(*b*). Each of these wave fronts, in turn, is diffracted by the second grating, G_2, to generate additional wave fronts, such as $E_{0,0},E_{0,1},E_{0,-1}$, etc. The diffracted beams are combined by a filtering lens to produce diffraction spots, such as D_0,D_{+1},D_{-1}, etc., in the focal plane of lens (filter plane). One of the diffraction spots, typically the first diffraction order, the D_{+1} spot, is chosen with an aperture for imaging onto the film plane.

The presence of the two gratings in the path of the optical wave front generates a lateral shift (or shearing) of the wave front. For example, the diffracted beam $E_{1,0}$, whose wave front is denoted as $S(x_1,x_2+\omega)$, is shifted from the beam $E_{0,1}$, whose wave front is denoted as $S(x_1,x_2)$, by a distance ω in the x_2 direction. The shift distance, ω, is expressed as $\Delta\tan\theta$, where $\theta=\arcsin(\lambda/p)$ is the diffraction angle and λ and p are the wavelength of light and the pitch of the gratings, respectively. For small angles of diffraction, $\omega\approx\Delta\theta\approx\Delta(\lambda/p)$, the condition for constructive interference of the original and shifted wave fronts is given by

$$S(x_1,x_2+\omega)-S(x_1,x_2)=n^{(2)}\lambda \quad n^{(2)}=0,\pm1,\pm2,\ldots \tag{1}$$

where $n^{(2)}$ is an integer that represents fringes associated with shearing along the x_2 direction. By dividing Eq. (1) by ω, taking ω to be sufficiently small, and substituting $\omega=\Delta(\lambda/p)$, it is seen that

$$\frac{\partial S(x_1,x_2)}{\partial x_2}=\frac{n^{(2)}\lambda}{\omega}=\frac{n^{(2)}p}{\Delta} \quad n^{(2)}=0,\pm1,\pm2,\ldots \tag{2}$$

Recalling that $S=2f$ and repeating the above-presented analysis for gratings aligned along the x_2 direction, it can be shown that the alternating dark and bright interference fringes correspond to constant values of components of the in-plane gradient of the wafer surface topography as follows:

$$\frac{\partial f(x_1,x_2)}{\partial x_\alpha}=\frac{n^{(\alpha)}p}{2\Delta} \quad n^{(\alpha)}=0,\pm1,\pm2,\ldots \tag{3}$$

where $\alpha,\beta\in\{1,2\}$. A relative rotation of the gratings to the wafer allows for both orthogonal components of slope to be recorded in the form of full-field slope maps. The three independent components of curvature tensor field, $\kappa_{\alpha\beta}$, can now be determined directly from two orthogonal CGS slope maps by partial differentiation along the x_1 and x_2 directions

$$\kappa_{\alpha\beta}(x_1,x_2)\approx\frac{\partial^2 f(x_1,x_2)}{\partial x_\alpha\partial x_\beta}\approx\frac{p}{2\Delta}\left\{\frac{\partial n^{(\alpha)}(x_1,x_2)}{\partial x_\beta}\right\} \tag{4}$$

In order to determine the full curvature tensor, the gradient fields in two orthogonal directions must be recorded. Equation (4), which applies to the shear (or twist) curvature component, κ_{12}, as well as the normal curvature components, κ_{11} and κ_{22}, is the equation governing the curvature tensor field at any in-plane location (x_1,x_2). It enables the global, full-field measurement of curvature for the film-substrate system. For thin film-thick substrate systems the full-field recording of all system curvature components is crucial since they can be related to the individual components of

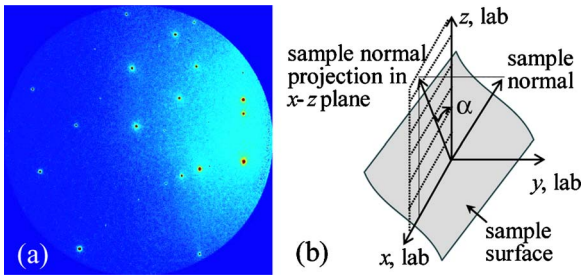


Fig. 2 (*a*) **Laue pattern from the single crystal Si wafer.** (*b*) **Definition of coordinate system and the projection angle α; slope in *xz* plane=tan(α).**

stress acting on the thin film, through simple plate theory [4–7]. This provides an easy and quick way of film stress measurement that can be instantaneously performed across an entire wafer surface.

X-Ray Microdiffraction (XRD)

X-ray microdiffraction is a local measurement technique which, in general, uses the lattice spacing in a crystalline structure as a local micro-strain gauge. The incoming beam diffracts from the crystalline lattice to form Laue patterns, which can be analyzed to measure the sample strain and stress. Synchrotron radiation, which was used in this experiment, has several advantages over traditional, commercially available laboratory-grade X-ray machines. These include its very small beam size ($\sim 1\times 1$ μm), higher energy, and ability to use a polychromatic (white) X-ray beam. In contrast, most lab X-ray machines feature large, monochromatic beam spot sizes of more than 100×100 μm and are thus unable to resolve spatial stress occurring in thin film structures. Also, while a polychromatic beam is used to determine the orientation and 3D deviatoric strain of each grain illuminated, a monochromatic beam takes an average measurement over all diffracting grains.

Laue patterns from a white beam form Laue spots (Fig. 2(*a*)), where each spot is generated by a specific X-ray energy selected by the given lattice plane. A sophisticated software program then deconvolutes these patterns and indexes them, identifying individual patterns from each grain [9]. From this analysis, one can determine the orientation matrix of each grain as well as its deviatoric strain. (The deviatoric stress is then found using Hooke's law [10].) When there are too many grains in an image, the software is unable to determine which spots correspond to which grains, and cannot analyze the image. Therefore, white x rays are used when the grain size is comparable to or larger than the beam size.

In the case of a single crystal specimen, the orientation matrix measured is always from the same grain. Once the crystal orientation is obtained at each location across the specimen, the relative slope and curvature are then determined by tracking the changes in the vector defining the grain normal with respect to the lab coordinate system. For a scan along the x axis (sample diameter), we are only concerned with the slope changes in the xz plane. This slope is equal to tan(α), where α is defined as the angle between the projection of the grain normal in the xz plane and the z axis in the lab reference frame (Fig. 2(*b*)).

For a radially symmetric sample on which the scan is performed along the diameter, where $y=0$, cylindrical coordinates can be used. The radial slope, $\partial f/\partial r=\tan(\alpha)$, and the circumferential curvatures κ_{rr} and $\kappa_{\theta\theta}$ are then determined from

$$\kappa_{rr}=\frac{\partial^2 f}{\partial r^2}=\frac{\partial(\tan\alpha)}{\partial r} \tag{5}$$

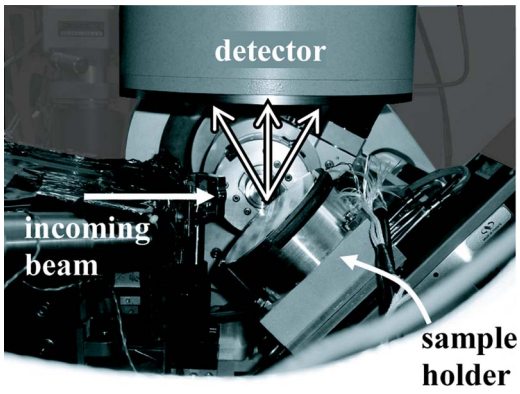


Fig. 3 The microdiffraction setup at the Advanced Light Source. The incoming X-ray beam is reflected from the sample surface and captured by the detector.

$$\kappa_{\theta\theta}=\frac{1}{r}\frac{\partial f}{\partial r}=\frac{1}{r}(\tan\alpha). \tag{6}$$

The experiment described here was performed at beamline 7.3.3 at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory in Berkeley, CA. At this beamline we can measure curvature and strain in the silicon substrate, as well as strain in thin films [9]. The reflection X-ray setup is shown in Fig. 3. In this experiment, white X-rays were used to measure the grain normal orientation across the sample.

Experiments

The same sample was used for both CGS and XRD measurements. This sample consisted of a 2-μm-thick, 25-mm-diam circular W film island in the center of a 100-mm-diam, 525-μm-thick Si 001 wafer (Fig. 4). The Young's modulus for Si and W are, respectively, 130 and 410 GPa, and the Poisson's ratio is 0.28 for both materials [11].

Both CGS and XRD data were collected at room temperature. Using the CGS method, slope, curvature, and stress maps of the sample were obtained in advance. At ALS, pointwise XRD measurements were made along almost the full diameter of the sample, 94.5 mm, with a 0.1 mm step size, yielding the single-crystal Si grain orientation at each point.

As discussed earlier, the geometry of the film-substrate system was chosen in such a way as to feature severe slope and curvature gradients across the wafer in the form of curvature discontinuities occurring at the edge of the film island. It was felt that the accurate recording of curvature jumps across such discontinuities will provide a test for validating CGS metrology through comparison to XRD. The spatial resolution of the XRD measurements (order of 1 μm) was definitely enough to cope with the abrupt spatial variations anticipated. Moreover, the geometry was intentionally kept simple enough so as to allow for straightforward scientific interpretation of the results. Furthermore, it was chosen in such a way as to allow for the development of closed-form analytical relations describing the substrate deformation. The direct comparison of the results of the two measurements with the results of

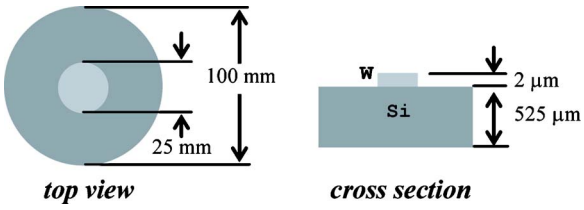


Fig. 4 Sample schematic

the analysis was anticipated to provide further guidance on the interpretation of the measured behavior and to shed light on the ability of elastic plate theory to analyze film-substrate systems which feature discontinuous film geometries.

Theoretical Predictions

The axisymmetric nature of the film substrate system allows for elastic plate theory and the thin film approximation to relate the system curvature fields to the stress in the circular island film. The theoretical development, whose details will be presented elsewhere [8], considers a circular film of thickness h_f and radius R_f deposited in the center of a circular substrate of thickness h_s and radius R_s. The film thickness is much smaller than that of the substrate ($h_f \ll h_s$). The Young's moduli of the film and the substrate are denoted by E_f and E_s, respectively, while their Poisson's ratios are denoted by ν_f and ν_s.

The analysis considers the special case in which both film and substrate are composed of isotropic linear elastic solids, characterized by the above-noted moduli. It is further assumed that the film features an in-plane isotropic, or equibiaxial, stress state characterized by a film stress σ_f such that $\sigma_{xx}=\sigma_{yy}=\sigma_f$ and $\sigma_{zz}=0$ in any proper orthogonal coordinate frame such that the x and y axes lie on the plane of the film (or the wafer). Alternatively, if a polar coordinate system centered at the island center is used, this is equivalent to $\sigma_{rr}=\sigma_{\theta\theta}=\sigma_f$ where σ_{rr} and $\sigma_{\theta\theta}$ are the radial and circumferential stress components, respectively. Perhaps the most severe restriction of the analysis is that stresses are not allowed to vary across the film (i.e., $\partial\sigma_f/\partial r=0$) and are restricted to remain spatially constant.

Under the above-mentioned assumptions an elastic boundary value problem can be formulated which relates the radially varying system curvatures to the constant and equibiaxial film stress σ_f. Because of the system's radial symmetry, the only nonzero component of the slope vector is $\partial f/\partial r$ while the only nonvanishing components of the curvature tensor are $\kappa_{rr}=\partial^2 f/\partial r^2$ and $\kappa_{\theta\theta}=(\partial f/\partial r)/r$, the radial and circumferential curvatures, respectively. Axisymmetry requires that the twist curvature component $\kappa_{r\theta}$ vanishes throughout the system.

The analysis shows that the sample curvature within the film-covered region is spatially constant and equibiaxial, i.e., $\kappa_{rr}=\kappa_{\theta\theta}=\kappa$, where κ is independent of the radial position r. In this region the sample curvatures are given by

$$\kappa_{rr}=\kappa_{\theta\theta}=\kappa=\frac{6h_f\sigma_f(1-\nu_s)}{E_s h_s^2}\left[1-\frac{1-\nu_s}{2}\left(1-\frac{R_f^2}{R_s^2}\right)\right] \quad 0\leqslant|r|\leqslant|R_f| \tag{7}$$

In this region the system deforms as a sphere. It should be noted that for $R_f=R_s$ (full film coverage of the wafer) Eq. (7) coincides with the predictions of the well-known Stoney formula [12] which relates constant film stress to constant system curvatures as described in Ref. [4]. This is an anticipated limiting behavior.

Outside the film covered region, however, the predictions are quite unexpected. Here the system curvatures κ_{rr} and $\kappa_{\theta\theta}$ are not equal and are strong functions of the radial position r. Despite the fact that this region is not covered by the film, its curvature components are nonzero and their magnitude depends on the magnitude of the stress of the film island as

$$\kappa_{rr}=\frac{\partial^2 f}{\partial r^2}=\kappa\frac{1-\dfrac{1+\nu_s}{1-\nu_s}\dfrac{R_s^2}{r^2}}{1+\dfrac{1+\nu_s}{1-\nu_s}\dfrac{R_s^2}{R_f^2}} \qquad \kappa_{\theta\theta}=\frac{1}{r}\frac{\partial f}{\partial r}=\kappa\frac{1+\dfrac{1+\nu_s}{1-\nu_s}\dfrac{R_s^2}{r^2}}{1+\dfrac{1+\nu_s}{1-\nu_s}\dfrac{R_s^2}{R_f^2}}$$
$$|R_f|\leqslant|r|\leqslant|R_s| \tag{8}$$

What is also remarkable is that comparison of the expressions for κ_{rr} from within and outside the film-covered region reveals a finite jump in radial curvature which involves a change of curvature sign across the circular interface $r=R_f$.

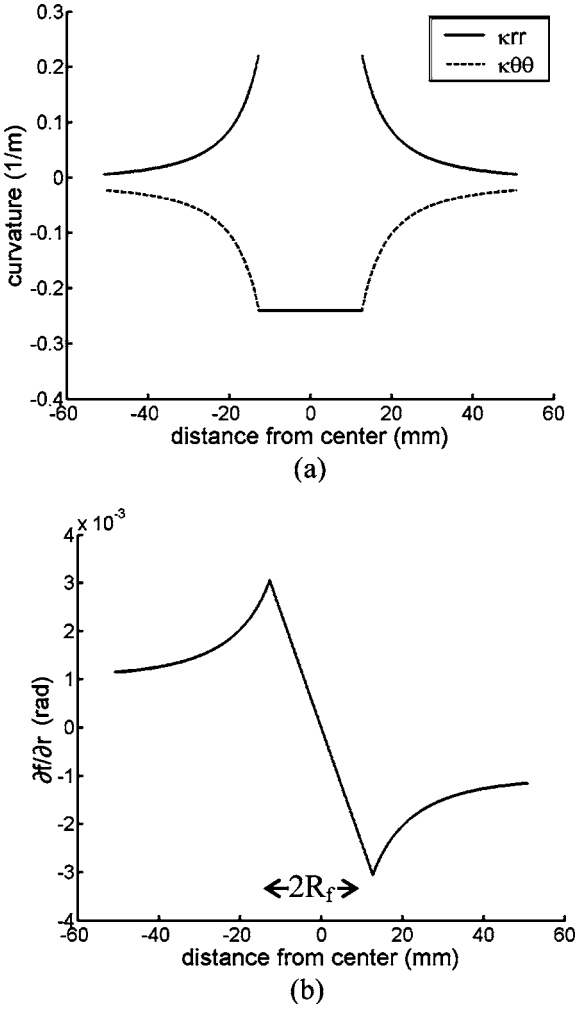


Fig. 5 Theoretical prediction of surface (*a*) curvature and (*b*) slope across the diameter of a radially symmetric circular wafer with a circular film island in the center. The assumed film stress is –2 GPa.

The radial slope component $\partial f/\partial r$ can now be computed from Eqs. (7) and (8) and is given by the following relations for the two regions:

$$\frac{\partial f}{\partial r}=\kappa_{\theta\theta}r=\kappa_{rr}r=\kappa r \quad 0\leqslant|r|\leqslant|R_f| \tag{9}$$

$$\frac{\partial f}{\partial r}=\kappa_{\theta\theta}r=\kappa\frac{1+\dfrac{1+\nu_s}{1-\nu_s}\dfrac{R_s^2}{r^2}}{1+\dfrac{1+\nu_s}{1-\nu_s}\dfrac{R_s^2}{R_f^2}}r \quad |R_f|\leqslant|r|\leqslant|R_s| \tag{10}$$

Within the film-covered region, there is a linear variation of slope with position while the variation of slope outside the film-covered region is more complex. At the interface $r=R_f$, the slopes are, as expected, continuous. Figure 5 shows the predicted variation of the two nonzero curvature components and the radial slope component. The figure is plotted by using the geometry and material parameters of the above-described W-Si system and an assumed film stress, σ_f, of −1.5 GPa. The above-discussed theoretically predicted features are obvious from the figure.

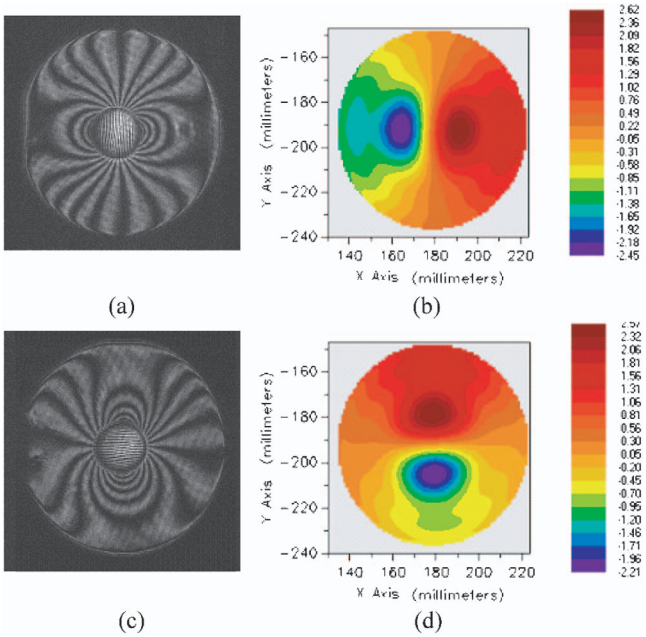


Fig. 6 CGS slope measurements in the *x* and *y* directions. Horizontal, $\partial f/\partial x$: (*a*) wafer image and (*b*) horizontal slope map. Vertical, $\partial f/\partial y$: (*c*) wafer image and (*d*) vertical slope map.

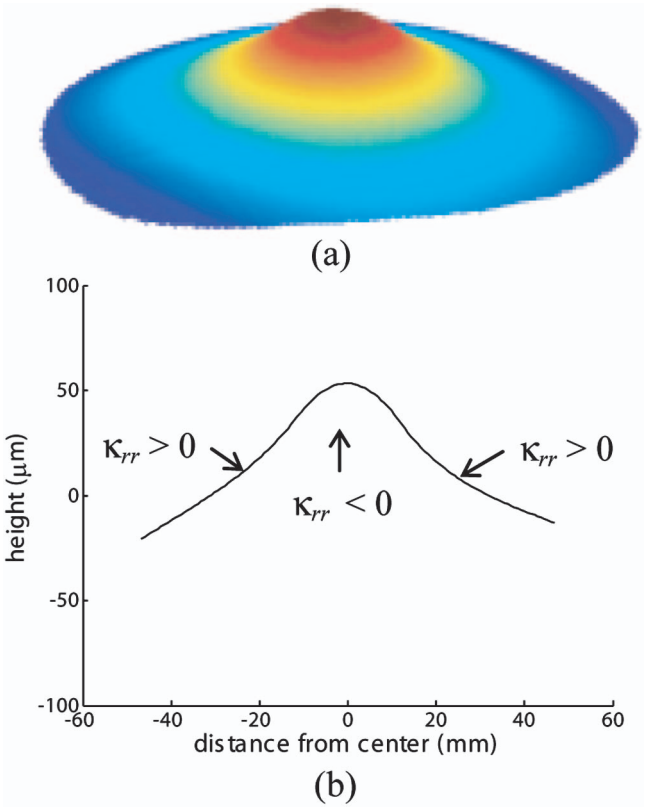


Fig. 7 Wafer topography, through integration of *x* and *y* slope maps: (*a*) full-field map; (*b*) radial cut through $y=0$

Results and Discussion

By using CGS interferometry, the same sample was measured to obtain slope interferograms and digitized full-field slope maps in the x and y directions (Fig. 6). Since these are in Cartesian coordinates, only the data along the diameters, the extracted linear data set from these maps, can be directly converted to radial slopes. It is worth noting that the Cartesian slope maps in each orthogonal direction differ by only a 90 deg rotation, as is expected for a radially symmetric sample.

The film island is distinctly visible on the x and y interferograms (Figs. 6(a) and 6(c)) as a circle in the middle of the picture. The fringe pattern in that region consists of dense, straight, evenly spaced lines that correspond to a constant equibiaxial curvature, as predicted by the analysis. Outside the film island, there is still a less dense but more complex fringe pattern, since there is a nonzero curvature in that area due to the circumferential constraints on the Si wafer. This is also qualitatively consistent with the analysis.

Further image analysis can at this point proceed in two ways—either to the wafer shape, through integration of the slope maps, or to curvature maps, through differentiation of slope components. Integration of the slope maps yields the surface topography of the wafer, as shown in Fig. 7. Consistent with theory (Fig. 5), the sign of the radial curvature component κ_{rr} changes across the film edge. The film region has a negative, constant curvature; there is a jump in curvature from negative to positive at the film edge; and there is a nonzero curvature outside the film.

The slope maps can also be used to calculate the two direct (κ_{xx} and κ_{yy}) and the twist (κ_{xy}) Cartesian curvature maps (Fig. 8). Again, the x and y curvature maps are very similar, with the noticeable difference being due to the graph color scale. Also note the diagonal symmetry in the twist curvature map. Both of these indicate a radially symmetric curvature state, as expected from the radial sample geometry.

The radial symmetry is verified by calculating the maximum, principal curvature map (Fig. 8(d)) by using the following relation [3]:

$$\kappa_{\max} = \left(\frac{\kappa_{xx} + \kappa_{yy}}{2}\right) + \frac{1}{2}\sqrt{(\kappa_{xx} - \kappa_{yy})^2 + 4\kappa_{xy}^2} \tag{11}$$

The obvious radial symmetry of this map shows that the axisymmetry assumption of the theory is relevant in this case and implies that $\kappa_{\max} = \kappa_{rr}$.

In order to obtain curvature by differentiation of the slope maps, the data are fit to a polynomial function by a numerical analysis. This polynomial fitting uses Zernike polynomials, a fifth-order polynomial set with 37 terms. While originally de-

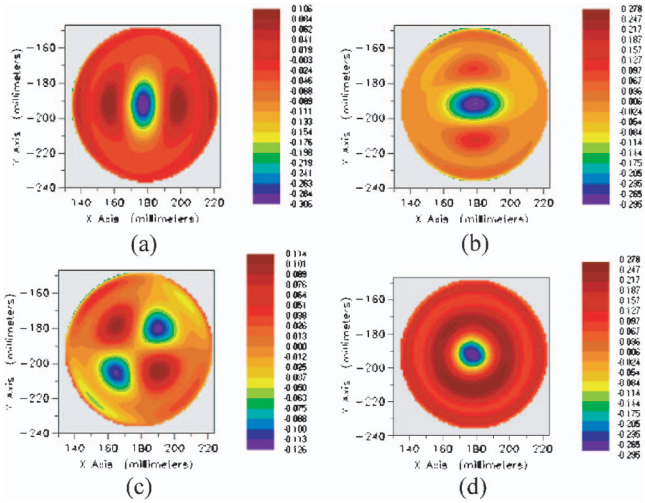


Fig. 8 CGS curvature maps in the *x*, *y* and twist directions: (*a*) horizontal map, $\kappa_{xx} = \partial^2 f/\partial x^2$, (*b*) vertical map, $\kappa_{yy} = \partial^2 f/\partial y^2$, (*c*) twist map, $\kappa_{xy} = \partial^2 f/\partial x \partial y$, (*d*) principal curvature, $\kappa_{\max} = \kappa_{rr}$

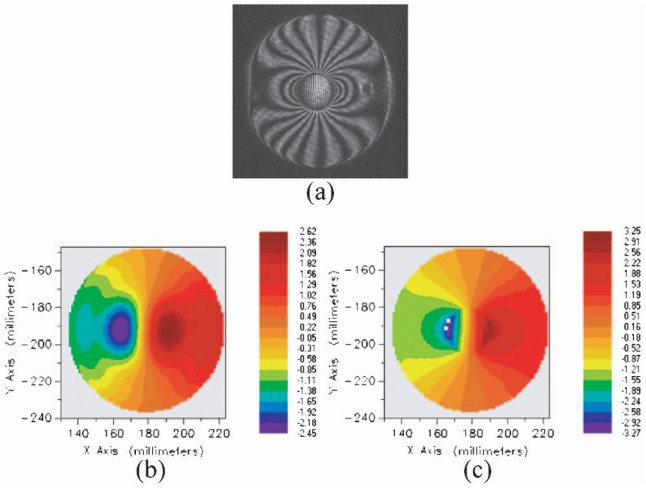


Fig. 9 (*a*) Horizontal interferogram and slope maps: (*b*) with filtering and smoothing and (*c*) with no smoothing (raw data)

signed to characterize optical aberrations, this polynomial set is often used to describe aspheric surfaces from interferometric data. In most cases, where spatial variations of curvature are gradual, this fitting process is advantageous since it filters out noise and allows numerical differentiation to be performed. This particular sample, however, has a highly nonuniform geometry, which corresponds to unusually high gradients in slope and jumps in curvature which are lost in the numerical smoothing process. Gradients in slope this large are not typically seen in modern industrial thin

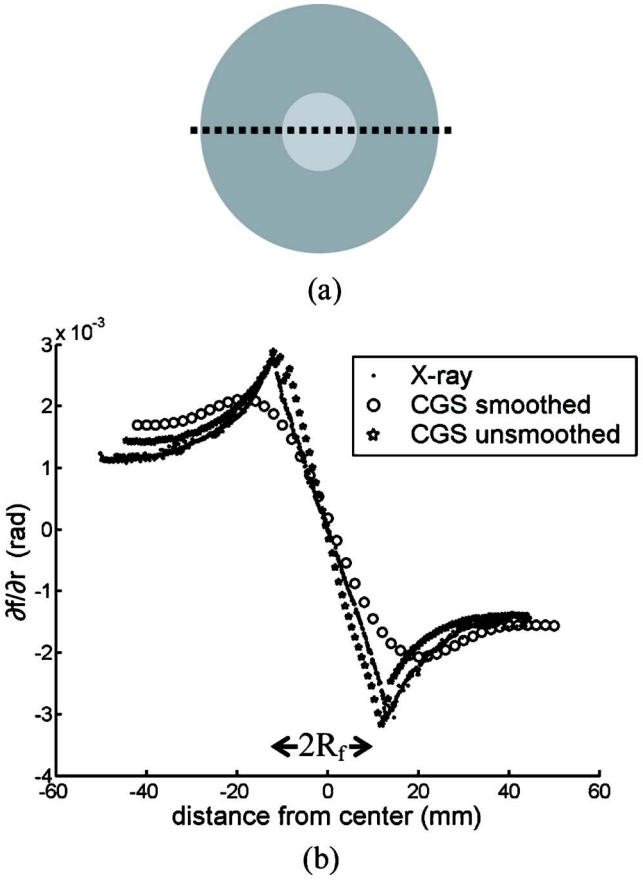


Fig. 10 Data were extracted from CGS slope maps across the sample diameter (*a*) and compared with XRD (*b*)

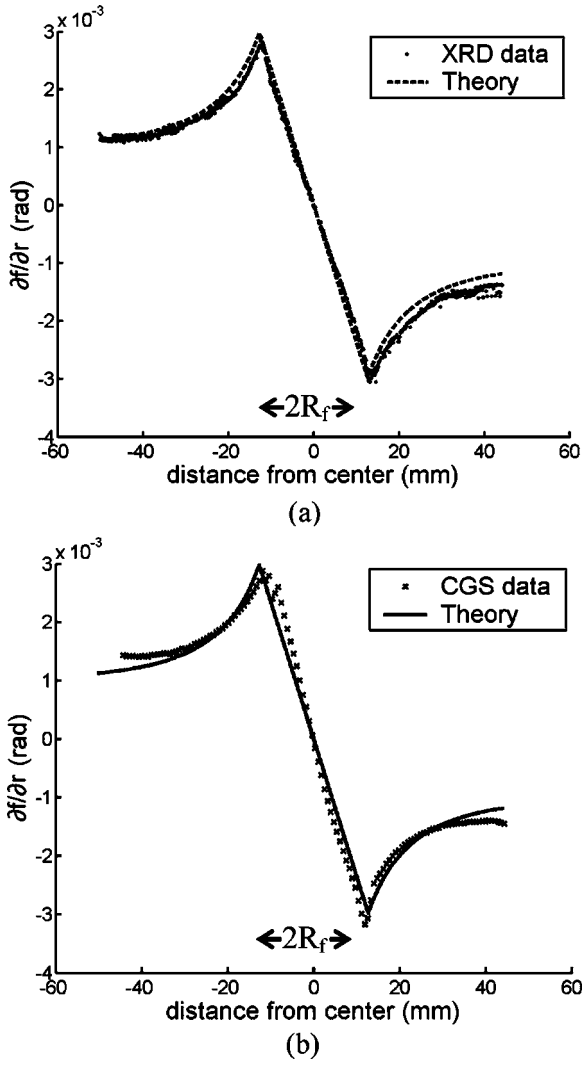


Fig. 11 Comparison of (*a*) XRD and (*b*) CGS data with theoretical predictions, using film stress as the fitting parameter

film-substrate systems, which feature small, closely spaced features, and therefore this over-smoothing situation does not, in practice, generally occur.

In this case, the film-covered region should have a constant curvature, or linear slope, which corresponds with straight, evenly spaced vertical lines on the horizontal slope map (Fig. 9). There should also be a sharp transition between the slope of the film covered and that of the bare Si regions. Instead, Fig. 9(*b*) features only approximately vertical lines in the film-covered regions, and only a gradual change to the bare Si region. To circumvent this problem, the numerical smoothing was artificially removed and the resulting unsmoothed digital slope map is displayed in Fig. 9(*c*). In this figure (with no filtering or smoothing), the anticipated straight evenly spaced lines within the film are clearly visible. The sharp transition between the constant and varying curvature inside and outside the film is also recorded. In fact, the film-covered circular island can be easily traced out on this map.

In order to directly compare CGS slope measurements with the XRD data taken along the sample diameter, a linear data set, also selected across the same diameter of the sample (Fig. 10(*a*)), was extracted from both slope maps in Fig. 6. Figure 10(*b*) shows the difference between the smoothed and unsmoothed CGS data, compared with XRD across the same region. In both the XRD and the unsmoothed CGS data, there is a sharp transition across the film island edges. This transition from one slope regime to the

other is predicted by theory in Eqs. (9) and (10) and is due to the discontinuous film geometry as discussed in the previous section. By contrast, the smoothed CGS slope data completely miss this sharp transition by oversmoothing the raw interferogram.

The rather remarkable agreement between the raw CGS data and the XRD measurements provides a clear indication that CGS interferometry, if properly interpreted, can provide an excellent means of measuring deformations even at the presence of severe geometric discontinuities. It should be noted at this point that at the vicinity of the film-covered area, X-ray microdiffraction data were taken every 0.1 mm, which accounts for the high density of data reported in Fig. 10(*b*). A total of 940 measurements were taken across the wafer diameter. This measurement, although very spatially resolved, was very tedious and took a total of 3 h to complete. In contrast, the CGS measurement was almost instantaneous and was only limited by the acquisition time of the CCD camera used to record the interferograms.

The XRD and CGS slope data were also compared with theoretical predictions (Fig. 11), using the film stress as the single fitting parameter in Eq. (11). The comparison is quite good—for a given film stress, the shape and magnitude of the slope curve for XRD data, CGS data and theoretical prediction are all extremely close. This implies that even in the presence of strong curvature discontinuities of the type displayed in Fig. 5, plate theory does a remarkable job in capturing the wafer deformation behavior.

Conclusions

Coherent gradient sensing (CGS) interferometry provides a full-field, real-time, in situ slope and curvature measurement over the entire wafer surface. Nonuniform deformations have been measured using CGS interferometry in a partially covered film (island pattern). Highly resolved X-ray microdiffraction has been used to obtain slope and substrate curvature data from the same wafer for validation of the CGS measurement. Slope and curvature values measured from CGS interferometry are highly consistent with X-ray microdiffraction data and also with theoretical predictions of discontinuous curvature states.

Acknowledgment

The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF0098 at Lawrence Berkeley National Laboratory.

References

[1] Rosakis, A. J., Singh, R. P., Tsuji, Y., Kolawa, E., and Moore, N. R., Jr., 1998, "Full-Field Measurements of Curvature Using Coherent Gradient Sensing: Application to Thin Film Characterization," Thin Solid Films, **325**(1-2), pp. 42–54.

[2] Lee, H., Rosakis, A. J., and Freund, L. B., 2001, "Full-Field Optical Measurement of Curvatures in Ultra-Thin-Film-Substrate Systems in the Range of Geometrically Nonlinear Deformations," J. Appl. Phys., **89**(11), pp. 6116–6129.

[3] Park, T. S., Suresh, S., Rosakis, A. J., and Ryu, J., 2003, "Measurement of Full-Field Curvature and Geometrical Instability of Thin Film-Substrate Systems Through CGS Interferometry," J. Mech. Phys. Solids, **51**(11-12), pp. 2191–2211.

[4] Freund, L. B., and Suresh, S., 2003, *Thin Film Materials*, Cambridge University Press, Cambridge, UK, Chaps. 2,3.

[5] Park, T. S., and Suresh, S., 2000, "Effects of Line and Passivation Geometry on Curvature Evolution During Processing and Thermal Cycling in Copper Interconnect Lines," Acta Mater., **48**(12), pp. 3169–3175.

[6] Park, T. S., Dao, M., Suresh, S., Rosakis, A. J., Pantuso, D., and Shankar, S., 2005, "Some Practical Issues of Curvature and Thermal Stress in Realistic Multi-level Metal Interconnect Structures," J. Electron. Mater. (submitted).

[7] Rosakis, A. J., Park, T. S., and Suresh, S., 2005, "Analytical and Numerical Modeling of Stress in Multi-level Interconnect Structures and Vias," Thin Solid Films (submitted).

[8] Rosakis, A. J., and Huang, Y., 2005, "A Thermoelastic Analysis of Film Stress and Non-Uniform Deformations in Wafers With Isolated and Periodic Film Islands" (unpublished).

[9] Tamura, N., MacDowell, A. A., Spolenak, R., Valek, B. C., Bravman, J. C., Brown, W. L., Celestre, R. S., Padmore, H. A., Batterman, B. W., and Patel, J. R., 2003, "Scanning X-Ray Microdiffraction With Submicrometer White Beam for Strain/Stress and Orientation Mapping in Thin Films," J. Synchrotron Radiat., **10**, pp. 137–143.

[10] Noyan, I. C., and Cohen, J. B., 1987, *Residual Stress*, Springer-Verlag, New York, Chap. 5.

[11] Gouldstone, A., Shen, Y.-L., Suresh, S., and Thompson, C. V., 1998, "Evolution of Stresses in Passivated and Unpassivated Metal Interconnects," J. Mater. Res., **13**(7), pp. 1956–1966.

[12] Stoney, G. G., 1909, "The Tension of Metallic Films Deposited by Electrolysis," Proc. R. Soc., London, **82**, pp. 172–175.

Mechanical Characterization of Released Thin Films by Contact Loading

Rongjing Zhang
Division of Engineering and Applied Science,
California Institute of Technology,
Pasadena, CA 91125

Doron Shilo
Faculty of Mechanical Engineering,
Technion—Israel Institute of Technology,
Haifa 32000, Israel

Guruswami Ravichandran

Kaushik Bhattacharya

Division of Engineering and Applied Science,
California Institute of Technology,
Pasadena, CA 91125

The design of reliable micro electro-mechanical systems (MEMS) requires understanding of material properties of devices, especially for free-standing thin structures such as membranes, bridges, and cantilevers. The desired characterization system for obtaining mechanical properties of active materials often requires load control. However, there is no such device among the currently available tools for mechanical characterization of thin films. In this paper, a new technique, which is load-controlled and especially suitable for testing highly fragile free-standing structures, is presented. The instrument developed for this purpose has the capability of measuring both the static and dynamic mechanical response and can be used for electro/magneto/thermo mechanical characterization of actuators or active materials. The capabilities of the technique are demonstrated by studying the behavior of 75 nm *thick amorphous silicon nitride* (Si_3N_4) *membranes. Loading up to very large deflections shows excellent repeatability and complete elastic behavior without significant cracking or mechanical damage. These results indicate the stability of the developed instrument and its ability to avoid local or temporal stress concentration during the entire experimental process. Finite element simulations are used to extract the material properties such as Young's modulus and residual stress of the membranes. These values for* Si_3N_4 *are in close agreement with values obtained using a different technique, as well as those found in the literature. Potential applications of this technique in studying functional thin film materials, such as shape memory alloys, are also discussed.* [DOI: 10.1115/1.2166652]

1 Introduction

An important aspect in advancing the application of Micro/Nano Electro-Mechanical Systems (MEMS/NEMS) is to increase the reliability of the component devices. This largely depends on the knowledge and understanding of the mechanical behavior of the component materials [1–3]. However, the mechanical characterization of active materials and associated MEMS devices has lagged behind the corresponding theory and design software [1]. Standard methods for characterization have yet to be established. For these purposes, mechanical characterization instruments, which allow testing structures with small dimensions such as thin free standing films, membranes, and cantilevers, are required. The interest in reducing the dimensions of these structures to increase the volume density warrants load and displacement sensitivity in the sub-mN and sub-μm scales. These requirements are not met by the traditional mechanical characterization tools, such as the large scale materials testing machines. On the other hand, nano-indentation systems [4] are suitable for local characterization of thin films on substrates, but not for free-standing micro device characterization. Therefore, there is a great interest in developing new techniques that allow for testing in the gap between the capabilities of the above mentioned well established instruments, between macro and nano scales.

Recently, many efforts in new testing methods were made by different research groups to explore the mechanical properties on free standing thin films [5–17]. These methods typically impose a fixed displacement by means of a nano-positioning motor [10–16] or an on-chip actuator [17] and measure the load. Displacement-control experiments are suitable for large specimens and/or for ductile materials, which exhibit relatively high fracture toughness. However, the fragile nature and nonlinear behavior of many component materials in MEMS structures raise the possibility of having a sudden unexpected load, which may cause failure under displacement control. For example, when starting the experiment and approaching the tip to the sample, there is always a sudden change in load which requires a feedback loop to prevent damage to the sample. Moreover, many thin MEMS structures present strong nonlinearity in load displacement relations. When these relations are unknown, there exists the possibility for high loading rate when apply a pre-determined displacement rate to these thin structures. Therefore, it is desirable to test fragile MEMS structures in a load-control instrument, i.e., an instrument that imposes a force and measures the displacement. Displacement-control instruments can be adapted for load control by means of a feedback loop. But, there are severe limitations on the response time and this is effective only for quasi-static tests instead of dynamic testing. Hence, there is a necessity to develop new techniques, which operate under load control for measuring mechanical properties of fragile MEMS devices.

Many of the MEMS structures can be studied only by dynamic load controlled experiments. These include micro-actuators, which are evaluated by the load they can overcome, and active materials such as shape memory alloys, electrostrictive, and magnetostrictive materials, in which the load (i.e., the stress and not the strain) determines the driving force for microstructural changes such as domain switching [18,19]. These advanced functional devices require a combination of a controlled and constant mechanical load and a dynamic electric/magnetic/thermal loading. Such experiments are currently beyond the scope of existing techniques.

This paper presents a new technique for measuring the mechanical response of thin free-standing MEMS structures under load control. The apparatus developed consists of inexpensive off-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received July 5, 2005; final manuscript received November 16, 2005. Review conducted by K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

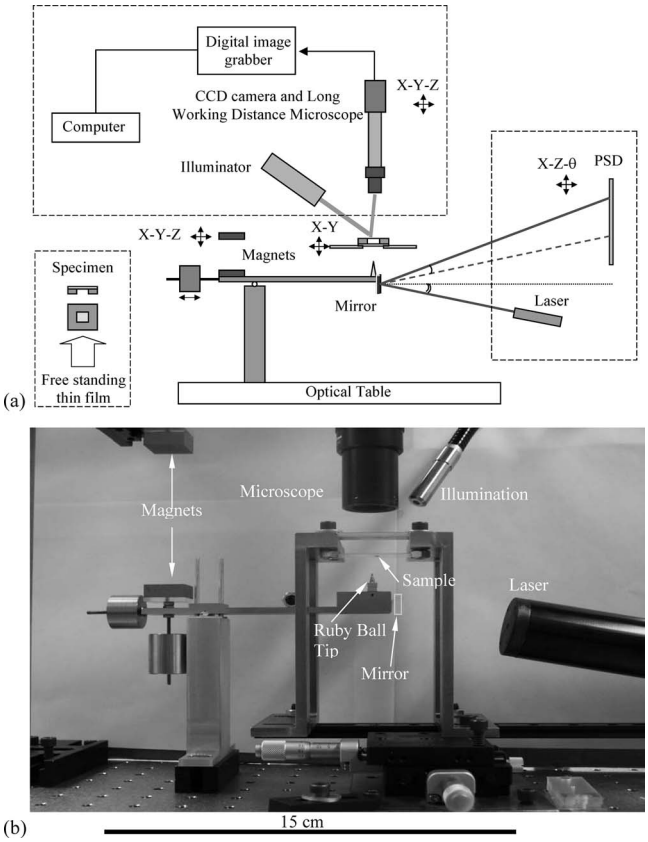


Fig. 1 (*a*) Schematic illustration of the experimental setup; (*b*) Photograph of the experimental apparatus

the-shelf products and can also be used for dynamic electro/magneto/thermo mechanical characterization experiments. The capabilities of the technique are demonstrated by testing the behavior of amorphous Si_3N_4 membrane structures, and the results demonstrate an ability to avoid local or transient stress concentration during the entire experiment.

2 Experimental Setup

The requirement of the loading and displacement range, as well as the fragility of the sample increase the difficulties in designing mechanical characterization methods for MEMS devices. The traditional methods for displacement measurement, such as the strain gauge and LVDT (linear variable displacement transducer), require partial or entire sensor in contact with the surface whose displacement needs to be measured. These methods are limited in use on the fragile sub-micron thick free-standing thin films and membranes. To overcome these difficulties, a noncontacting method using a laser beam and a position sensitive detector (PSD) are used.

Designing an appropriate loading method for testing thin films is also challenging. In principle, load control can be obtained if the spring constant of the apparatus, i.e., the mechanical structure that transmits the load from the actuator to the specimen, is significantly smaller than that of the specimen. In this situation, the applied force, F, is directly determined by the displacement imposed by the actuator, S, via $F \cong K_0 S$ where K_0 is the effective spring constant of the apparatus. However, this condition requires a very small value of K_0, which results in a low resonant frequency of the apparatus and limits the application of dynamic experiments. This problem is overcome by applying the load via a magnetostatic interaction in which the response time is much faster.

A schematic of the experimental setup is shown in Fig. 1(*a*) and a photograph of the principal apparatus is shown in Fig. 1(*b*). The load is applied by a tip, which is attached to one end of a substantially rigid beam. The force applied to the sample is proportional to the magnetostatic force being applied to the other end of the rigid beam. The late force is controlled by changing the distance between a pair of permanent magnets; one (lower) is attached to the beam and the other (upper) is attached to an external post and is able to move vertically. The upper magnet is aligned in the horizontal plane, using a *X-Y* micro-stage, such that it is situated exactly above the lower magnet, and hence no lateral forces are applied on the rigid beam.

A wide variety of methods have been suggested in the literature for measuring specimen displacements (or strains) in micromechanical testing apparatuses. These include imaging techniques [10–13], interference patterns [14,15], diffraction spots [16], and capacitance measurements [17]. These methods are limited by low sampling rates, or not having high enough resolution for large deformation, or the special complicated manufacture and sensing process of the censor. Some of these methods need special treatment (e.g., reflective coating) of the samples, which can alter the sample structure. In order to enable dynamic experiments, the loading tip displacement, which is the same as the displacement of the sample at the contact point, is measured by monitoring the deflection of a single laser beam, which is reflected from a mirror attached to the end of the rigid beam. The reflected laser beam is sensed by a Position Sensitive Detector (PSD) (Hamamatsu Co., S3979), which is attached to an external post, and the resulting voltage signal is recorded using an oscilloscope (LDS Nicolet Technologies, Model 40) (see Fig. 1). The PSD circuit allows sampling rates of up to 100 kHz and has a typical precision of about 1 μm. Due to a geometric amplification (the distance between the mirror and the PSD is typically 10 times larger than the distance between the tip and the pivot), a typical resolution of 0.1 μm can be easily obtained in measuring the tip displacement. A *Y-Z-θ* stage was designed to mount the PSD and is used in the alignment of the sensor.

The sample is mounted on an *X-Y-Z* micro-stage and its alignment with respect to the tip is monitored by a long working distance optical microscope equipped with a CCD camera connected to a computer with digital frame grabbing hardware (EPIX Inc., PIXCI SV4) and software (EPIX Inc., XCLIBV2.2-DWT-U and XCAPLITE-WIN-V2.2). A wide variety of tip shapes and dimensions are commercially available, which can provide a variety of loading conditions, including point load and line load conditions.

Note that all the controlling and monitoring components, i.e., the upper magnet, the laser, the PSD, and the microscope, are isolated from the rigid beam. This fact significantly reduces the vibration and noise and protects the very fragile structures that are being tested.

The magnetostatic force is actually determined by the difference $\Delta d = d_0 - d$, where d_0 is the initial distance between the magnets at the point where the tip first comes into contact with the sample and d is the distance between the magnets at some arbitrary moment during the experiment (see illustration in Fig. 2). The working conditions are chosen in which $d_0 = 100$ mm and Δd varies up to 13 mm. Under these conditions the sample (or load cell) displacements at the μm scale are negligible in comparison to Δd, and hence $\Delta d \cong \Delta z$, where $\Delta z = z_0 - z$ relates to the absolute readings of the upper magnet position (see Fig. 2). Thus, the load is determined directly by the upper magnet position and load-control conditions are achieved. Moreover, since $d_0 \gg \Delta z$, there is nearly a linear relation between F and Δz, as is demonstrated by the calibration curve presented in Fig. 3, which was measured by placing a load cell (Omega Engineering Inc., LCFA-50g) instead of the sample. It can be deduced from the curve in Fig. 3, a 1 μm change of Δz results in load change of 1.5 μN. Thus, load sensitivity in the μN scale can be easily obtained by moving the upper magnet with a micro-positioning device. Note that different cali-

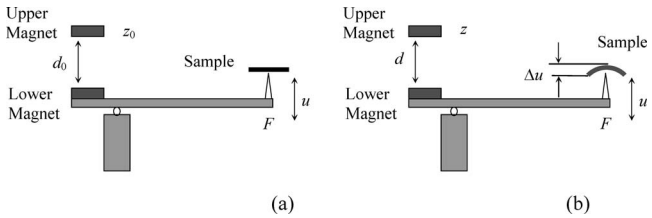


Fig. 2 Schematics of the loading system, (*a*) at the point where the tip first comes into contact with the sample; (*b*) at some arbitrary moment during the experiment. d_0 is the initial distance between the upper and lower magnets at the moment as shown in (*a*); *d* is the distance between the magnets at some arbitrary loading; z_0 is the initial reading of the upper magnets position; *z* is the absolute position at some arbitrary loading moment.

bration curves, which provide different load range and different sensitivity, can be obtained by choosing different values for d_0 or by placing different magnets.

In principle, the sample should be placed at the same height as the load cell in order to obtain the same d_0 value. Nevertheless, changes of d_0 in the range of ±1 mm result only in subtle changes of the calibration curve and hence small changes of the sample height will not have a significant effect on the calibration curve. On the other hand, an accurate determination of z_0, i.e., a clear identification of the point where the tip first comes into contact with the sample, is essential in order to accurately determines Δz. Figure 4 demonstrates that z_0 can be determined with an accuracy which is equivalent to the resolution of the micro-positioning device that is used to move the upper magnet. In this figure, two distinct regions are clearly observed in the plot of the tip displacement, u, as a function of z. During the initial movement of the magnet, the tip does not touch the sample and hence its displacement is proportional to the displacement of the upper magnet and the slope is relatively large. At the point where the tip contacts the sample, the slope changes abruptly, as the tip displacement is now equal to the sample displacement and is in the μm-scale. The obvious change in slope makes it very easy to identify the contact point. The resolution of determining the contact point has the same resolution of the z measurement, i.e., the resolution is on the order of 0.1 μm. As a result, the error in contact force is on the order of sub-μN.

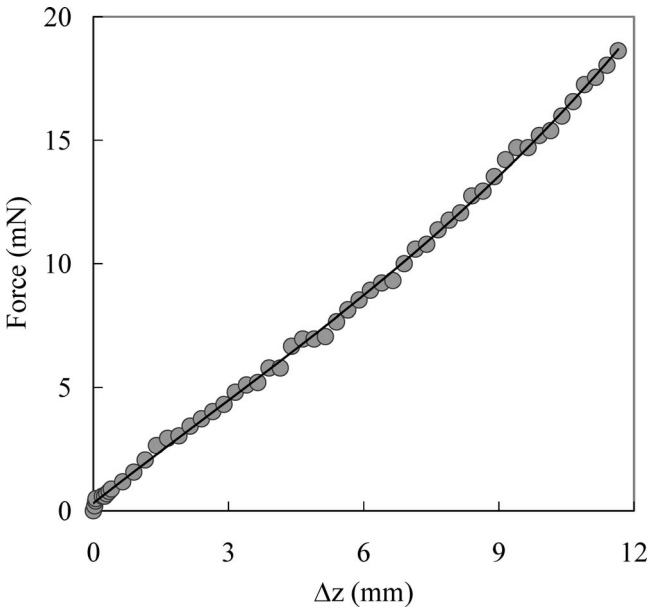


Fig. 3 The applied force, *F* in mN, as a function of the change in the position (displacement) of the upper magnet, Δz in mm for the identical z_0 as in the experiments. This response curve was obtained by applying force on a load cell.

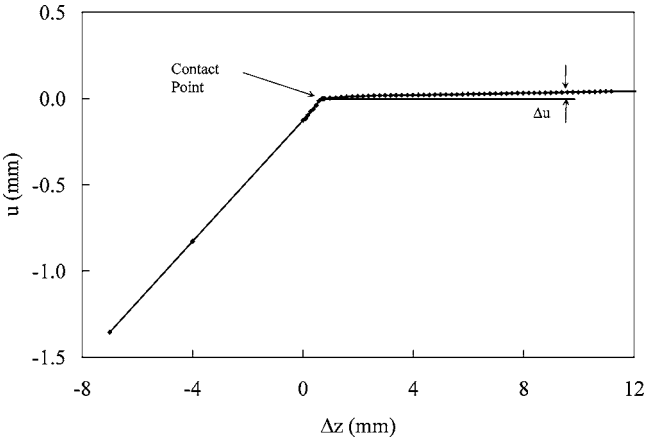


Fig. 4 The tip displacement, *u*, as a function of the position of the upper magnet, *z*. The slope of the curve changes abruptly at the point where the tip first comes into contact with the sample. The maximum value of Δu is less than 50 μm, for a change in position of the upper magnet which is on the order of 10 mm, 3 orders of magnitude larger than Δu. As a consequence, the load is completely determined by *z* and the setup is working in load-control mode.

3 Results

The results presented in this paper were obtained using a ruby ball tip, having a radius, $R=150$ μm. Large tip radius is desired in testing square and rectangular membranes in order to limit the indentation stress and prevent membrane rupture. As shown below, although the tip radius is a significant fraction of the membrane span, the radius of the contact region, r_c, is much smaller than the span, a. This fact significantly simplifies the mechanical analysis.

Figure 5 shows the applied force as a function of the measured tip displacement during loading and unloading processes. The two curves coincide with each other with no significant hysteresis, which means that the membrane behaves elastically. The largest

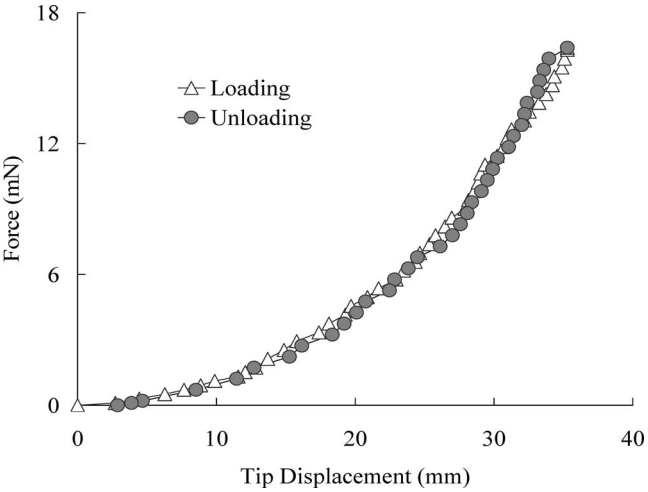


Fig. 5 The mechanical response of a 75 nm thick free-standing amorphous silicon nitride (Si_3N_4) film during loading and unloading

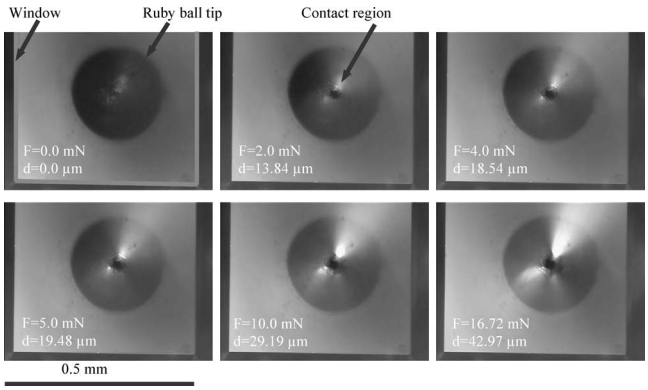


Fig. 6 A series of optical images of a Si_3N_4 membrane under different loads (corresponding load (*F*) and displacement (*d*) are indicated). These images were captured by a CCD camera mounted on the microscope above the membrane. In each image, the gray square is the Si_3N_4 membrane (window); the circular region is the ruby ball tip; the small dark region in the center which expands when loading increases is the contact area.

tip displacement is 37.5 μm, which is about 500 times the film thickness, t. This means that the bending stresses are negligible in comparison to the stresses due to stretching, i.e., to a very good approximation the thin film behaves as a membrane. Figure 6 shows a series of optical images of the membrane, which were taken under different loads. The deflected membrane forms a tent-shape where the angle of deflection, β, increases as the load increases. The repeatability of the new technique is demonstrated in Fig. 7, which presents the load-displacement curves of five membranes located in different regions of the same wafer (see Fig. 7, insert). All the curves are almost coincident with each other.

A geometrical calculation under the assumption of membrane theory shows that the radius of the contact region is given by $r_c = R \sin\beta$ (see Fig. 8). Therefore, for reasonable values of $\beta(0-15$ deg), $r_c \ll R$ and hence $r_c \ll a$, where a is again, the span of the film. Under these conditions the deflection angle can be computed, $\tan\beta = 2u/a$. Thus, for the largest tip displacement of $u = 37.5$ μm, β and r_c take values of 10.2 deg and 26 μm, respectively.

For the membrane theory approximation, the stresses at the

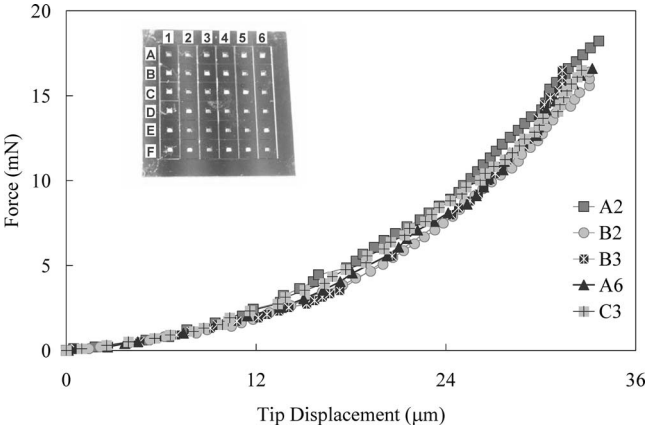


Fig. 7 A wafer consisting of 6×6 devices (insert) was tested at various locations. The mechanical responses of five membranes located at different regions of the same wafer are highly repeatable.

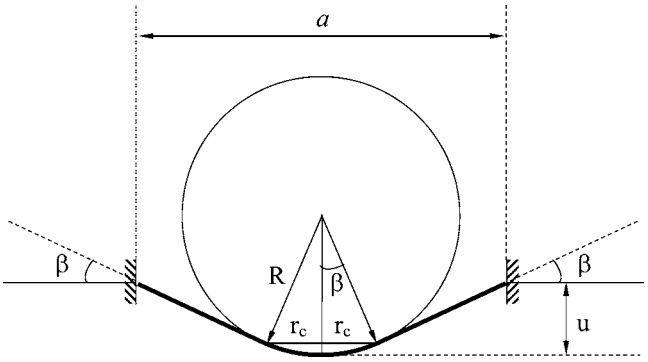


Fig. 8 Schematic diagram of a ridged spherical loading tip deforming a free-standing membrane. The relevant geometric variables are labeled.

contact region are uniform and the balance of the total forces along the z-direction yields the following expression for the membrane (stretching) stress,

$$\sigma = \frac{F}{2\pi t r_c \sin\beta} \cong \frac{Fa^2}{8\pi t R u^2} \tag{1}$$

where t is the membrane thickness, i.e., 75 nm, that of the thin film. The expression on the right hand side in (1) was obtained by making the small angle approximation, i.e., $\sin\beta \cong 2u/a$. A substitution of the largest tip displacement, $u = 37.5$ μm, at the highest force, $F = 16.6$ mN, results in very large membrane stress, $\sigma = 9.2$ GPa.

The attainment of such high stresses and large deflections reflect the high quality of the Si_3N_4 membranes. This also attests to the capabilities of the developed technique to avoid local or transient stress concentrations during the entire approach, loading, and unloading phases of the experiment. It illustrates the capability of the new technique for studying highly fragile and micro devices and structures.

4 Data Analysis and Results

4.1 Finite Element Analysis. The complexity of modeling the mechanical problem associated with the technique developed here is due to the contact loading process and large deformation associated with this problem. As a result, the loading boundary evolves with unknown loading distribution, and the governing equations are strongly coupled with severe nonlinearity even for a simple structure such as a membrane, for which the effect of bending is neglected [20]. Begley [21] recently investigated the axially symmetric case of contact problem and obtained a closed form solution while considering the effect of the indenter size. For the two-dimensional (2D) membrane under pressure loading, the solution to the resulting strongly coupled second-order partial differential equations were obtained numerically [20]. Yet, finding analytical solutions for the combination of these two cases remains as an open problem. In the present investigation, the finite element method is used to simulate tthe mechanical problem of the experiment, namely a spherical indenter loading a thin membrane. The complete analysis consists of two steps. The first step, i.e., the forward analysis, is to simulate an F (force)-d (displacement) curve by assigning a set of known material properties. Following this analysis, the shape factors which describe the response (F-d) curve are obtained by curve fitting and dimensional analysis. During the second step, the reverse analysis, material properties of the membrane are obtained by best fitting the experimental curve.

The simulations were performed using a commercial code, ABAQUS Standard Version 6.4.1 [22]. The geometry of the model has two parts: The indenter and the film. The indenter is simulated

by a rigid sphere with 300 μm in diameter and the free standing film has square geometry of the 470×470 μm with the thickness of 0.1 μm.

The thin film is discretized using 6 or 8 noded 3D continuum elements (C3D6 and C3D8), which can model large deformation problems and provide high accuracy in problems involving contact. There are 14,007 nodes and 7338 elements in the thin film and the region over which the mesh is refined covers the center area of the thin film with 100 μm in diameter. Within this region, the largest element is about 2.5×2.5 μm, and the smallest element size is 0.3×0.3 μm.

The thin film was clamped on all four sides to simulate the boundary condition of being bonded to a rigid substrate. The indenter was confined such that only translation in the z direction (normal to the thin film) was allowed. No additional symmetry conditions were imposed because a complete thin film and indenter geometry was used instead of part of it.

The loading is applied by specifying the axial displacement of the indenter. In the contact pair, the top surface of the film is defined as the slave surface and the spherical surface of the indenter as the master surface. The contact between the thin film and the spherical indenter is assumed to be governed by Coulomb friction law. Because both the materials are ceramics, which tend to have large adhesion, the high stress inside the film indicate that the contact pair could have very strong interaction. For these reasons, rough surface interaction is assumed and the friction coefficient for the contact was set to be 0.8. The effect of the friction coefficient on the results will be discussed later.

The initial material properties for the thin film amorphous Si_3N_4 are chosen to approximate the properties close to that of the bulk ceramic Si_3N_4. Accordingly, the elastic modulus and the Poisson's ratio are chosen to be 300 GPa and 0.22, respectively [23]. The linear elastic material model (elastic constitutive model of ABAQUS) with large deformation capability is chosen based on the experimental observation where there is no hysteresis in the loading-unloading process as indicated by the load-displacement curves (Fig. 5) and the film appeared intact without any cracks after loading when examined using a scanning electron microscope (SEM). The equi-biaxial residual stresses of free-standing thin films are significant parameters which depend largely on the sample fabrication method. The samples used in current investigation were fabricated by the low pressure chemical vapor deposition (LPCVD) Si_3N_4 on Si substrate and then etched away part of the substrate to make the film free-standing. In the analysis, the residual stresses are specified to be 370 MPa in the simulations corresponding to the nominal value often found in the literature for this type of thin film [23].

The indentation displacement is increased monotonically to a maximum displacement of 50 μm in 200 steps. This maximum displacement is larger than the measured experimental displacement. The geometrically nonlinear solver for large deformation in ABAQUS is employed for solving the boundary value problem. At the end of each step, the stress component in the z-direction (vertical) of the indenter is integrated in the contact area to obtain the total indentation force. Thus, the F-d curve is obtained.

4.2 Mechanical Properties Analysis. The first step in extracting the mechanical properties is to obtain the shape factors describing the load-displacement response by curve fitting of the FEA result. The F-d curve is shown in Fig. 9 for the geometrical and material parameters specified in the previous section. It is critical to choose the right form of the relation between the indentation force (F) and the center displacement (d). The solution to corresponding one-dimensional problem [21,24] suggests that two terms should be included in the F-d relation: A linear term and a cubic term with respect to d. The linear term dominates the small deflection regime and is strongly affected by the residual stress. The cubic term that dominates the large deflection regime is strongly influenced by the modulus of the material. The research

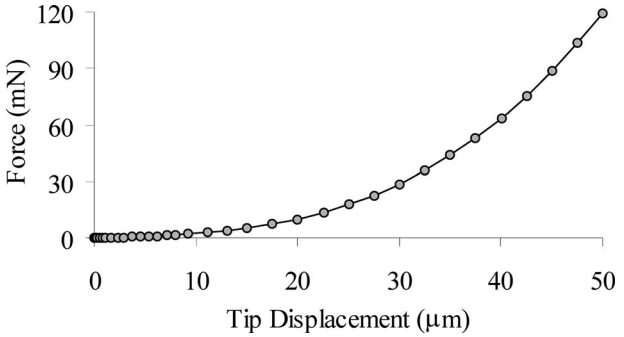


Fig. 9 Load-deflection (*F-d*) curve obtained from the finite element simulation and is used to obtain the shape factors in Eq. (2). The solid curve is the fit and the solid dots are the finite element analysis data.

on a square membrane under similar contact conditions still remains an open problem. However, it is reasonable to assume that the solution to the square membrane is similar to the circular membrane, except for coefficients (shape factors) to describe the difference in the response (F-d) curves.

Based on the arguments outlined above, the F-d relation is assumed to have the following form,

$$F = C_1 at\left[\sigma_0\left(\frac{d}{a}\right) + C_2\frac{Et^3}{(1-\nu^2)a^3}\left(\frac{d}{a}\right)^3\right] \tag{2}$$

where C_1 and C_2 are dimensionless coefficients.

Equation (2) can be rewritten in the following form,

$$F = A_1 d + A_3 d^3 \tag{3}$$

where A_1 and A_3 are coefficients corresponding to the linear term and cubic term, respectively. The least squares fitting of the F-d curve in Fig. 9 results in the coefficients, $A_1=0.134$ mN/μm and $A_3=9.047\times10^{-4}$ mN/μm^3. Then, for the assumed material properties in the finite element simulations, $E=300$ GPa, $\nu=0.22$ and $\sigma_0=370$ MPa, the shape factors C_1 and C_2 are evaluated to be 3.62 and 1.82×10^{11}, respectively. These shape factors will be used to determine Young's modulus and residual stress in the next step, i.e., the inverse analysis.

Figure 10 shows the least square fitting of the experimental F-d curve. By fitting this curve using Eq. (3), the fitting coefficients, A_1^e and A_3^e, are found to be 0.123 mN/μm and 2.33×10^{-4} mN/μm^3, respectively. The superscript e in the fitting coefficients is used to denote that they correspond to the experi-

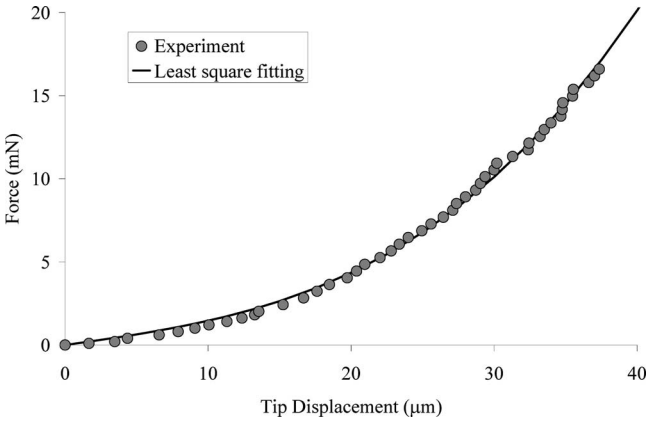


Fig. 10 Least square fitting of force-displacement (*F-d*) curve of the experimental data for extracting material properties using Eqs. (4) and (5)

Table 1 Comparison of properties for silicon nitride film

Reference	Test Method	Young's Modulus E (GPa)	Residual Stress σ_0 (MPa)	Difference in E (%)
Current work	Contact loading	243	417	0
Edwards et al. [23]	Micro tension	255±5	NA	4.7
Zhang et al. [25]	Pressure buldge	255.3	372	4.7

mentally determined values. Then, the initial stress σ_0 and the elastic modulus E of the film can be obtained as $\sigma_0=417.3$ MPa and $E=243.2$ GPa, by using the following equations:

$$\sigma_0 = \frac{A_1^e}{C_1 t} \tag{4}$$

$$E = \frac{A_3^e(1-\nu^2)a^5}{C_1 C_2 t^4} \tag{5}$$

where C_1 and C_2 are obtained in the first step of the analysis, and equal 3.62 and 1.82×10^{11}, respectively. The values of Young's modulus and residual stress agree well with the characterization results of the same sample by a different technique, the pressure bulge test [25]. The value of the Young's modulus of thin film Si_3N_4 found here is found to be in good agreement with those found in literature [23] (see Table 1).

4.3 Uncertainty Analysis. For a given *F-d* curve, the solution of E and σ_0 are unique, i.e., there are unique values of fitted E and σ_0. The accuracy of the above analysis directly depends on the accurate measurement of the *F-d* curve. The errors in the tests come from the displacement and the force measurements. Typical error from displacement measurements is about 0.1 μm over the displacement range of 30 μm, which is less than 0.5%. Errors in the force measurements include a 5% error from the load cell and a less-than 0.1% error from the micrometer. Therefore, the total error in the analysis is about 5.6%.

During the first step of the analysis where the two shape factors C_1 and C_2 are determined, the sensitivity of C_1 and C_2 to the variation of input material properties was investigated. The results showed that a ±5% change in either the Poisson's ratio (ν) or the residual stress (σ_0) would lead to variations of less than ±5% of the shape factors values. When ν reaches 0.25, the maximum value in the reasonable range for Si_3N_4 [23], i.e., a 13% increase value, the result only shows a 3% change in C_1 and C_2. However, these two shape factors are more sensitive to variation in the Young's modulus, E. A 5% change in E will lead a 9% change in C_2 and less than 7% change in C_1. The effect of friction coefficient is also analyzed. To change its value from 0.8 to 1.0 will only lead a 3% change in both C_1 and C_2. Moreover, for a frictionless condition, where this coefficient is zero, it will result in less than 7% change in C_1 and 9% change in C_2.

During the second step of the analysis of determining the material properties, the sensitivity of the predicted material properties to variations in the shape factors obtained from the first step analysis is studied as well. The results showed that a ±5% change of C_1 will cause $\mp$5% change in values of E and σ_0, while a ±5% change in C_2 will introduce $\mp$5% change in E and leave σ_0 unchanged.

In order to verify the method used in current work to obtain material properties, the predicted material properties were used as input in FEM analysis. The calculated *F-d* curve agrees well with the experimental data. Both the first and second step of the analysis gave unique results of parameters, (C_1, C_2), and (E, σ_0), respectively.

Moreover, the effect of the residual stress on the mechanical behavior of the overall structure cannot be ignored. From (2), one can see that the pre-stress makes the film stiffer, i.e., for the loading with the same magnitude, the pre-stressed (tensile) film will have less center deflection. This feature can be used in MEMS devices to improve the performance. The pre-stress can be as much as several hundred MPa, and this can cause a difference in the center deflection as much as 20%, which could affect the performance of the MEMS devices significantly.

5 Conclusions

A new technique for the mechanical characterization of released thin films under indentation load has been developed. This technique can be used to apply load in the μN-mN range by either load or displacement control. The load and displacement can be measured to an accuracy of within 0.1 mN and 0.1 μm, respectively. The capability and reliability of this new technique has been demonstrated by studying Si_3N_4 free-standing membranes. The elastic modulus and residual stress of Si_3N_4 free standing thin film are around 250 GPa and 400 MPa, respectively. These values are in close agreement with values obtained using a different technique and as well as those found in the literature. The experimental setup has the capability of testing bridge shape samples by using wedge-shaped loading tips instead of spherical (ball) tips. It also has the capability for performing dynamic measurements by replacing the permanent magnets in the current loading apparatus with computer controlled electro-magnets.

Acknowledgment

The research support provided by the Army Research Office through a MURI grant (No. DAAD19-01-1-0517) to the California Institute of Technology on *Engineered Complexity in Ferroelectric Devices* is gratefully acknowledged.

References

[1] Spearing, S. M., 2000, "Materials Issues in Microelectromechanical Systems (MEMS)," Acta Mater., **48**, pp. 179–196.
[2] Sharpe, W. N., 2001, Mechanical Properties of MEMS Materials, in *The MEMS Handbook*, M. Gad-el-Hak, ed., CRC Press, Boca Raton, FL.
[3] Senturia, S. D., 2001, *Microsystem design*, Kluwer Publishers, Boston.
[4] 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**(6), pp. 1564–1583.
[5] Srikar, V. T., and Spearing, S. M., 2003, "A Critical Review of Microscale Mechanical Testing Methods Used in the Design of Microelectromechanical Systems," Exp. Mech., **43**, pp. 238–247.
[6] Haque, M. A., and Saif, M. T. A., 2003, "A Review of MEMS-Based Microscale and Nanoscale Tensile and Bending Testing," Exp. Mech., **43**(3), pp. 248–255.
[7] Kraft, O., and Volkert, C. A., 2001, "Mechanical Testing of Thin Films and Small Structures," Adv. Eng. Mater., **3**, pp. 99–110.
[8] Zhang, T. Y., Su, Y. J., Qian, C. F., Zhao, M. H., and Chen, L. Q., 2000, "Microbridge Testing of Silicon Nitride Thin Films Deposited on Silicon Wafer," Acta Mater., **48**, pp. 2843–2857.
[9] Vlassak, J. J., and Nix, W. D., 1992, "A New Bulge Test Technique for the Determination of Young's Modulus and Poisson's Ratio of Thin Films," J. Mater. Res., **7**(12), pp. 3242–3249.
[10] Read, D. T., and Dally, J. W., 1993, "A New Method for Measuring the Strength and Ductility of Thin Films," J. Mater. Res., **8**(7), pp. 1542–1549.
[11] Chasiotis, I., and Knauss, W., 2002, "A New Microtensile Tester for the Study of MEMS Materials with the Aid of Atomic Force Microscopy," Exp. Mech., **42**(1), pp. 51–57.
[12] Haque, M. A., and Saif, M. T. A., 2002, "In situ Tensile Testing of Nano-scale Specimens in SEM and TEM," Exp. Mech., **42**, pp. 123–128.
[13] Emery, R. D., and Povirk, G. L., 2003, "Tensile Behavior of Free-standing Gold Films. Part I. Coarse-Grained Films," Acta Mater., **51**, pp. 2067–2078; Acta Mater., **51**, pp. 2079–2087.
[14] Espinosa, H. D., Prorok, B. C., and Fischer, M., 2003, "A Methodology for Determining Mechanical Properties of Freestanding Thin Films and MEMS Materials," J. Mech. Phys. Solids, **51**, pp. 47–67.
[15] Sharpe, W. N., and Gianola, D. S., 2004, "Techniques for Testing Thin Films in Tension," Exp. Tech., **28**, pp. 23–27.
[16] Ruud, J. A., Josell, D., Spaepen, F., and Greer, A. L., 1993, "A New Method for Tensile Testing of Thin Films," J. Mater. Res., **8**(1), pp. 112–117.
[17] Zhu, Y., Moldova, N., and Espinosa, H. D., 2005, "A Microelectromechanical

Load Sensor for In Situ Electron and X-ray Microscopy Tensile Testing on Nanostructures," Appl. Phys. Lett., **86**, 013506.
[18] Shu, Y. C., and Bhattacharya, K., "Domain Patterns and Macroscopic Behavior of Ferroelectric Materials," Philos. Mag. B, **81**(12), (2001).
[19] Burcsu, E., Ravichandran, G., and Bhattacharya, K., "Large Electrostrictive Actuation of Barium Titanate Single Crystals," J. Mech. Phys. Solids, **52**, 823–846 (2004).
[20] Katsikadelis, J. T., Nerantzaki, M. S., and Tsiatas, C. C., 2001, "The Analog Equation Method for Large Deflection Analysis of Membranes. A Boundary-only Solution," Comput. Mech., **27**, pp. 513–523.
[21] Begley, M. R., and Mackin, T. J., 2004, "Spherical Indentation of Freestanding Circular Thin Films in the Membrane Regime," J. Mech. Phys. Solids, **52**, pp. 2005–2023.
[22] ABAQUS, www.abaqus.com
[23] Edwards, R. L., and Sharpe, W. N., 2004, "Comparison of Tensile and Bulge Tests for Thin-film Silicon Nitride," Exp. Mech., **44**, pp. 49–54.
[24] Zhang, Y., and Wise, K. D., 1994, "Performance of Non-Planar Silicon Diaphragms Under Large Deflections," J. Microelectromech. Syst., **3**(2), pp. 59–68.
[25] Zhang, R., Shilo, D., De Basombes, J. T., Bhattcharya, K., and Ravichandran, G., 2005, "Mechanical Characterization of Multi-Layer Free Standing Thin Films Using Pressure Bulge Test," Exp. Mech., in preparation.

Material Characterization and Modeling of Single-Wall Carbon Nanotube/Polyelectrolyte Multilayer Nanocomposites

Gang Huang

Bo Wang

Hongbing Lu[1]
e-mail: hongbin@ceat.okstate.edu

School of Mechanical and Aerospace Engineering,
Oklahoma State University,
Stillwater, OK 74078

Arif Mamedov

Sachin Gupta

Nomadics,
1024 S. Innovation Way,
Stillwater, OK 74074

*Strong single-wall carbon nanotubes (SWNTs) possess very high stiffness and strength. They have potential for use to tailor the material design to reach desired mechanical properties through SWNT nanocomposites. Layer-by-layer (LBL) assembly technique is an effective method to fabricate SWNT/polyelectrolyte nanocomposite films. To determine the relationship between the constituents of the SWNT/polymer nanocomposites made by LBL technique, a method has been developed to extend the recent work by Liu and Chen (Mech. Mater., **35**, pp. 69–81, 2003) for the calculation of the effective Young's modulus. The work by Liu and Chen on the mixture model is evaluated by finite element analysis of nanocomposites with SWNT volume fraction between 0% and 5%. An equivalent length coefficient is introduced and determined from finite element analysis. A formula is presented using this coefficient to determine the effective Young's modulus. It is identified that the current work can be applied to SWNT loadings between 0% and 5%, while Liu and Chen's approach is appropriate for relatively high SWNT volume fractions, close to 5%, but is not appropriate for relatively low SWNT volume fractions. The results obtained from this method are used to determine the effective Young's modulus of SWNT/polyelectrolyte nanocomposite with 4.7% SWNT loading. The material properties are characterized using both nanoindentation and tensile tests. Nanoindentation results indicate that both the in-plane relaxation modulus and the through-thickness relaxation modulus of SWNT nanocomposites are very close to each other, despite the orientation preference of the SWNTs in the nanocomposites. The steady state in-plane Young's relaxation modulus compares well with the tensile modulus, and measurement results are compared with Young's modulus determined from the method presented.*
[DOI: 10.1115/1.2206196]

1 Introduction

Single-wall carbon nanotubes (SWNTs) possess exceptional stiffness (Young's modulus ~1000 GPa [1]) and tensile strength (tensile strength ~50 GPa [2]). Since the 1990s, they have been used as reinforcing constituents to enhance the mechanical properties of polymer matrix [3–7]. SWNTs have the potential for use to tailor the design of nanostructured materials to reach the desired mechanical properties. SWNT dispersion plays a critical role in enhancing both material stiffness and tensile strength. In recent years, the method of layer-by-layer (LBL) assembly of SWNT nanocomposites has been successfully used to disperse SWNT uniformly over a very wide range of loadings in polymer matrix. The weight fraction of SWNTs in nanocomposites has reached as high as 50% [8,9], leading to an enhancement in the Young's modulus of 100 times or more and an improvement in tensile strength of ten times or more over the neat polymer resin. Understanding the structure-property relation is essential in material design and optimization. One of the critical issues is to evaluate the effective material properties such as stiffness, strength, and toughness for new nanocomposites based on their constituents.

Numerous approaches have been used to determine the structure-property relations. These include, but are not limited to, molecular dynamics (MD) simulations [10], equivalent-continuum modeling technique [11–13], micromechanics approach [14–23], and continuum mechanics approach using representative volume elements [24,25]. MD simulations can accurately determine the chemical and physical properties of nanostructures at atomistic scale with the use of appropriate interatomic potentials. However, it is limited by the attainable time and length scales [10], preventing it from predicting accurately the behavior of the overall bulk SWNT nanocomposites. Micromechanical approaches have been used to determine the effective material properties for SWNT nanocomposites [14–18]. Harik [19,20] used the continuum beam model for carbon nanotubes and nanorods to develop mechanics of carbon nanotubes. The Mori-Tanaka theory [21], which evaluates the average internal stress in the matrix of a material containing precipitates such as SWNTs, has been utilized extensively for the evaluation of the effective properties of composites. Fisher, Bradshaw, and Brinson [22,23] investigated the elastic modulus of a polymer reinforced with wavy carbon nanotubes (CNTs) using the Mori-Tanaka method, and found that the waviness in CNTs leads to much less effective stiffening effects in a nanocomposite reinforced by CNTs rather than straight CNTs.

Recently Liu and Chen [24,25] presented a method to use a rectangular representative volume element (RVE) to determine the Young's modulus and Poisson's ratio based on the strength of materials analysis. In their model, they used two matrix elements at both ends to connect CNT embedded matrix elements in the middle in serial connection to develop a formula for the effective Young's modulus of CNT nanocomposites. Their formula, as will

[1]Author to whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received December 20, 2004; final manuscript received April 24, 2006. Review conducted by K. M. Liechti. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California – Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

be discussed further in Sec. 2, has been found to be effective at relatively high CNT volume fraction (5% or higher).

In this investigation, we extend the method presented by Liu and Chen [24,25] by introducing an equivalent length coefficient to consider the three-dimensional end effects of SWNTs in matrix to develop a modified mixture model for the calculation of the Young's modulus. The equivalent length coefficient will be determined from finite element analysis of nanocomposites with different SWNT loadings between 0% and 5%.

A concurrent effort is to manufacture SWNT/polyelectrolyte multilayer nanocomposites. Identical processing conditions are used to fabricate LBL multilayer polyelectrolyte films and multilayer SWNT/polyelectrolyte nanocomposite films with SWNT loading of 4.7%. Both tensile and nanoindentation measurements are used to characterize the mechanical properties of films. In nanoindentation, both the in-plane properties and the through-thickness properties are determined using viscoelastic contact mechanics analysis; and the steady state values are compared with tensile data. The in-plane Young's modulus is compared with the calculated results to examine the modified mixture model.

2 Mixture Model for Effective Young's Modulus

We extend the serial model presented by Liu and Chen [24,25] to develop a modified mixture model for the effective Young's modulus of SWNT nanocomposites. In this mixture model, a representative volume element (RVE) is established through the consideration of SWNTs embedded in a rectangular prism matrix. Both the nanotubes and the matrix are considered as homogeneous, isotropic and linearly elastic materials, and they have a perfect adhesion at their interface.

Figure 1 shows a schematic diagram of the mixture model with the use of a SWNT/matrix element in the middle (Fig. 1(*a*)). In Fig. 1(*a*), a nanotube or multiple nanotubes with a length, l_0 are embedded in the center of the matrix. The model in Fig. 1(*a*) is represented by an equivalent model as shown in Fig. 1(*b*), whereas the shaded area in the middle of Fig. 1(*b*) is a homogenized area with the same modulus as determined using the strength of materials approach from Fig. 1(*c*). The equivalent condition assures that both models in Figs. 1(*a*) and 1(*b*) give the same Young's modulus in the longitudinal direction. It may be noted that the three-dimensional end effects are present in the model in Fig. 1(*a*) near both ends of the SWNT, but are absent in the model of Fig. 1(*b*) as a uniaxial stress state is assumed following the Euler-Bernoulli assumptions. To allow both models in Figs. 1(*a*) and 1(*b*) to give the same Young's modulus in the longitudinal direction, the length l_0^{eq} of the shaded area in Fig. 1(*b*) is not necessarily equal to the length of the SWNT, l_0, in Fig. 1(*a*); rather there is a relation between them. The equivalent length l_0^{eq} is calculated from

$$l_0^{eq} = \xi \cdot l_0 \tag{1}$$

where ξ is an equivalent length coefficient, and $\xi = \xi(l_0/l_1, f_{NT})$; it is determined from equating the solution of the Young's modulus from the mixture model in Fig. 1(*a*) to the model in Fig. 1(*b*). In the case where $\xi = 1$, it indicates that the end effects near SWNTs in Fig. 1(*a*) are negligible so that the model in Fig. 1(*a*) can be considered as a uniaxial model as shown in Fig. 1(*b*) with $l_0^{eq} = l_0$. The coefficient ξ will be determined from finite element analysis and analysis of the mixture model in Fig. 1(*b*) using strength of materials.

We next develop an equation for the Young's modulus of the mixture model in Fig. 1(*b*) using the strength of materials approach. Applying the plane assumption as indicated in the Euler-Bernoulli assumptions to the model in Fig. 1(*c*) leads to identical strain in the SWNT and matrix. Under this assumption, the load

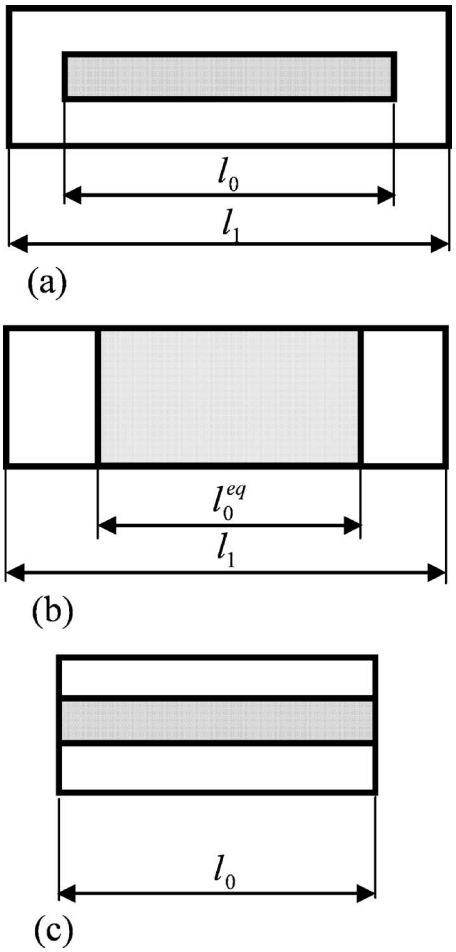


Fig. 1 Schematic of the mixture model

applied on the CNT and the matrix will be distributed on CNT and matrix based on their respective stiffness, resulting in the effective Young's modulus, E_e^0,

$$E_e^0 = E_{NT} f_{NT}^0 + E_M (1 - f_{NT}^0) \tag{2}$$

where E_{NT} and E_M are the Young's moduli for the SWNT and the matrix, respectively, f_{NT}^0 is the volume fraction of SWNT. Since the densities for both the SWNT and the matrix (polyelectrolyte) are very close to each other, the volume fraction is considered approximately the same as the weight fraction.

In Fig. 1(*a*), the volume fraction of SWNTs is f_{NT}, and in Fig. 1(*c*) the volume fraction is f_{NT}^0, their relation is

$$f_{NT} = f_{NT}^0 \cdot \frac{l_0}{l_1} \tag{3}$$

The model in Fig. 1(*c*) provides the Young's modulus for the homogeneous element (shaded area in Fig. 2(*b*)) in the middle of

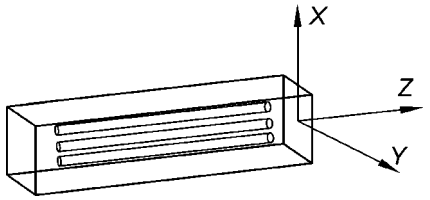


Fig. 2 First set of FEM model for matrix embedded with three SWNTs

Fig. 1(*b*). For the serial model in Fig. 1(*b*), the axial stress in the homogeneous element in the middle of Fig. 1(*b*) is the same as that in the matrix at both ends of the model. The overall deformation in the model of Fig. 1(*b*) is the summation of the deformation in the homogeneous element in the middle and the deformation in the remaining matrix. With these considerations, the effective Young's modulus of nanocomposites in the entire model of Fig. 1(*b*) is

$$E_e = \left[\frac{1}{E_M}\left(1 - \frac{l_0^{eq}}{l_1}\right) + \frac{1}{E_e^0}\frac{l_0^{eq}}{l_1} \right]^{-1} \tag{4}$$

Considering Eqs. (1)–(3), Eq. (4) can be rearranged to the following form

$$E_e = E_M + f_{NT}(E_{NT} - E_M)\left[\frac{1}{1 + \left(\frac{E_{NT}}{E_M} - 1\right) \cdot f_{NT} \cdot \left(\frac{1}{\xi} \cdot \frac{l_1}{l_0} - 1\right)} \right] \tag{5}$$

Equation (5) is the formula to evaluate the effective Young's modulus of nanocomposites. When $\xi = 1$, Eq. (5) is reduced to the formula presented by Liu and Chen [24,25]. In the case of a low volume fraction in which the condition of dilute CNT dispersion in a nanocomposite can be justified, the effective Young's modulus, given by the dilute approximation approach [13], is

$$E_e = E_M + f_{NT}(E_{NT} - E_M)A^{NT,\mathrm{dil}} \tag{6}$$

where $A^{NT,dil}$ is the dilute strain concentration factor, determined from an RVE with one inclusion.

For a two-phase composite with a very low volume fraction of identical inclusions, the effective Young's modulus can be determined using Eq. (6). With the increase in the volume fraction of inclusions, the interaction between inclusions becomes more significant, voiding the dilute inclusion approximation. An alternative approach, not used in the current study, is the application of the Morri-Tanaka model in micro-mechanics analysis for CNT reinforced nanocomposites. For the details of this approach, please refer to the work by Bradshaw et al. [23].

3 Finite Element Analysis of an RVE

In this section we will use finite element analysis to model RVEs to determine the equivalent length coefficient ξ in Eq. (5) and the effective Young's modulus of LBL SWNT/polyelectrolyte nanocomposite film. In the model SWNTs are considered as straight circular cylinders lying on the film plane.

In the mixture model as shown in Fig. 1(*b*), the Young's modulus of the homogeneous element in the middle is determined from the model in Fig. 1(*c*). The three-dimensional RVE in Fig. 1(*a*) will be investigated under tension to determine its effective Young's modulus using ABAQUS/Standard code. SWNTs are modeled as solid circular cylinders with a diameter of 1.2 nm. The length of SWNTs is modeled as 40 nm to accommodate the computational capacity (the length of actual nanotubes used in the LBL film is between 600–700 nm). Our results indicate that when the SWNT length has exceeded 40 nm, further increase in the SWNT length under the same f_{NT} does not change the effective Young's modulus, so that a length of 40 nm used in analysis for SWNT is adequate. The matrix is a rectangular prism with a square cross-sectional area A and a length l_1. A perfect bonding is assumed between the SWNTs and the matrix. Both the matrix and the SWNTs are assumed to be homogeneous, isotropic, and linearly elastic. The Young's modulus of the polyelectrolyte is 300 MPa, as measured from tensile tests that will be described in Sec. 4, and the Young's modulus of SWNTs is 1000 GPa [1]; the Poisson's ratio used for both polyelectrolyte and the SWNT is assumed as 0.3. It was found that the variation of Poisson's ratio does not significantly affect the effective Young's modulus [22].

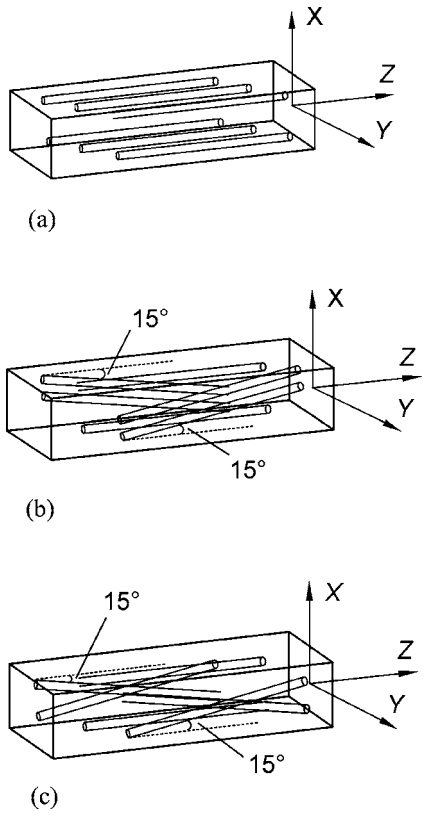


Fig. 3 Second set of FEM models

Two sets of models are used in numerical analysis for SWNT nanocomposites. In the first set, a parallel SWNT or multiple parallel SWNTs are embedded in the center of the matrix in the loading direction. The other set uses carbon nanotubes dispersed in the matrix pseudo randomly. In the first set of models, one, two, or three tubes in the Z axis (see Fig. 2 for three tubes in a model) are considered in each model. The volume fraction of a single SWNT in this model is 1.67%. The finite element mesh contains 55,560 nodes and 282,595 tetrahedral elements. Similarly, two or three SWNTs are embedded in parallel in the matrix to construct other composite models with volume fractions, 3.33% and 5%, respectively. It may be noted that the change in the cross-sectional area or the length of matrix will induce different volume fractions as well. In the second set of models, three representative three-dimensional finite element models are used to consider the case of pseudo-randomly distributed SWNTs in the matrix, as shown in Fig. 3. Figure 3(*a*) shows a matrix with six parallel SWNTs separated by different spacing in the three directions. Figure 3(*b*) shows six SWNTs parallel to the film plane (front surface of the RVE); the two parallel SWNTs in the front form a 30 deg angle with the two parallel SWNTs in the back, while the two parallel SWNTs in the middle are in the loading direction (Z axis). Figure 3(*c*) shows six SWNTs parallel to the film plane (front surface of the RVE), the two SWNTs in the front form 30 deg angle with each other, and the two SWNTs in the back also form 30 deg with each other; the two parallel SWNTs in the middle are aligned with the loading direction (Z axis).

In each tensile model of length l_1, one end is constrained in the Z axis, and the other end is subjected to an axial displacement, Δ applied at all nodes on the end surface. The summation of all nodal reaction forces is F_R. The effective Young's modulus in the RVE is then computed as

$$E_e^{\mathrm{FEM}} = \frac{F_R l_1}{A\Delta} \tag{7}$$

The results from finite element analysis are used in Eq. (7) to calculate the effective Young's modulus. We have investigated the effective Young's modulus in the longitudinal direction for nanocomposites with SWNT volume fractions from 0.1% to 5%.

4 Tensile and Nanoindentation Measurements

We conducted two independent tests, namely nanoindentation and small scale tensile tests, to find the effective mechanical properties of the neat resin and the nanocomposite materials. It is expected that nanoindentation would provide material properties in both the in-plane and through-thickness directions of the nanocomposite films, while the tensile tests would provide the in-plane properties. We describe in this section first the material preparation, and then the two types of tests.

4.1 Material Preparation. Poly(dimethyldiallylammonium chloride) (PDDA, Molecular weight Mw = 200,000 – 350,000; 20% wt. in water) and Poly(acrylic acid) (PAA, Mw = 450,000) used as organic components of LBL assembly were bought from Aldrich (Milwaukee, WI). All chemicals were used without further purification. Single-wall carbon nanotubes in the form of 15 mg/g aqueous gel material were purchased from SouthWest Nanotechnologies (Norman, OK) and were oxidized according to procedures described elsewhere [8,26,27]. The nutshell procedures were as follows: 0.7 g of gel material were refluxed in 10 mL of 2.6 M HNO_3 solution at 85 °C for 45 h; then, the solution was centrifuged at 900 rpm and the black precipitate was washed by de-ionized water (DI water) and centrifuged again. This procedure was repeated two or three times until the supernatant had *p*H=5 or more. After titration of supernatant to *p*H =8.0 with 1 M NaOH, it was shaken with the precipitate and centrifuged again. The resulting precipitate was separated and redispersed in 100 mL of DI water by bath sonication for 50 min.The resulting solution was centrifuged at 5000 rpm for 4 h to remove undispersed SWNT bundles and impurities. Supernatant solution, containing ~0.1% oxidized SWNTs, was used for layer-by-layer deposition. Layer-by-layer deposition was performed from aqueous dispersions according to procedures developed by Decher for oppositely charged polymers [28] using automated deposition machine, developed at Nomadics, Inc. A typical deposition cycle consisted of: (1) deposition of PDDA for 10 min; (2) rinsing; (3) deposition of PAA/SWNT mixture for 30 min; and (4) final rinsing. Both rinsing steps consisted of several flushes of deposition chamber for a total of 3 min. The concentrations of deposited species were 0.5% for both polymers and ~0.1% for oxidized SWNTs. Deposition was performed on both glass slides and Poly(vinyl chloride) (PVC) tubing. When desired degree of film thickness was obtained (usually 80–100 deposition cycles), composites were air dried. Free-standing films were prepared by careful removal of LBL multilayers from PVC tubing during the drying step. The LBL multilayer film contains 4.7% (weight percent) of SWNTs based on thermogravimetric analysis. A JOEL JXM 6400 scanning electron microscope (SEM) was used to observe the cross section of a 100 layer SWNT/polyelectrolyte film; the accelerating voltage and working distance used in imaging were 25 kv and 45 mm, respectively. A SEM image of the cross section of a 100 layer film is shown in Fig. 4(*a*). The top surface of a monolayer SWNT/polyelectrolyte film was observed using an atomic force microscope (AFM) (Digital Instruments, Multimode Scanning Probe Microscope with controller Nanoscope III a) under tapping model. Figure 4(*b*) shows an AFM top surface image of a single layer of SWNT/polyelectrolyte coating with 4.7% of SWNT on a glass slide. Before each test, each nanocompositefilm sample was cured at 90 °C in a vacuum oven for 2 h to reduce moisture concentration. All samples were prepared in identical

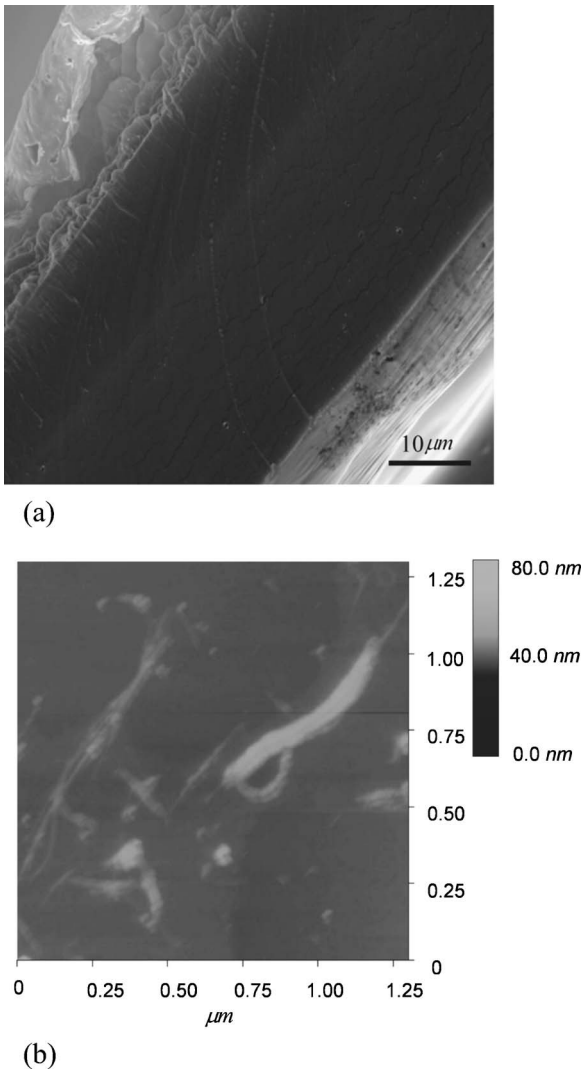


Fig. 4 SEM and AFM images (*a*) SEM image of cross section of LBL SWNT composites film, (*b*) AFM image of surface of a monolayer film with 4.7% of SWNTs

procedures. All tensile and nanoindentation tests were conducted in laboratories with a room temperature of 22 °C and a relative humidity nearly 50%.

4.2 Tensile Measurements. The in-plane properties of the SWNT/polyelectrolyte multilayer nanocomposite film were measured using small scale tensile tests. A small scale testing apparatus (United Testing Systems) was used to measure the stress-strain relation, including the Young's modulus. An American society for testing and materials (ASTM) D1708 test method was used in tensile tests. A displacement control with a velocity 0.25 μm/s was applied. Dumbbell shaped film tensile specimens were die cut from the multilayer neat resin and nanocomposite films following the ASTM D1708 standard. The nominal gauge length of the test specimens was 21.75 mm. The nominal width and thickness of the specimen were 4.75 and 0.056 mm, respectively. Actual dimensions were measured and used in the computation of stress and strain.

4.3 Nanoindentation Measurements. Nanoindentation tests were conducted using an MTS Nano Indenter XP system. The resolutions of displacement and load were 0.01 nm and 50 nN, respectively. A Berkovich indenter tip was used; it was assumed as a conical nanoindenter with a half-cone angle of 70.3 deg in the

analysis for determining the Young's relaxation modulus. For measurements of in-plane properties, a SWNT nanocomposite film was embedded into a polybed matrix, cured and then sectioned by ultramicrotome to reveal the cross section. The sample was then annealed at 90°C to reduce the effects from prior thermal/mechanical history. The nanoindentation tests were conducted after the drift rate dropped to a set value (0.05 nm/s in this study) to reduce error induced by thermal drifting. A constant rate loading history with a loading rate 0.065 mN/s was prescribed. After the indenter tip had made contact with the sample surface, the load and the displacement were recorded at a sampling rate of five data points per second.

4.4 Analytical Prerequisite in Measurements of Viscoelastic Properties Using Nanoindentation. Nanoindentation allows measurements of both the in-plane and through-thickness properties. In this section we provide formulas used for determining the linear viscoelastic properties using nanoindentation [29]. The Berkovich indenter is considered here, and is modeled as a rigid conical indenter. When a rigid conical indenter indents into a linearly viscoelastic half space, the indentation involves a time-dependent contact area between the indenter and the workpiece. For this moving boundary problem, Lee and Radok [30] proposed a hereditary integral operator to determine the time-dependent stresses and deformations in the case where the contact area between the indenter and the work material does not decrease with time. Applying this technique, the load-displacement relation for viscoelastic indentation with conical indenter is [29]

$$h^2(t) = \frac{\pi(1-\nu)\tan\alpha}{4}\int_0^t J(t-\xi)\left[\frac{dP(\xi)}{d\xi}\right]d\xi \tag{8}$$

where $P(t)$, $h(t)$ are load and displacement, respectively; α is the angle between the cone generator and the substrate plane, ν is the Poisson's ratio, and $J(t)$ is the creep compliance in shear at time t and can be expressed by the generalized Kelvin model,

$$J(t) = J_0 + \sum_{i=1}^{N} J_i(1 - e^{-t/\tau_i}) \tag{9}$$

where J_0, J_i are compliance numbers, τ_i retardation times. Under $P(t)=\nu_0 t$ with ν_0 being a constant loading rate, substituting Eq. (9) into Eq. (8) leads to

$$h^2(t) = \frac{1}{4}\pi(1-\nu)\tan\alpha\left[\left(J_0 + \sum_{i=1}^{N} J_i\right)P(t) - \sum_{i=1}^{N} J_i(\nu_0\tau_i)\left(1 - e^{-\frac{P(t)}{\nu_0\tau_i}}\right)\right] \tag{10}$$

After fitting Eq. (10) to the load-displacement curve from nanoindentation, the best-fit parameters, $J_0, J_i (i=1,\ldots,N)$ and τ_i can be obtained. The creep compliance can be subsequently determined using Eq. (9). Once $J(t)$ is obtained, other viscoelastic functions, such as the uniaxial relaxation modulus $E(t)$, can be determined. For example, the creep function in shear, $J(t)$, can be converted to $E(t)$ through the following relation

$$\int_0^t E(\tau)J(t-\tau)d\tau = 2(1+\nu)t \tag{11}$$

under condition of a constant Poisson's ratio ν.

5 Results and Discussion

We present in this section the measurement results from both tensile tests and nanoindentation tests, followed by the results on the effective Young's modulus determined using the method described in Sec. 2.

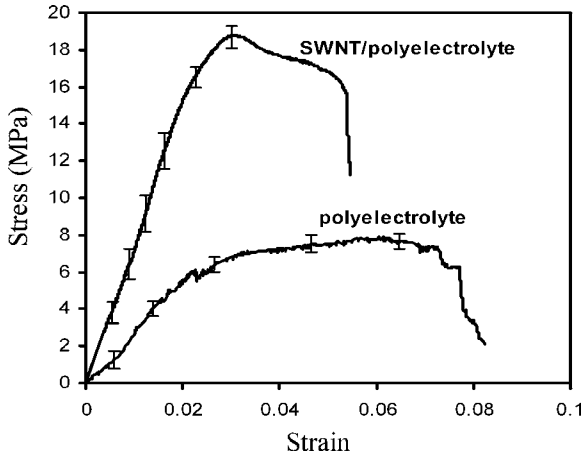


Fig. 5 Stress-strain curves from tensile testing

5.1 Measurement Results. The stress-strain curves for SWNT/polyelectrolyte multilayer composite film and the neat resin film (PAA+PDDA) are depicted in Fig. 5 from small scale tensile testing. For each type of film, three tests were conducted to obtain the average stress-strain curve; the error bars indicate the range of data obtained in tensile tests. The Young's moduli of the multilayer polyelectrolyte film and the SWNT/polyelectrolyte multilayer nanocomposite film are determined as 300 and 770 MPa, respectively. The tensile strengths for the neat polyelectrolyte and the SWNT/polyelectrolyte films are 6.6 and 18.6 MPa, respectively. With the use of 4.7% of SWNT in polyelectrolyte, the stiffness (indicated by the Young's modulus) and the tensile strength have been increased by 1.57 and 1.82 times, respectively, over the neat polyelectrolyte resin.

Nanoindentation tests were conducted to measure the Young's relaxation modulus in both in-plane direction and through-thickness direction. Figure 6 shows the nanoindentation load-displacement curves for SWNT/polyelectrolyte films in both directions. It is noted that while there are some differences between different curves for each type of nanoindentation tests, in general the curves are close to each other, indicating that the inhomogeneity in the material is not significant, primarily due to the fact that the indent area is very large compared to the size of a SWNT so that there are many SWNTs underneath an indenter tip, leading to an average behavior in the nanoindentation data. The relatively consistent results also provide evidence showing reasonably good dispersion of SWNTs in the nanocomposites. From Eqs. (9)–(11), the Young's relaxation modulus in the in-plane direction is determined as

$$E(t) = 787 + 325e^{-0.1t} + 254e^{-0.01t} + 208e^{-0.001t}\ \text{MPa} \tag{12}$$

and the Young's relaxation modulus in the through-thickness direction is

$$E(t) = 725 + 3053e^{-0.1t} + 2984e^{-0.01t} + 197e^{-0.001t}\ \text{MPa} \tag{13}$$

It is seen from Eqs. (12) and (13) that the Young's relaxation moduli in the steady state are 787 and 725 MPa in the in-plane and the through-thickness directions, respectively. The steady state values of the Young's relaxation modulus are the average values from four tests as shown in Fig. 6; the standard deviations are 75 and 18 MPa for the in-plane modulus and through-thickness modulus in the steady state, respectively. The in-plane modulus in the steady state is 8.6% higher than the through-thickness modulus in the steady state. This relatively small difference warrants the consideration of the film as a nearly isotropic material, and justifies the use of isotropic viscoelastic contact mechanics analysis to extract the viscoelastic properties of the SWNT/polyelectrolyte film.

The default average Young's moduli given by the MTS Nano

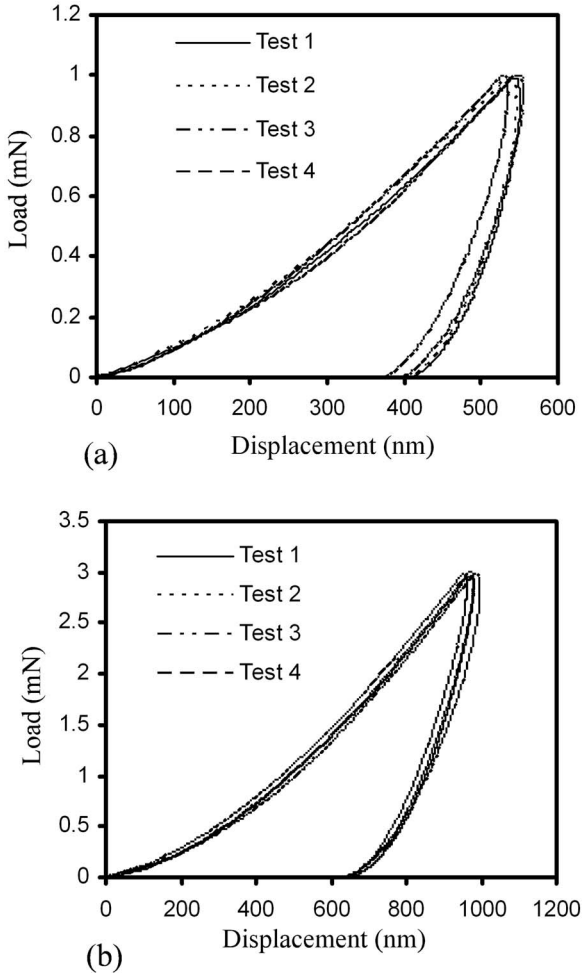


Fig. 6 Load-displacement curves from nanoindentation tests of SWNT/polyelectrolyte films in the in-plane and through-thickness directions. (*a*) In-plane load-displacement curves, (*b*) Through-thickness load-displacement curves.

XP system for the in-plane and the through-thickness directions are 4250 and 3470 MPa, respectively, determined from the unloading in the indentation load-displacement data, much higher than the steady state values as determined from viscoelastic analysis. It may be noted that the default results for Young's modulus are based on elastic-plastic analysis of the nanoindentation load-displacement curves, without taking into account of the time-dependent behavior of a viscoelastic material and could lead to significant overestimate of the Young's modulus [29]. Consequently, it is necessary to use viscoelastic analysis, as described in Sec. 4.4 to extract the Young's relaxation modulus.

5.2 Determination of Equivalent Length Coefficient in Mixture Model. Two sets of tensile models are used in finite element analysis. For the first set of models, a parallel SWNT or multiple parallel SWNTs are embedded in the matrix along the loading direction. For the other set, SWNTs dispersed in the matrix pseudo randomly are used. Numerical results from these models will be analyzed and discussed in this section. The first set of models with different geometries and volume fractions, with one of the models using three SWNTs in the matrix as shown in Fig. 2, are numerically analyzed to calculate the effective Young's modulus using Eq. (7). The equivalent length coefficient in Eq. (5), $\xi = \xi(l_0/l_1, f_{NT})$, depends on the size factor, l_0/l_1, and the volume fraction, f_{NT}. Finite element method (FEM) results are used in Eq. (7) to determine the Young's modulus. Substituting the Young's modulus to Eq. (5) leads to ξ for the given l_0/l_1 and f_{NT} used in FEM simulations. The geometry parameters of some models are listed in Table 1. For a fixed f_{NT} the equivalent length coefficient as a function of l_0/l_1 is plotted in Fig. 7(*a*). Results indicate that ξ does not change much with the increase of l_0/l_1, but ξ is much less than unity at small volume fractions, and ξ increases with the volume fraction. For a fixed l_0/l_1, the equivalent length coefficient ξ is plotted as a function of f_{NT} in Fig. 7(*b*). This figure indicates that ξ increases with the increase of f_{NT} and ξ appears to approach unity as f_{NT} becomes close to 5%. It is noted that $\xi=1$ corresponds to the condition that the end effects of SWNTs in Fig. 1(*a*) are negligible, so that the Liu-Chen approach [24,25] is appropriate. As shown in Fig. 7, ξ is much less than 1 when the volume

Table 1 Geometry of some FEM models

FE Models	l_0(nm)	l_1(nm)	f_{NT}
No. 1	40	73.3	0.047
No. 2	40	66.7	0.047
No. 3	40	56	0.047
No. 4	40	53.3	0.025
No. 5	40	50	0.025
No. 6	40	48	0.025
No. 7	40	73.3	0.01
No. 8	40	66.7	0.01
No. 9	40	56	0.01

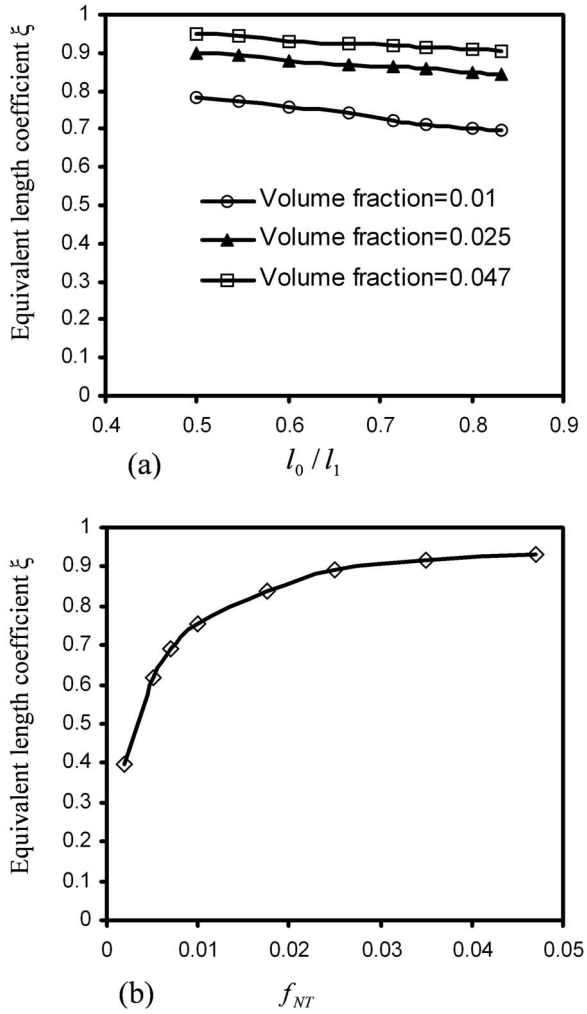


Fig. 7 Equivalent length coefficient $\xi=\xi(l_0/l_1, f_{NT})$; (*a*) ξ versus l_0/l_1 curves, (*b*) ξ versus f_{NT} curves

Table 2 Comparison of the Young's moduli from FEM and measurements

	Young's modulus (MPa)
Tensile testing	770
Nanoindentation, steady state (In plane)	787
Nanoindentation, steady state (Through thickness)	725
Eq. (5) model in set 1	831
FEM model a in set 2	843
FEM model b in set 2	734
FEM model c in set 2	711

fraction is very low. The finite element results indicate that the Liu-Chen approach cannot provide accurate prediction under small SWNT volume fractions. Consequently, it is determined that the Liu-Chen approach [24,25] is suitable for SWNT nanocomposites with relatively high volume fractions, but not suitable for relatively low SWNT volume fractions. The formula presented in this work is capable of predicting the Young's modulus from a very low SWNT volume fraction to a volume fraction near 5%.

5.3 Effective Material Properties of Nanocomposites Based on Representative Models. In the first set of models, SWNTs are aligned with the longitudinal direction of the representative volume element (RVE). In most situations, however, SWNTs are not expected to be aligned perfectly and uniformly distributed in the matrix. In the second set of models, SWNTs are pseudo-randomly distributed in the matrix. Table 2 gives the effective longitudinal Young's modulus E_e for three different models. Results indicate that these models give different effective Young's modulus data. But the difference is not huge, with the maximum difference of 16.9% in the Young's modulus data among these models. Consequently, the approach presented in this work could be used to estimate the Young's modulus based on the constituents in a SWNT nanocomposite.

5.4 Comparison of Numerical and Measurement Results. With the equivalent length coefficient ξ determined, Eq. 5 can be used to determine the effective Young's modulus for nanocomposites embedded with identical SWNTs. Figure 8 shows the results of the effective Young's modulus of SWNT nanocomposites over a range of relatively low volume fractions of SWNTs; for comparison, also plotted in Fig. 8 are the results of the effective Young's modulus data obtained from FEM simulations and the dilute approximate approach (Eq. (6)), where the dilute concentration factor is calculated by

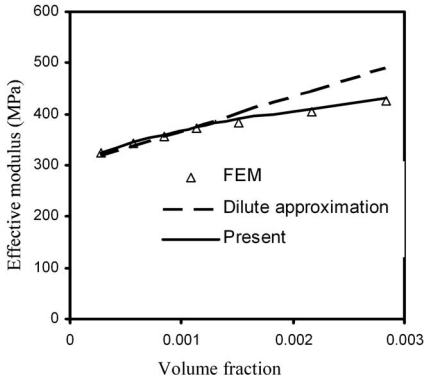


Fig. 8 Comparison of the effective Young's modulus determined from the dilute approximation approach, FEM, and the formula Eq. (5)

$$A^{NT,\mathrm{dil}} = \frac{\langle \epsilon^{NT} \rangle|_{\tilde{\epsilon}}}{\tilde{\epsilon}} = \frac{\int_{V_{NT}} \epsilon^{NT}|_{\tilde{\epsilon}} dV}{\tilde{\epsilon} \int_{V_{NT}} dV} \approx \frac{\sum_{e=1}^{n} \bar{\epsilon}^{e}|_{\tilde{\epsilon}} v_e}{\tilde{\epsilon} \sum_{e=1}^{n} v_e} \tag{14}$$

where $\langle \epsilon^{NT} \rangle$ and V_{NT} stand for the volumetric average strain and volume of inclusions, respectively, $\tilde{\epsilon}$ the homogeneous strain in the far field; $\bar{\epsilon}^{e}$ and v_e, provided from finite element analysis, are the strain component at element centroid and the volume of element, respectively. As shown in Fig. 8, the effective Young's modulus determined by Eq. (5) has a reasonably good agreement with the data from finite element analysis, while the effective Young's moduli calculated by the dilute approximate approach agree well with results from Eq. (5) at SWNT volume fraction lower than 0.13% but deviate from the results of Eq. (5) and finite element analysis at higher volume fractions, which manifests the fact that the dilute approximate approach is only applicable at very low volume fraction of inclusions, as expected.

Equation (5) is also applied to compute the effective Young's modulus for composite model as shown in Fig. 2, with SWNT volume fraction of 4.7%; the results are listed in Table 2. Also listed in Table 2 are results from finite element analysis and measurement results for the effective Young's moduli of the SWNT nanocomposites. The steady state value of the Young's relaxation modulus is close to the Young's modulus measured from tensile tests. But the Young's relaxation modulus at initial stage is much higher than that from tensile testing, as shown in Eqs. (12) and (13) due to viscoelastic effects.

As shown in Table 2, both the results from formula Eq. (5) and the numerical results on the effective Young's modulus are consistent, and are in a reasonably good agreement with tensile data, and nanoindentation data in the steady state. Consequently, the formula Eq. (5) with the use of ξ determined from finite element analysis can be used to estimate the effective Young's modulus of SWNT nanocomposites. However, it should be noted that some assumptions are used in this study, such as a perfect bonding between SWNTs and polymer matrix, assuming nanotubes in straight form instead of wavy form, simplification of REV elements with a size-dependent geometry coefficient that might not be suitable for highly randomly distributed CNTs with very high volume fractions. A more complete model should consider higher level of complexities, such as interlocking of SWNTs, randomness of CNT shapes, very high SWNT volume fractions, viscoelastic effects in the matrix, and the SWNT/polymer nanocomposites.

6 Conclusions

Based on the strength of materials and the finite element analysis, a formula has been derived from a mixture model to determine the effective in-plane Young's modulus for SWNT nanocomposite film between 0% and 5%. The Liu-Chen approach is examined and found that it is suitable for relatively high volume fractions, close to 5%, but not appropriate for relatively low volume fractions. The formula with the use of an equivalent length coefficient allows determination of the effective Young's modulus from very low SWNT volume fraction to about 5% SWNT volume fraction. Both tensile and nanoindentation measurements are used to determine the material properties of the neat LBL polyelectrolyte film and SWNT/polyelectrolyte film. Nanoindentation data show that both neat resin and nanocomposite films show viscoelastic effects, the Young's modulus of the in-plane modulus in the steady state is very close to the Young's modulus determined from tensile tests. The Young's relaxation modulus in the in-plane direction is only slightly larger than the Young's relaxation modulus in the film through-thickness direction, indicating that the nanocomposite film is nearly isotropic, despite the preferred orientation (parallel to the film plane) of the SWNTs. With the use of 4.7% of SWNTs in the polyelectrolyte matrix, both the modulus and strength have

been increased by over 150% over the neat resin. The increase in modulus due to reinforcement of SWNTs can be determined approximately by the model presented.

Acknowledgment

The work presented is sponsored by NASA Langley Research Center under Grant No. NNL04AA4ZG with Dr. Thomas S. Gates as program director. The authors acknowledge the support from the NSF under Grant No. CMS-9985060 for the acquisition of a nanoindenter for this research. The authors are grateful to Dr. Brian P. Grady at the University of Oklahoma for conducting small scale tensile tests.

References

[1] Saito, R., Dresselhaus, G., and Dresselhaus, M. S., 1998, *Physical Properties of Carbon Nanotubes*, Imperial College Press, London.
[2] Yu, M. F., Files, B. S., Arepalli, S., and Ruoff, R. S., 2000, "Tensile Loading of Ropes of Single Wall Carbon Nanotubes and Their Mechanical Properties," Phys. Rev. Lett., **84**, pp. 5552–5555.
[3] Odegard, G. M., Gates, T. S., and Herring, H. M., 2005, "Characterization of Viscoelastic Properties of Polymeric Materials through Nanoindentation," Exp. Mech., **45**(2), pp. 130–136.
[4] Alexandre, M., and Dubois, P., 2000, "Polymer-Layered Silicate Nanocomposites: Preparation, Properties and Uses of a New Class of Materials," Mater. Sci. Eng., R., **28**, pp. 1–63.
[5] Wang, Y., Sun, C., Sun, X., Hinkley, J., Odegard, G. M., and Gates, T. S., 2002, "Nano-Scale Finite Element Analysis of Polymer Networks," AIAA J., **1318**, pp. 1–10.
[6] Yu, M. F., Files, B. S., Arepalli, S., and Ruoff, R. S., 2000, "Tensile Loading of Ropes of Single Wall Carbon Nanotubes and Their Mechanical Properties," Phys. Rev. Lett., **84**(24), pp. 5552–5555.
[7] Li, F., Cheng, H. M., Bai, S., Su, G., and Dresselhaus, M. S., 2000, "Tensile Strength of Single-Walled Carbon Nanotubes Directly Measured from their Macroscopic Ropes," Appl. Phys. Lett., **77**(20), pp. 3161–3163.
[8] Mamedov, A. A., Kotov, N., Prato, M., and Guldi, D. M., 2002, "Molecular Design of Strong Single-Wall Carbon Nanotubes/Polyelectrolyte Multiplayer Composites," Nat. Mater., **1**, pp. 190–194.
[9] Kotov, N. A., Mamedov, A. A., Guldi, D. M., Tang, Z., Prato, M., Wicksted, J., and Hirsch, A., 2004, *Dekker Encyclopedia of Nanoscience and Nanotechnology*, Dekker, New York, pp. 2607–2613.
[10] Lau, K., Chipara, M., Ling, H., and Hui, D., 2004, "On the Effective Elastic Moduli of Carbon Nanotubes for Nanocomposite Structures," Composites, Part B, **35**, pp. 95–101.
[11] Odegard, G. M., Gates, T. S., Nicholson, L. M., and Wise, K. E., 2002, "Equivalent-Continuum Modeling of Nano-Structured Materials," Compos. Sci. Technol., **62**(14), pp. 1869–1880.
[12] Odegard, G. M., Gates, T. S., Wise, K. E., Park, C., and Siochi, E. J., 2003, "Constitutive Modeling of Nanotube–Reinforced Polymer Composites," Compos. Sci. Technol., **63**, pp. 1671–1687.
[13] Valavala, P. K., and Odegard, G. M., 2005, "Modeling Techniques for Determination of Mechanical Properties of Polymer Nanocomposites," Rev. Adv. Mater. Sci., **9**, pp. 34–44.
[14] Christensen, R. M., 1997, "Stress Based Yield/Failure Criteria for Fiber Composites," Int. J. Solids Struct., **34**(5), pp. 529–543.
[15] Christensen, R. M., 2000, "Mechanics of Cellular and Other Low-Density Materials," Int. J. Solids Struct., **37**, pp. 93–104.
[16] Christensen, R. M., 1996, "On the Relationship of Minimal Conditions to Low Density Material Microstructures," J. Mech. Phys. Solids, **44**(12), pp. 2113–2123.
[17] Yang, C., Huh, H., and Hahn, T. H., 2003, "Evaluation of Effective Material Properties of Composite Materials Using Special FEM," J. Mater. Process. Technol., **140**, pp. 185–190.
[18] Hu, N., Wang, B., Tan, G. W., Yao, Z. H., and Yuan, W. F., 2000, "Effective Elastic Properties of 2-D Solid With Circular Holes: Numerical Simulation," Compos. Sci. Technol., **60**, pp. 1811–1823.
[19] Harik, V. M., 2002, "Mechanics of Carbon Nanotubes: Applicability of the Continuum-Beam Models," Comput. Mater. Sci., **24**, pp. 328–342.
[20] Harik, V. M., 2001, "Ranges of Applicability for the Continuum Beam Model in the Mechanics of Carbon Nanotubes and Nanorods," Solid State Commun., **120**, pp. 331–335.
[21] Mori, T., and Tanaka, K., 1973, "Average Stress in Matrix and Average Elastic Energy of Materials With Misfitting Inclusions," Acta Metall., **21**, pp. 571–574.
[22] Fisher, F. T., Bradshaw, R. D., and Brinson, L. C., 2003, "Fibre Waviness in Nanotube-Reinforced Polymer Composites- I: Modulus Predictions Using Effective Nanotube Properties," Compos. Sci. Technol., **63**, pp. 1689–1703.
[23] Bradshaw, R. D., Fisher, F. T., and Brinson, L. C., 2003, "Fibre Waviness in Nanotube-Reinforced Polymer Composites- II: Modeling via Numerical Approximation of the Dilute Strain Concentration Tensor," Compos. Sci. Technol., **63**, pp. 1705–1722.
[24] Liu, Y. J., and Chen, X. L., 2003, "Evaluation of the Effective Material Properties of Carbon Nanotube-Based Composites Using a Nanoscale Representative Volume Element," Mech. Mater., **35**, pp. 69–81.
[25] Chen, X. L., and Liu, Y. J., 2003, "Square Representative Volume Elements for Evaluating the Effective Material Properties of Carbon Nanotube-Based Composites," Mech. Mater., **29**, pp. 1–11.
[26] Hammond, P. T., 2004, "Form and Function in Multiplayer Assembly: New Applications at the Nanoscale," Nat. Mater., **16**(15), pp. 1271–1293.
[27] Lui, J., Rinzler, A. G., Dai, H., Hafner, J. H., Bradley, R. K., Boul, P. J., Lu, A., Iverson, T., Shelimov, K., Huffman, C. B., Rodriguez-Macias, F., Shon, Y.-S., Lee, T. R., Colbert, D. T., and Smalley, R. E., 1998, "Fullerene Pipes," Science, **280**(5367), pp. 1253–1256.
[28] Decher, G., 1997, "Fuzzy Nanoassemblies: Toward Layered Polymeric Multicomposites," Science, **277**, pp. 1232–1237.
[29] Lu, H., Wang, B., Ma, J., Huang, G., and Viswananthan, H., 2003, "Measurement of Creep Compliance of Solid Polymers by Nanoindentation," Mech. Time-Depend. Mater., **7**, pp. 189–207.
[30] Lee, E. H., and Radok, J. R. M., 1960, "The Contact Problem for Viscoelastic Bodies," ASME J. Appl. Mech., **27**, pp. 438–444.

Peter A. Gustafson
University of Michigan,
Ann Arbor, MI 48109
e-mail: petegus@umich.edu

Stephen J. Harris

Ann E. O'Neill

Physical and Environmental Science Department,
Ford Research and Advanced Engineering Center,
Dearborn, MI 48121

Anthony M. Waas
Professor of Aerospace Engineering
Fellow ASME
University of Michigan,
Ann Arbor, MI 48109
e-mail: dcw@umich.edu

Measurement of Biaxial Stress States in Silicon Using Micro-Raman Spectroscopy

Micro-Raman spectroscopy is used to determine the multiaxial stress state in silicon wafers using a strategy proposed by Narayanan, et al. (J. Appl. Phys. ***82****, 2595–2602 (1997)) Previously, this strategy was validated when silicon was subjected to uniaxial stress in the laboratory frame (Harris, et al. J. Appl. Phys.* ***96****, 7195–7201 (2004)). In the present work, silicon wafers have been analyzed that were subjected to biaxial stress states in the laboratory frame. The predicted curves for the initially degenerate F_{2g} peaks were found to fall within the variability of the measured curves. Stress ratios were found to be predictable. Stress magnitudes were also found to be predictable, but are subject to uncertainty greater than 25%. To perform these tests, an apparatus has been developed which can provide controlled ratios of biaxial stress in a simple and compact test geometry. This fixture was used under a microscope, enabling in situ measurement of biaxial stress states.* [DOI: 10.1115/1.2187527]

1 Introduction

There exist a large number of applications where it is desirable to determine the unknown stress state of a silicon structure. An example from the automotive industry is the silicon particles present in aluminum alloys. Cast aluminum components are now commonly used to achieve significant vehicle weight savings. Therefore, it is important to understand the failure mechanisms associated with these alloys. It has been shown that the initial ductile fracture in cast aluminum occurs at the onset of cracking of the eutectic silicon particles [1]. Due to their prominent role in ductile failure, it is desirable to understand the complete stress states and the mechanisms of load transfer that are present in silicon particles. Several theories of load transfer have been proposed [2–4], however the assumptions inherent in those theories make them inadequate to predict load transfer in aluminum alloys [5].

Micro-Raman spectroscopy has been used extensively to analyze local stresses and strains in a variety of materials [6–14]. However, conventional use of the technique has been limited to simple stress states (e.g., uniaxial or equibiaxial), where the Raman shift can be calibrated against a known stress magnitude. Full knowledge of a general stress tensor (six components) cannot be achieved by measuring a single Raman shift. However, Narayanan et al. [15] proposed a strategy for determining all three plane stress components by taking advantage not only of the shift, but also of the polarization and intensities of the three Raman components. In prior work [5], Micro-Raman spectroscopy provided an in situ measurement of particle stress (three components) for isolated eutectic silicon particles on the surface of cast aluminum. The technique was validated for a uniaxial stress state. Though relevant for simple loading, a more general validation is required for the complex stress fields found in aluminum alloys and other structures containing silicon.

In this paper, experiments on ⟨111⟩-normal silicon wafers which support the strategy of Narayanan et al. [15] for the case of biaxial stress are described. The eventual aim is to provide experimental in situ measurements of the stress states of individual silicon particles before and during ductile fracture experiments on cast aluminum. This would include measurements in cracked particles, whose presence can have an important effect on matrix flow [16].

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received December 16, 2004; final manuscript received January 18, 2006. Review conducted by K. M. Liechti. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California – Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Background

The process of predicting the Raman shift from a known state of strain has been widely reported [15,17,18]. For brevity, only key equations are presented. Each phonon mode j is associated with a frequency ω_j and an effective force constant K_j^0, which is the second derivative of the crystal potential energy with respect to the mode normal coordinates. Due to the cubic symmetry of silicon, there are three phonons which are degenerate. However, non-hydrostatic strain causes a change in the force constants, breaking the symmetry, lifting the degeneracy, and splitting the phonons into three separate frequencies. The splitting can be represented by a 3×3 matrix $[\Delta K]$ in the degenerate modes of the eigenfrequencies, where the eigenvalues and eigenvectors of the $[\Delta K]$ matrix are related to the Raman frequency shift measured by the spectrometer. The relationship between $[\Delta K]$, the phonon deformation potentials (p, q, r), and the applied strain ε in the crystal reference frame is given by

$$\begin{pmatrix} \Delta K_{11} \\ \Delta K_{22} \\ \Delta K_{33} \\ 2\Delta K_{23} \\ 2\Delta K_{13} \\ 2\Delta K_{12} \end{pmatrix} = \begin{pmatrix} p & q & q & 0 & 0 & 0 \\ q & p & q & 0 & 0 & 0 \\ q & q & p & 0 & 0 & 0 \\ 0 & 0 & 0 & r & 0 & 0 \\ 0 & 0 & 0 & 0 & r & 0 \\ 0 & 0 & 0 & 0 & 0 & r \end{pmatrix} \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \varepsilon_{23} \\ \varepsilon_{13} \\ \varepsilon_{12} \end{pmatrix} \tag{1}$$

The relationship between the eigenvalues, represented by λ, and the frequency shifts, $\Delta\omega$, is given by

$$\lambda = 2\omega_0\Delta\omega \tag{2}$$

where ω_0 is the Raman frequency of the F_{2g} mode in the unstressed condition, reported to be 520 cm^{-1} [17].

The splitting, which carries complete information about the strain that caused it [19], is not large enough for a Raman spec-

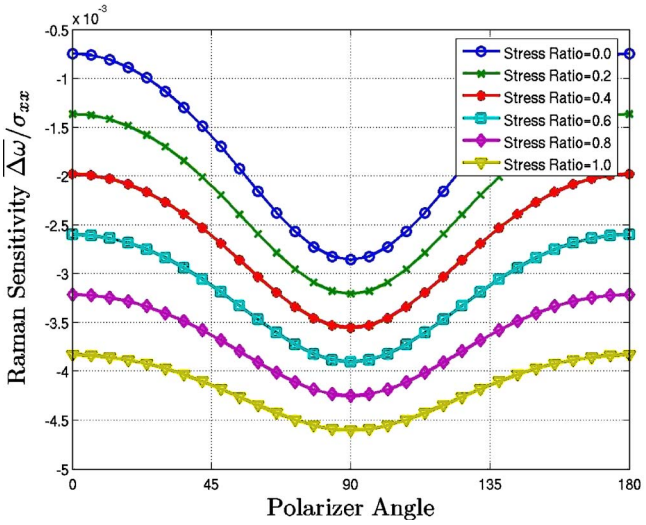


Fig. 1 Predictions of observed Raman shift ($\overline{\Delta\omega}$) per stress σ_{xx} for several ratios of σ_{yy}/σ_{xx}, based on Anastassakis PDPs

trometer to resolve in silicon. However, Narayanan et al. [15] suggested that, to a first approximation, the measured frequency shift $\overline{\Delta\omega}$ is the intensity-weighted average of the shifts of the three components, where the components are represented by $\Delta\omega_i$, their intensities by I_i, and the total (sum) intensity by I_T.

Since the three components have different polarizations, inserting a polarizer in the optical path selects different combinations of the three components, shifting the apparent frequency as described by

$$\overline{\Delta\omega} = \sum_{i=1}^{3} \frac{\Delta\omega_i I_i}{I_T} \tag{3}$$

For a given strain, knowing the values of the phonon deformation potentials (PDPs) allows prediction of the individual components and the overall shift seen through a polarizer. Using the PDPs provided by Anastassakis [20], Fig. 1 shows the predicted apparent Raman shift $\overline{\Delta\omega}$ per unit applied stress σ_{xx} in the laboratory reference frame, for various levels of applied stress σ_{yy}. It has been experimentally verified that the dependence on polarizer angle follows the predictions of Eq. (3).

Predicting stress based on Raman shift depends on inverting the process described; first measuring the line center position as a function of polarizer angle, and then calculating the corresponding stresses. The strategy can be successful if the inversion is unique.

The curves shown in Fig. 1 can be described with two parameters: $\overline{\Delta\omega_0}$, the shift observed through a 0 deg polarizer, and $\overline{\Delta\omega_{90}}$, the shift observed through a 90 deg polarizer. Narayanan et al. [15] proposed the use of Eqs. (4*a*) and (4*b*) in order to determine an unknown biaxial stresses in silicon

$$\overline{\Delta\omega_0} = a\sigma_{xx} + b\sigma_{yy} \tag{4a}$$

$$\overline{\Delta\omega_{90}} = c\sigma_{xx} + d\sigma_{yy} \tag{4b}$$

where a, b, c, and d are constants. Defining R as the stress ratio,

$$R \equiv \frac{\sigma_{xx}}{\sigma_{yy}} \tag{5}$$

and after dividing by σ_{xx}, the result is Eqs. (6*a*) and (6*b*), relating Raman sensitivity $\overline{\Delta\omega_0}/\sigma_{xx}$ to stress ratio R,

$$\frac{\overline{\Delta\omega_0}}{\sigma_{xx}} = a + bR \tag{6a}$$

$$\frac{\overline{\Delta\omega_{90}}}{\sigma_{xx}} = c + dR \tag{6b}$$

In this work, the focus is on verification of this technique by examining several ratios of biaxial stress against these equations. To do so, a novel biaxial stressing device is used.

3 Controlling Stress Ratio Through the Use of Elliptical Test Specimens

To validate Eqs. (6*a*) and (6*b*), a controlled biaxial stress field must be created in silicon and presented to the micro-Raman spectrometer. A pressurized, clamped elliptical plate facilitates the creation of a controllable biaxial stress field on the plate surface. The mechanics of the device manufactured for this purpose are presented in the following paragraphs.

Consider a thin flat plate of uniform, linearly elastic, isotropic material that is clamped elliptically at its edges and loaded with a uniform pressure. Such an elliptical plate with semi-axes $2a$ and $2b$ is defined in $x-y$ space as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

When clamped and subjected to a uniform pressure load, the disk has the following boundary conditions:

$$w = \frac{\partial w}{\partial n} = 0 \tag{7}$$

In Eq. (7), w is the out of plane deflection of the plate, and n is the normal to the ellipse edge at the plate boundary. Assuming small displacements and neglecting shear stresses on the middle plane, the following solution for deflection [21] is obtained for the out of plane deflection $w(x,y)$:

$$w(x,y) = k\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)^2 \tag{8}$$

where k is a constant given by

$$k = \left(\frac{q}{8D}\right)\left(\frac{a^4b^4}{3a^4 + 2a^2b^2 + 3b^4}\right)$$

and D is the flexural rigidity of the plate, defined as

$$D = \frac{Et^3}{12(1-\nu^2)}$$

where E is the flexural modulus of the plate, ν the Poisson's ratio, and t the plate thickness. For a circular plate, where the ellipse ratio $a:b$ is 1:1, the deflection due to a uniform pressure load q at radius r is given by Eq. (9), which is identical to the small deflection solution provided by Timoshenko [22]. The infinite strip solution is also satisfied when appropriate a and b are inserted into Eq. (8) [21],

$$w(r) = \frac{q(a^2 + r^2)^2}{64D} \tag{9}$$

Using the above equations, the bending and twisting moment distributions are given by

$$M_{xx}(x,y) = 4Dk\left[\frac{1}{a^2} - 3\frac{x^2}{a^4} - \frac{y^2}{a^2b^2} + \nu\left(\frac{1}{b^2} - 3\frac{y^2}{b^4} - \frac{x^2}{a^2b^2}\right)\right] \tag{10a}$$

$$M_{yy}(x,y) = 4Dk\left[\frac{1}{b^2} - 3\frac{y^2}{b^4} - \frac{x^2}{a^2b^2} + \nu\left(\frac{1}{a^2} - 3\frac{x^2}{a^4} - \frac{y^2}{a^2b^2}\right)\right] \tag{10b}$$

$$M_{xy}(x,y) = (\nu - 1)\frac{8Dk}{a^2b^2}xy \tag{10c}$$

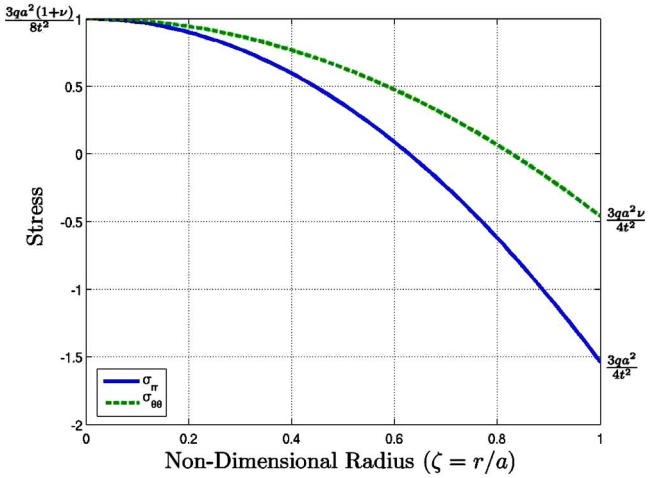


Fig. 2 Stresses σ_{rr} and $\sigma_{\theta\theta}$ in an isotropic, clamped, circular plate (diameter 2*a*, thickness *h*, Poisson's ratio ν) subjected to uniform pressure load (*q*). There is an area in the center of the plate subjected to a uniform biaxial stress field.

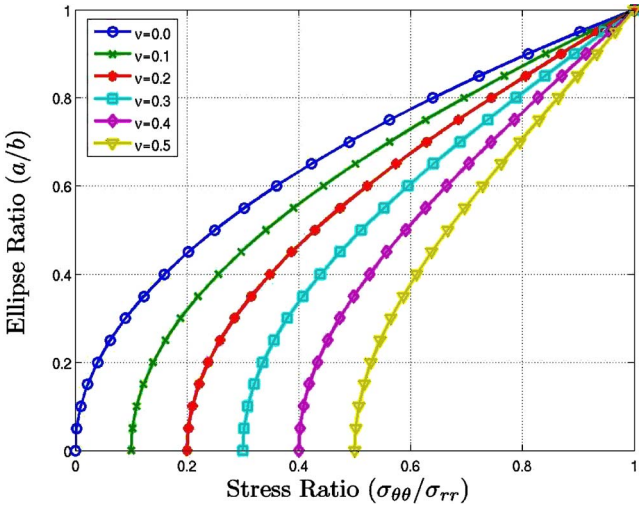


Fig. 3 Theoretical relationship between stress ratio and ellipse ratio, plotted for a range of Poisson's ratios

At the center of the disk ($x=y=0$), the bending moments are given by

$$M_{xx} = 4Dk\left[\frac{1}{a^2} + \nu\frac{1}{b^2}\right] \tag{11a}$$

$$M_{yy} = 4Dk\left[\frac{1}{b^2} + \nu\frac{1}{a^2}\right] \tag{11b}$$

$$M_{xy} = 0 \tag{11c}$$

Using the moments above, the stresses at the center of the disk surface are given by Eqs. (12*a*) and (12*b*), providing a convenient method for calculating the stress ratio at the crown of the plate

$$\sigma_{xx} = 6\frac{M_{xx}}{t^2} \tag{12a}$$

$$\sigma_{yy} = 6\frac{M_{yy}}{t^2} \tag{12b}$$

For a circular plate, the bending moments (in cylindrical coordinates) are given by Eqs. (13*a*) and (13*b*), providing stresses that are plotted in Fig. 2,

$$M_{rr}(r) = \frac{q}{16}[a^2(1+\nu) - r^2(3+\nu)] \tag{13a}$$

$$M_{\theta\theta}(r) = \frac{q}{16}[a^2(1+\nu) - r^2(1+3\nu)] \tag{13b}$$

The radial and tangential stresses found at the center of the plate ($r=0$) are of equal magnitude as given in Eq. (14). Thus an equibiaxial stress state is found at the crown of a circular disk with the loading described,

$$\sigma_{rr}^{\max} = \sigma_{\theta\theta}^{\max} = \frac{3qa^2}{8t^2}(1+\nu) \tag{14}$$

It is evident in Fig. 2 that a region in the center of the disk is subjected to an approximately uniform biaxial stress field. Similar regions of approximately uniform stress are also present in an elliptical disk. These areas can be exploited for the purpose of Raman spectroscopy.

Equations (13*a*) and (13*b*) provide a convenient method for calculating the major and minor axis ratio that is required in order to achieve a desired stress ratio. Define ρ as the ellipse ratio:

$$\rho \equiv \frac{b}{a} \tag{15}$$

Combining Eqs. (5), (13*a*), and (13*b*) leads to Eq. (16), which relates the stress ratio to the ellipse ratio as plotted in Fig. 3,

$$R = \frac{[b^2/(ab)^2] + \nu[a^2/(ab)^2]}{[a^2/(ab)^2] + \nu[b^2/(ab)^2]} = \frac{\rho^2 + \nu}{1 + \nu\rho^2} \tag{16}$$

To determine the ellipse ratio required for a desired stress ratio, Eq. (16) has two solutions, given by

$$\left[\rho = \sqrt{\frac{\nu - R}{R\nu - 1}}, \rho = -\sqrt{\frac{\nu - R}{R\nu - 1}}\right] \tag{17}$$

Only the first of these two solutions is valid.

4 Finite Element (FE) Model

The solution presented in Sec. 3 was developed for an isotropic material, characterized by two constants. However, silicon is a cubic material characterized by three constants. In order to confirm the usefulness of the elliptical plate technique, a finite element model was developed in the commercial code ABAQUS, and is shown in Fig. 4.

In order to correlate with testing, the finite element model included the rubber gaskets that are present in the experimental setup at the periphery of the elliptical plates. Displacement (w) is constrained on the top and bottom surfaces of the rubber gaskets, where the rubber is clamped (see Fig. 4). Cubic material properties were assumed for the silicon, where the stiffness matrix C_{ij} in the crystal orientation is given in Eq. (18). The rubber gasket was not of interest and was therefore modeled very simply with linear elastic isotropic material properties. The Poisson's ratio was chosen to be 0.45, as it provided the best correlation with strain gauge measurements. A range of rubber moduli were applied in different models to simulate the inconsistencies that are present in the clamp-down technique for the silicon wafers. Quadratic solid elements were used for all materials,

$$C_{ij} = \begin{pmatrix} 166 & 64 & 64 & 0 & 0 & 0 \\ 64 & 166 & 64 & 0 & 0 & 0 \\ 64 & 64 & 166 & 0 & 0 & 0 \\ 0 & 0 & 0 & 80 & 0 & 0 \\ 0 & 0 & 0 & 0 & 80 & 0 \\ 0 & 0 & 0 & 0 & 0 & 80 \end{pmatrix} \text{GPa} \tag{18}$$

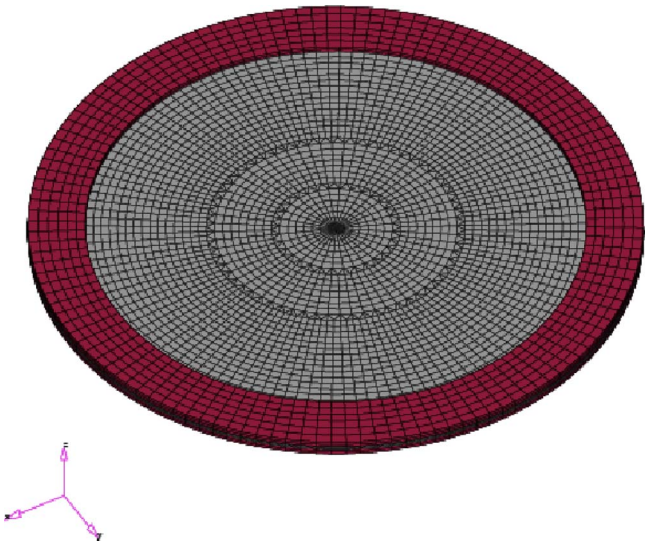


Fig. 4 Isometric view of the ABAQUS model. The silicon and rubber gaskets are modeled with solid elements.

The simplistic geometry of the setup and the plate-like sample might seem to be efficiently modeled with shell elements. However, 3D elements are required to properly capture the differences between the laboratory reference frame and the orientation of the silicon crystal reference frame. The $\langle 111 \rangle$ direction in silicon is oriented along the normal to the plate surface. The ABAQUS non-linear geometry assumption was used, as it was found in the model and experimentally that a linear solution is not adequate for the highest pressures applied during the experiments. A mesh convergence study was completed to ensure adequate mesh density.

The best correlations between the FE results and the experimental results are presented in Table 1. These results confirm that the aspect ratio ($a:b$) controls the ratio of biaxial strain and stress at the crown of the silicon wafer. The predicted pressure-strain response compares favorably with the measured strain. Exact correlation was not essential due to the test-to-test variation in the clamped boundary condition. The FE results are used as a comparative verification of measured strains. Therefore, the results in Table 1 confirm that this device is an effective tool for providing controlled biaxial stress states in silicon.

5 Experimental Setup

Silicon wafers (125 mm in diameter and 0.62 mm thick) were clamped in the elliptical fixture shown in Fig. 5. The fixture consists of a steel base plate with air passages and a set of elliptical clamping plates. The geometry of the ellipse creates the controlled stress fields in each specimen as described above. For each test, a silicon sample was seated between rubber gaskets and clamped between the base plate and an elliptical plate. Air pressure was supplied from a nitrogen cylinder and pressure regulator. The applied pressure varied from zero to the maximum pressure, where the maximum pressure ranged between 60 and 180 KPa depending on the elliptical plate. Pressure was measured with an MKS Baratron pressure transducer model #750B14TFB2GA (0–10,000 Torr). In addition to the elliptical plates, a micro three-point bend device was used to obtain Raman shifts for approximately uniaxial stress states, as previously reported in Harris et al. [5].

Table 1 Initial slope of strain versus applied pressure response for the finite element model and for measured strains

Plate ratio ($a:b$)	$\partial\epsilon_{xx}/\partial P$ measured [$\mu\epsilon$/kPa]	$\partial\epsilon_{xx}/\partial P$ ABAQUS [$\mu\epsilon$/kPa]	$\partial\epsilon_{yy}/\partial P$ ABAQUS [$\mu\epsilon$/kPa]	Strain ratio $\epsilon_{yy}/\epsilon_{xx}$
1.0:1	10.71	10.71	10.55	0.98
1.5:1	10.20	10.37	6.33	0.61
2.0:1	6.09	6.12	1.88	0.31
3.0:1	3.44	3.47	0.52	0.15

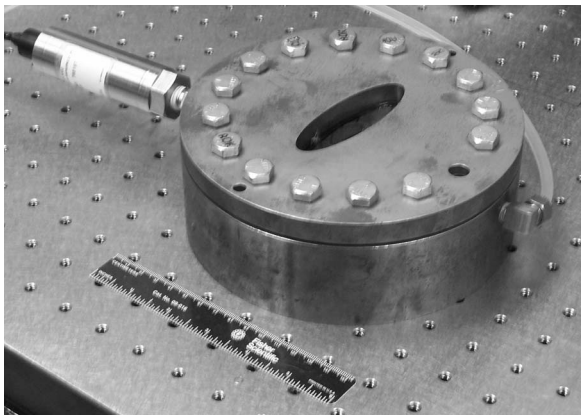


Fig. 5 Elliptical plate loading device, consisting of a steel base plate with air feed and an elliptical clamping plate. Rubber gaskets are used to seat the clamped silicon wafer.

Raman measurements were carried out in backscattering mode with a Renishaw spectrometer using the 633 nm line from a He-Ne laser. The device was oriented so that the minor axis of the ellipse (X axis in the lab coordinate frame) was perpendicular to the polarization of the incident beam. The silicon sample, with $\langle 111 \rangle$-normal, was oriented so that the $\langle 110 \rangle$ crystal axis was along the minor axis of the ellipse.

The laser was focused at the center of the ellipse through a 20× objective lens with a 2 cm working distance, which resulted in a 2–3 μm diameter laser spot, measured through the microscope. The relatively long focal length helped ensure that the incident and scattering angles were nearly perpendicular to the surface, so that the polarization states were well-defined [15]. Because the optical absorption coefficient of silicon at 633 nm is approximately 3.9103 cm^{-1}, only the top few microns of the silicon wafer (out of a total thickness of 620 μm) were sampled and are considered to be subjected to plane stress. The laser power at the focus was approximately 2 mW, corresponding to an average intensity of about 40 kW/cm^2. Because of the relatively small absorption coefficient of silicon at the He-Ne wavelength, the laser power was spread over a relatively large volume, and the predicted temperature rise in the illuminated volume is negligible. The back-scattered Raman light was collimated and passed through a rotatable polarizer before entering the spectrometer, which has a 2400 groove/mm grating. A charge coupled device (CCD) detected the light, and the grating produced a dispersion of 0.66 cm^{-1}/pixel. The silicon line near 520 cm^{-1} is a convolution of its intrinsic shape and the spectrometer instrument function. The width of the 520 cm^{-1} line is about 4 cm^{-1}. The silicon line was fit between 500 and 540 cm^{-1} using the five-parameter pseudo-Voigt profile given in

$$y(\omega) = y_0 + a\left[c\left(\frac{1}{1+(\omega-\omega_0/b)^2}\right) + (1-c)e^{-0.5\left(\frac{\omega-\omega_0}{b}\right)^2}\right] \tag{19}$$

Typically, c was close to 0.5 (equal parts Gaussian and Lorentzian), and the R^2 correlation coefficient for the fit was between 0.9990 and 0.9999. Although neither the resolution of the spectrometer nor the pixel density of the CCD is particularly high, the repeatability of the spectra and the quality of the Voigt fit were excellent because of the extremely high signal-to-noise ratio

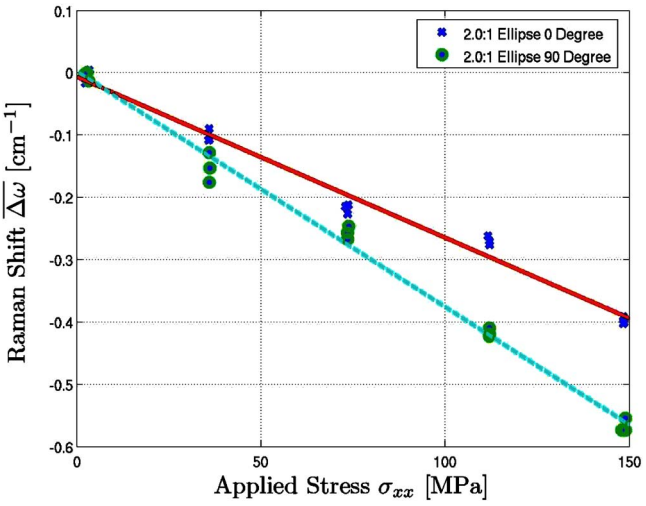


Fig. 6 Measured Raman shift $\overline{\Delta\omega}$ plotted as a function of applied stress σ_{xx} with an in-line 0 deg and 90 deg polarizer. These data, taken with the 2:1 elliptical plate, are an example of the output for each experiment.

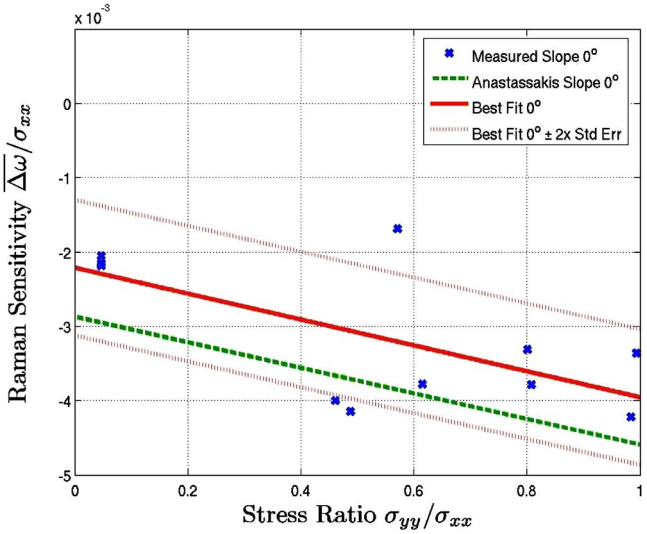


Fig. 7 Raman sensitivity $\overline{\Delta\omega_0}/\sigma_{xx}$ plotted versus stress ratio σ_{yy}/σ_{xx} for measured data (0 deg polarizer) and for predicted result using Anastassakis PDPs. The predictions fall within the variability in the measured data.

($\sim 10^3$). As a result, the standard error for the line center position ω_0 was quite small, between 0.002 and 0.003 cm^{-1} for the grating used. Raman spectra were accumulated for whatever time was required to give 20,000 to 30,000 counts at the peak of the silicon line, typically between 20 and 100 s depending on the orientation of the in-line polarizer.

It was found during experimentation that small variations in the focal plane of the laser had a substantial effect on the measured Raman shift, as much as 0.1 cm^{-1} for a 5 μm shift in the focal plane. Therefore, the focus was tuned before each measurement by maximizing the intensity of the scattered light into the spectrometer. In order to correct for drift in the spectrometer, a pair of plasma lines from the laser, at 434 and 649 cm^{-1}, were used as internal frequency standards.

For each data set, three data points were taken at each pressure as the gauge pressure was ramped from zero to the maximum pressure. Data were taken with the polarizer set at 0 deg and 90 deg relative to the polarization of the laser. The polarizer is slightly wedged, which causes the path of the light entering the spectrometer to be bent by about 0.5 mrad. Because entrance angle affects the frequency measurement of the spectrometer, rotating the wedged polarizer by 180 deg about its own axis changes the observed frequency $\overline{\Delta\omega}$. To avoid this problem, two polarizers were placed in the light path, in series, oriented such that the effect is cancelled [5].

Strains were measured as close as was possible to the crown of the bent silicon wafer. Surface strain was measured in the lab frame using standard strain rosettes. In the case of the uniaxially stressed wafer, a uniaxial gauge was used to measure the strain, since a strain rosette would not fit onto the sample.

Figure 6 presents an example of the measured Raman shifts plotted against applied σ_{xx} (lab frame) for 0 deg and 90 deg polarizers using a 2:1 ratio clamping plate. To determine the stress levels reported in the data, the stress tensor is calculated using the measured strains and the material stiffness rotated into the correct orientation. The measured Raman shift is linear with applied stress for each case. Stresses and strains are nonzero after initial clamping of the silicon wafer, due to the nonuniformity in the gaskets and bolt tensions. Since the slope of the Raman shift versus applied stress curve is the result of interest, the initial Raman shift is not reported. For the approximately uniaxial data taken on the three point bending mechanism, the transverse stress was approximated using a finite element model [5].

6 Results

In Figs. 7 and 8, measured stress ratios and their corresponding sensitivities are plotted for 0 deg and 90 deg polarizers. Additionally, a least squares linear curve fit is plotted through each data set. The constants a,b,c,d of Eqs. (6*a*) and (6*b*) are calculated and reported in Table 2. The standard errors in these constants are calculated using the formulas provided by Blum [23]. The error bars shown in these figures are calculated from the best fit line ± twice the linear standard error function. It should be mentioned that the error bars are calculated based on the variability of the data in the whole curve, whereas the variability is likely to be dependent on the applied stress ratio. In prior work [5], it was reported that the stress can be calculated to within 10% for uniaxial stress. In this work, there is likely more variability in the middle stress ratios than in the uniaxial or biaxial stress ratios. This is an artifact of the test geometry, where the higher aspect

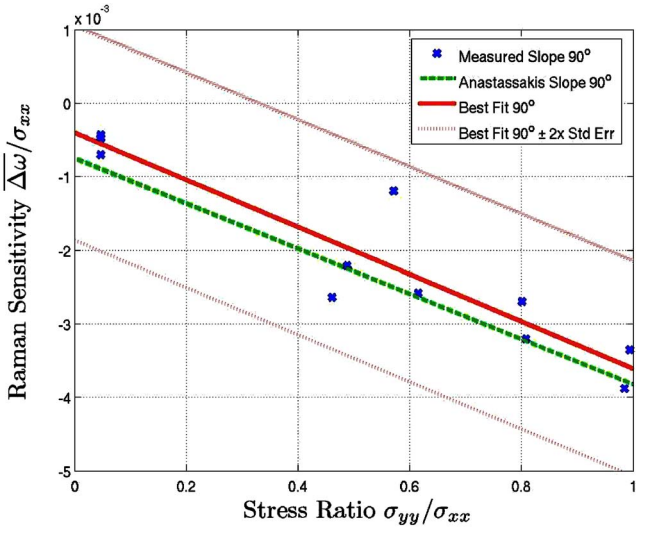


Fig. 8 Raman sensitivity $\overline{\Delta\omega_{90}}/\sigma_{xx}$ plotted versus stress ratio σ_{yy}/σ_{xx} for measured data (90 deg polarizer) and for predicted result using Anastassakis PDPs. The predictions fall within the variability in the measured data.

Table 2 Slope and intercept results in the $\overline{\Delta\omega}/\sigma_{xx}$ versus stress ratio plane

Constants from Eqs. (6*a*) and (6*b*)	Slope	Intercept	Slope Std error	Intercept Std error
0° Eq. (6*a*)	a	b	-	-
90° Eq. (6*b*)	c	d	-	-
0° Anastassakis[a]	$-1.72E-3$	$-2.87E-3$	-	-
90° Anastassakis[a]	$-3.07E-3$	$-7.53E-4$	-	-
0° Chandrasekhar[b]	$-1.42E-3$	$-2.30E-3$	-	-
90° Chandrasekhar[b]	$-2.52E-3$	$-5.55E-4$	-	-
0° measured	$-1.74E-3$	$-2.21E-3$	$5.44E-4$	$3.30E-4$
90° measured	$-3.21E-3$	$-4.09E-4$	$3.40E-4$	$2.06E-4$

[a]Reference [20]
[b]Reference [24]

ratio ellipses have crowns that are closer to the clamped boundary conditions (where clamping is not consistent from experiment to experiment). There are also larger stress gradients for these ellipses, therefore the stresses are less uniform around the laser focal point and strain gauges. It is worth noting that one data point, taken with the 2:1 ellipse, significantly increases the variability in the data. The error bars increase by 31 and 45% for the 0 deg and 90 deg polarizers, respectively, over what they would be if the data point were removed. There are correspondingly large increases in the standard errors of the slope and intercept constants of Eqs. (6*a*) and (6*b*).

The results shown in Figs. 7 and 8 provide comparison between the measured Raman sensitivities to the predictions using the PDPs provided by Anastassakis [20]. In these figures, the variability in the measured data is evident. However, the trends are in agreement with the published PDPs. The predictions based on Chandrasekhar [24] are not plotted, though they are also within the measurement variability. The range of measured sensitivity data encompasses the two sets of predicted curves, even though the predicted values themselves differ by as much as 25%. Figure 9 indicates that the ratio of $\overline{\Delta\omega_0}/\overline{\Delta\omega_{90}}$ increases monotonically with increasing stress ratio. Further, it is shown that the Raman shift ratio is in good agreement with that which would be predicted using either the Anastassakis or Chandrasekhar PDPs.

The shift ratio is most sensitive when the stress ratio is small, as it is for uniaxial stress. The curve flattens and becomes less sensitive to stress ratio for values close to 1.0. The Raman shift ratio is sensitive to measurement variability for small stress ratios. This is because the measured shifts are smaller (particularly for $\Delta\omega_0$), while measurement variability stays constant. However, it appears that variability in the shift ratio is smaller than the variability in Raman sensitivity. Therefore, it can be concluded that the variability present in the Raman shift data affects the 0 deg and 90 deg polarizers proportionally, and that the stress ratio can be estimated accurately.

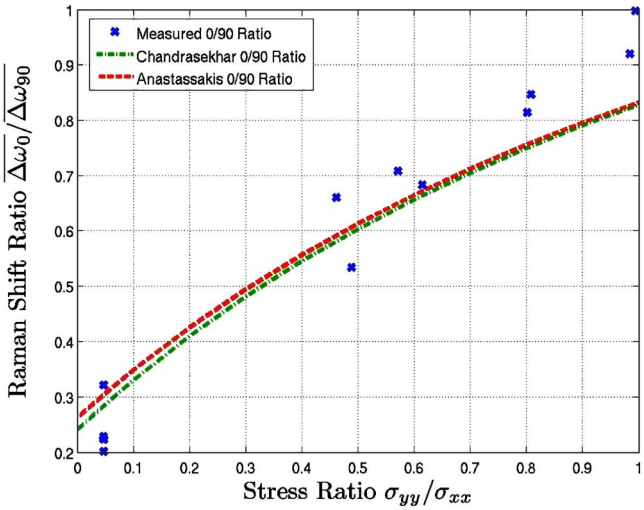


Fig. 9 Measured Raman shift ratio $\overline{\Delta\omega_0}/\overline{\Delta\omega_{90}}$ plotted as a function of stress ratio σ_{yy}/σ_{xx}. The variability present in the shift ratio data is smaller than the variability of the sensitivity data from which it was calculated.

The results shown in Figs. 7–9 can be used to determine the normal stress ratio and magnitude for ⟨111⟩-normal silicon subjected to an unknown plane stress state. To do so, first measure the Raman shifts using 0 deg and 90 deg polarizers to determine the stress ratio from Fig. 9. Then, using the known stress ratio, the stress magnitude can be determined by solving Eqs. (6*a*) and (6*b*) for stress, using the a,b,c,d constants provided in Table 2. Using this technique, two of the three planar stress values can be determined.

Stress prediction is relatively unreliable due to the measurement variability observed. Uncertainties of up to 25% are present simply based on the choice of PDPs used in the calculation. There are several additional sources of error that must be considered. Notably, variations in clamping pressure at the boundary lead to a wide range of initial stress conditions during expiremental testing. For the purpose of Raman sensitivity calculation, each sample was assumed to maintain a constant stress ratio as the pressure loading was increased. However the initial stress conditions caused a measurable deviation in each sample sequence. This effect was most pronounced in the 2:1 and 3:1 ellipse ratios. The effect of the boundary condition was minimized by including direct measurement of strain using gauges, however the boundary condition uncertainty could have created strain fields which were not adequately captured by the gauge. A second potential error source was the strain gauge rosettes. During the course of experimentation, difficulty was observed in establishing and maintaining good adhesion between the gauge and the polished silicon wafer. Despite diligent efforts to monitor bond integrity, a portion of the strain rosette could have lost adhesion during any given sample sequence. Additionally, a small amount of gauge error was introduced due to the location of the stain gauge rosette. The laser focused directly on the silicon wafer, therefore the rosette was required to be located several millimeters from the focal point of the laser. Since the in plane spatial derivatives of strain are small, this error source is expected to be small. Finally, the Raman shift was found to be very sensitive to laser focus. Though a procedure was established to insure consistent focus, the measurements were subject to focal variability and drift.

Additional measurements must be performed in order to quantify the confidence intervals present in this technique. Those measurements are the subject of future work. Verification of the micro-Raman technique for stress states that contain shear (in the laboratory reference frame) is also slated for future work.

7 Concluding Remarks

A compact device has been developed which presents controlled ratios of biaxial stress to a Raman spectrometer. The device has been used to generate biaxial stress ratios for verification of the micro-Raman technique. The measured Raman sensitivity data have a relatively large amount of variability. However, the trends compare favorably with predictions using the PDPs provided by either Chandrasekhar [24] or Anastassakis [20]. Predictions using either set of constants fall within the margin of error of the measured Raman shifts. Measurement variability affects the $\Delta\omega_0$ and $\Delta\omega_{90}$ proportionally, therefore the Raman shift ratio $\Delta\omega_0/\Delta\omega_{90}$ has less variability than the sensitivities used to calculate it. The Raman shift ratio, $\Delta\omega_0/\Delta\omega_{90}$, is more sensitive to changes in stress ratio at low stress ratios (uniaxial) than at high stress ratios (biaxial). It is also more sensitive to measurement variability at

low stress ratios than high stress ratios. The stress ratio σ_{yy}/σ_{xx} can be calculated from the measured $\overline{\Delta\omega_0}/\overline{\Delta\omega_{90}}$ ratio. Stress magnitudes can be calculated using the equations provided. Predictions of stress magnitude are expected to have a level of uncertainty greater than 25%.

References

[1] Wang, Q. G., Caceres, C. H., and Griffiths, J. R., 2003, "Damage by Eutectic Particle Cracking in Aluminum Casting Alloys A356/357," Metall. Mater. Trans. A, **34**(12), pp. 2901–2912.

[2] Clyne, T., 1989, "A Simple Development of the Shear Lag Theory Appropriate for Composites With a Relatively Small Modulus Mismatch," Mater. Sci. Eng., A, **122**(183), pp. 183–192.

[3] Eshelby, J. D., 1957, "The Determination of the Elastic Field of an Ellipsoidal Inclusion and Related Problem," Proc. R. Soc. London, Ser. A, **241**(376), pp. 376–396.

[4] Mori, T., and Tanaka, K., 1973, "Average Stress in Matrix and Average Elastic Energy of Materials with Misfitting Inclusions," Acta Metall., **21**(5), pp. 571–574.

[5] Harris, S. J., ONeill, A. E., Gustafson, P. A., Boileau, J., Yang, W., Weber, W. H., Majumdar, B., and Ghosh, S., 2004, "Measurement of the State of Stress in Silicon with Micro-Raman Spectroscopy," J. Appl. Phys., **96**(12), pp. 7195–7201.

[6] DeWolf, I., Ignat, M., Pozza, G., Maniguet, L., and Maes, H. E., 1999, "Analysis of Local Mechanical Stresses In and Near Tungsten Lines on Silicon Substrate," J. Appl. Phys., **85**(9), pp. 6477.

[7] DeWolf, I., 1996, "Micro-Raman Spectroscopy to Study Local Mechanical Stress in Silicon Integrated Circuits," Semicond. Sci. Technol., **11**(2), pp. 139–154.

[8] DeWolf, I., Maes, H. E., and Jones, S. K., 1996, "Stress Measurements in Silicon Devices through Raman Spectroscopy: Bridging the Gap Between Theory and Experiment," J. Appl. Phys., **79**(9), p. 7148.

[9] DeWolf, I., Pozzat, G., Pinardi, K., Howard, D. J., Ignat, M., Jain, S. C., and Maes, H. E., 1996, "Experimental Validation of Mechanical Stress Models by Micro-Raman Spectroscopy," Microelectron. Reliab., **36**(11-12), pp. 1751–1754.

[10] Ajito, K., Sukamto, J. P. H., Nagahara, L. A., Hashimoto, K., and Fujishima, A., 1995, "Strain Imaging Analysis of Si Using Raman Microscopy," **13**, pp. 1234–1238, AVS, May 1995.

[11] Anastassakis, E., and Liarokapis, E., 1987, "Polycrystalline Si Under Strain: Elastic and Lattice-Dynamical Considerations," J. Appl. Phys., **62**(8), pp. 3346–3352.

[12] Anastassakis, E., Pinczuk, A., Burstein, E., Pollak, F. H., and Cardona, M., 1970, "Effect of Static Uniaxial Stress on the Raman Spectrum of Silicon," Solid State Commun., **8**(2), pp. 133–138.

[13] Englert, Th., Abstreiter, G., and Pontcharra, J., 1980, "Determination of Existing Stress in Silicon Films on Sapphire Substrate Using Raman Spectroscopy," Solid-State Electron., **23**(1), pp. 31–33.

[14] Kobayashi, K., Inoue, Y., Nishimura, T., Hirayama, M., Akasaka, Y., Kato, T., and Ibuki, S., 1990, "Local-Oxidation-Induced Stress Measured by Raman Microprobe Spectroscopy," J. Electrochem. Soc., **137**(6), pp. 1987–1989.

[15] Narayanan, S., Kalidindi, S. R., and Schadler, L. S., 1997, "Determination of Unknown Stress States in Silicon Wafers Using Microlaser Raman Spectroscopy," J. Appl. Phys., **82**(5), pp. 2595–2602.

[16] Brockenbrough, J. R., and Zok, F. W., 1995, "Role of Particle Cracking in Flow and Fracture of Metal Matrix Composites," Acta Metall. Mater., **43**(1), pp. 11–20.

[17] Anastassakis, E. M., 1980, *Dynamic Properties of Solids*, North-Holland, Amsterdam.

[18] Loudon, R., 1964, "Raman Effect in Crystals," Adv. Phys., **13**(52), pp. 423–430.

[19] Loechelt, G. H., Cave, N. G., and Menendez, J., 1999, "Polarized Off-Axis Raman Spectroscopy: A Technique for Measuring Stress Tensors in Semiconductors," J. Appl. Phys., **86**(11), pp. 6164–6180.

[20] Anastassakis, E., Cantarero, A., and Cardona, M., 1990, "Piezo-Raman Measurements and Anharmonic Parameters in Silicon and Diamond," Phys. Rev. B, **41**(11), pp. 7529–7535.

[21] Ugural, A. C., 1981, *Stresses in Plates and Shells*, McGraw-Hill, New York.

[22] Timoshenko, S., and Woinowsky-Krieger, S., 1959, *Theory of Plates and Shells*, McGraw-Hill, New York.

[23] Blum, J., and Roseblatt, J., 1972, *Probability and Statistics*, Saunders, Philadelphia, PA.

[24] Chandrasekhar, M., Renucci, J., and Cardona, M., 1978, "Effects of Interband Excitations on Raman Phonons in Heavily Doped N-Si," Phys. Rev. B, **17**, pp. 1623–1633.

The Effect of Molecular Mass Distribution on Time-Dependent Behavior of Polyamides[1]

I. Emri
Center for Experimental Mechanics,
University of Ljubljana,
Cesta na Brdo 49,
Ljubljana, Slovenia
e-mail: ie@fs.uni-lj.si

B. S. von Bernstorff
BASF Aktiengesellschaft,
Ludwigshafen, Germany
e-mail: bernd-steffen.bernstorff-von@basf-ag.de

In this paper we show that the time-dependent properties of polyamides may be significantly modified by altering the "initial kinetics" of the material through the modification of the molecular mass distribution. In our investigations we have used polyamide 6 (PA6) as a testing material. We have shown that the molecular weight distribution determines the time scale of material structure formation and, hence, its time-dependent properties. The melting temperature of the bimodal molecular mass distribution of polyamide 6 material is shifted for about 30°C towards higher temperatures in comparison to the conventional monomodal molecular mass distribution of PA6. Modification of the molecular mass distribution also improves the time and temperature dependence of the mechanical properties of PA6. It seems that understanding the effect of the molecular weight distribution on the process of material structure formation may well become a new approach in development of new generation polymeric materials.
[DOI: 10.1115/1.2173008]

1 Introduction

Polyamides are one of the most important structural materials for engineering applications ranging from injection molded automotive parts, extruded foils for packaging, to melt-spun fibers for textiles. In these applications their time-dependent mechanical properties determine the functionality of the produced parts, and it is therefore important to understand how these properties may be modified and improved.

Macroscopic properties of materials depend on their inherent structures. Thus, by modifying the structure of materials one may alter their macroscopic properties. Vivid demonstration of this are different structural arrays of carbon atoms, which may appear on macro scale as graphite, diamonds,[2] fullerenes, or carbon nanotubes [1–3]. Each of these structural arrays exhibit orders of magnitude different physical properties. The "building blocks" are atoms and modification of the final structure depends on the processing history of the C atoms.

For polymers structure formation evolves over several steps, each beginning with different size and complexity of their building blocks. Structure formation takes place at atomistic scales forming higher order structure up to macroscopic dimensions and it evolves at different rates, i.e., at different time-space scales, usually in three different steps, as it is schematically shown in Fig. 1.

In the case of polyamides the initial building blocks are carbon (C), hydrogen (H), nitrogen (N), and oxygen (O) atoms. In the first processing step they are assembled into monomers, e.g., caprolactam for polyamide 6. The monomers are then becoming the building blocks for the polymerization, the next step of polymer structure formation. In these first two steps the topology of individual molecules, their masses, and mass distribution are being determined [4,5]. These molecular characteristics, along with the inherent chemical potentials, define the so-called "initial kinetics" of macromolecules [6], which is the readiness of molecules to move, to interact, and to form structures. Macromolecules are the building blocks of the last step of material structure formation, the melt processing in injection molding to form bulky parts, or extrusion of foils or fibers. The chain lengths (masses) of interacting macromolecules define the time-space scale at which the process of their interactions take place [6,7], and, thus, affect the structure formation of polymers. Therefore, by modifying the chain length or chain length distribution of macromolecules one may modify the time-space scale of molecular interactions and consequently the final material structure.

The goal of this work was to investigate the possibility to modify physical properties of a given polymer without altering the processing conditions when forming the final products. In this paper we show that the time-dependent properties of polyamide 6 (PA6) may be significantly improved by altering their molecular mass distribution. Material time-dependent properties were monitored through the shear relaxation modulus, $G(t)$, and the corresponding relaxation spectrum, $H(\tau)$. In addition, we have used optical microscopy to analyze the changes of the structure, and differential scanning calorimetry (DSC) to monitor differences of thermal properties.

Thermomechanical boundary conditions of the processing step significantly influence the process of polymers structure formation, and their final macroscopical properties [8–15]. We have therefore developed a special specimen preparation procedure to exclude the effect of variations in boundary conditions on the end structure of the material.

2 Modification of the Initial Kinetics of Polyamide PA6

In our approach we have changed the initial kinetics of PA6 by modifying its molecular mass distribution. Modifying the molecular mass was achieved by varying the average chain length of the PA6 polymer chains. This approach is known for some time, how-

[1]This paper is dedicated to Professor Dr. Wolfgang G. Knauss of California Institute of Technology, at the occasion of his 70th birthday.

[2]Artificial diamonds have been researched since the early 1950s. The first man involved was American physicist Percy Williams Bridgman. His research included extensive studies of materials subjected to high pressure. He won a Nobel Prize in physics in 1946 for his achievements but never actually created synthetic diamonds. However, on December 8, 1954, a scientist from General Electric subjected black carbon powder to pressures of 50,000 atmosphere for 16 h and made two small synthetic diamonds.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received February 11, 2005; final manuscript received January 4, 2006. Review conducted by R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of applied Mechanics, Department of Mechanical and Environmental Engineering, University of California – Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

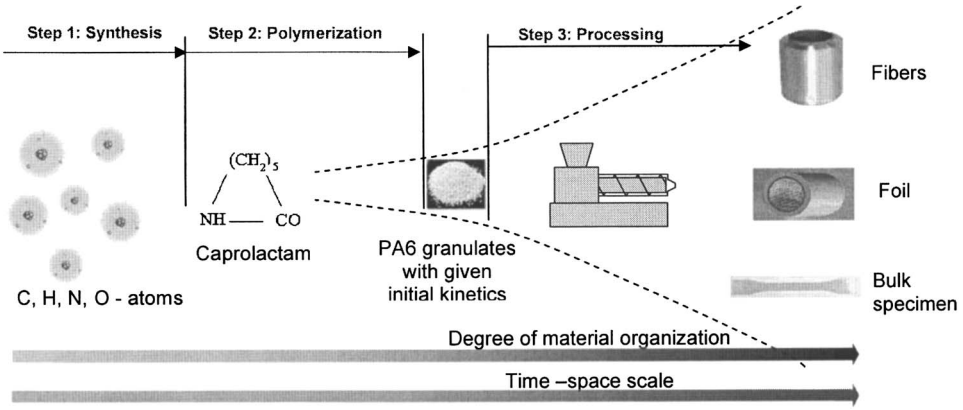


Fig. 1 Different phases of polymer structure formation

ever, it was used for those materials only that do not tend to reamidate, such as polyamides. In our investigations we have done just that. Starting from the same PA6 monomer units we produced two different PA6 materials, with a monomodal and bimodal molecular mass distributions. Both distributions are presented in Fig. 2. The monomodal distributed polyamide is a standard commercial PA6 material BS400N of BASF [16]. Details on the novel procedure of preparing the bimodal PA6 material are described in our patents [17–21].

In melt processing technologies processing parameters are commonly related to the relative viscosity (RV) of the material, which is a function of its average molecular mass. The molecular mass distribution of the bimodal material was therefore tuned such that its relative viscosity, RV=2.44, became almost equal to the relative viscosity of the standard BS400N with RV=2.4. Modification of the molecular mass distribution will, thus, affect the time scale at which molecular interactions take place during the structure formation [7], and ultimately results in different material end structures.

3 Analysis of the Time Dependency of Polymers

The mechanical behavior of polymeric materials in solid state is generally characterized in terms of their time-dependent properties in shear or in simple tension. We will use the strain-response functions exemplified by the (shear) modulus because in this case the viscoelastic and the viscous contributions are not separated from one another [22]. The shear-relaxation modulus is given by

$$G(t) = \{G_e\} + \int_{-\infty}^{\infty} H(\tau)\exp(t/\tau)d\ln\tau$$

$$= G_g - \int_{-\infty}^{\infty} H(\tau)[1-\exp(t/\tau)]d\ln\tau \quad (1)$$

where the viscoelastic and the viscous component are combined into a single expression, $G_e=\lim_{t\to\infty}G(t)$ being the equilibrium at long times, and $G_g=\lim_{t\to 0}G(t)$ the glassy modulus, which describes the behavior at short times, $H(\tau)$ is the relaxation spectrum, while t and τ denote time, and relaxation time, respectively. The braces around G_e signify that G_e is absent if the material is rheodictic [22]. The two viscoelastic constants, G_g and $\{G_e\}$, are linked by the relation [22]

$$G_g - \{G_e\} = \int_{-\infty}^{\infty} H(\tau)d\ln\tau \quad (2)$$

This relation explains why there exist two forms in the relation (1). The relaxation spectrum $H(\tau)$ carries complete information on time dependency of viscoelastic and viscous properties of the material, which comprise the information for short- and long-time behavior. It also mirrors the time scale of individual structural rearrangements in a solid polymer exposed to an outside strain excitation. Thus, the relaxation spectra may be considered as transfer functions, representing the cumulative macroscopic response of the underlying molecular rearrangements caused by the applied excitation. Modifications in material structure should be therefore mirrored in the relaxation spectrum.

Relaxation spectrum, $H(\tau)$, cannot be measured directly but must be calculated from the relaxation measurements. This requires an inverse solution of the Eq. (1), which is a Fredholm integral equation of the first kind. The inverse solution is an ill-posed problem, which may be solved with one of the algorithms developed over the last decade [23]. We have used the Emri-Tschoegl algorithm to calculate the relaxation spectra of the two materials.

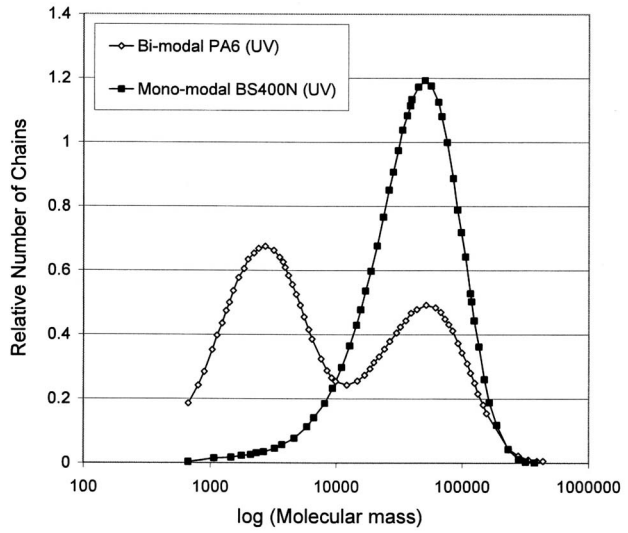


Fig. 2 Changing the initial kinetics of PA6 by modifying its molecular mass distribution, measured with ultraviolet detector, for monomodal BS400N and bimodal PA6

4 Experiment

4.1 Specimen Preparation. To analyze the effect of changing the initial kinetics of the material on its structure formation any influence from variation in boundary conditions must be excluded. Cylindrical specimens with diameter $D=3.0$ mm and length $L=56.50$ mm were cast from both materials. The casting procedure starts with filling up a glass tube with PA6 pellets, Fig. 3(*a*). Next

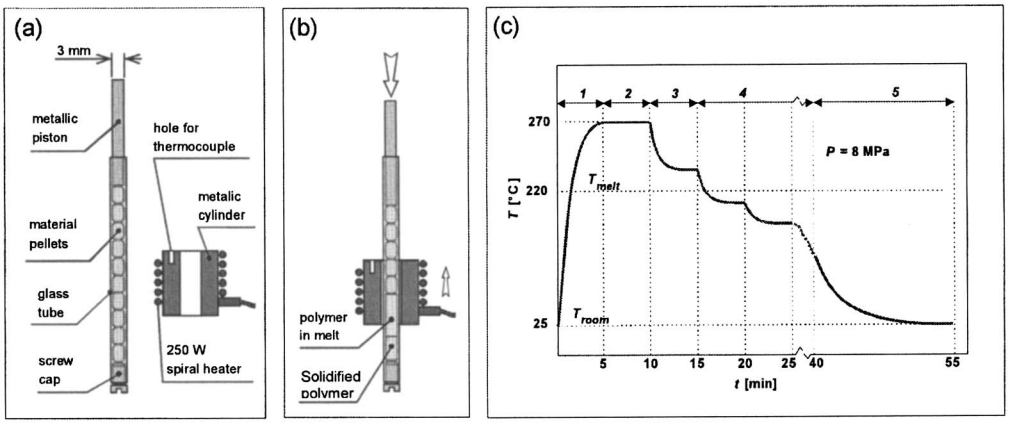


Fig. 3 Schematics of the specimen preparation

we place a spiral heater at the bottom of the tube, and simultaneously apply pressure of about 8 MPa, Fig. 3(*b*), to start the melting process, shown as step 1 in Fig. 3(*c*). In step 2 we wait for approximately 5 min, so that the pellets within the heater are completely molten. As the pellets melt, the applied pressure pushes the above unmolten pellets into the melt, thus removing any possible trapped air. In step 3 we move the heater up by 1/2 of its height, Fig. 3(*b*). Following this, steps 2 and 3 are repeated several times until all the pellets are molten. After this, the heater is removed from the glass tube and the specimen is left to naturally cool down to room temperature, step 5. The plot in Fig. 3(*c*) shows the temperature history for the first pellet segment, whereas step 4 represents the repetition of steps 2 and 3.

Figure 4 shows polarized light microscopy of the two specimens, which clearly reveal the difference in the morphology of the two materials. The bimodal material yields much finer structure than the standard PA6 material, when exposed to the same boundary (processing) conditions during the process of solidification. Obviously the molecular weight distribution affects the material structure formation in the process of solidification.

Material structure may be considered as a "frozen-in" energy, which means that the melting temperature of the bimodal material with its finer and more complex structure should occur at higher temperatures than that of less structured monomodal PA6. This is vividly presented in Fig. 5, which shows a comparison of the two DSC measurements. Measurements were performed on a standard Perkin-Elmer DSC7 apparatus using the heating rate of 20°C/min.

By increasing the temperature of a material we essentially increase the kinetic energy of individual molecular segments [7]. Since the monomer units of the monomodal and the bimodal material are chemically identical, the inertia of the individual monomer units is the same for both materials. As a consequence the local structural disintegrations should begin for both materials at the same temperature. This occurs at about 140°C, as seen in Fig. 5. However, the bimodal polyamide 6 will require more energy to "undo" its more complex structure. As a consequence the heat flow peak of the bimodal material is shifted towards higher temperatures for about 30°C, see Fig. 5.

For engineering applications it is of great importance to control the time-dependent material properties, as they affect the durability of structure made from this material. The effect of bimodality on the material inherent structure should then also be detectable in the material time-dependent mechanical properties.

4.2 Measurements of $G(t)$. Measurements of $G(t)$ were performed on a special apparatus, which is shown schematically in Fig. 6. The apparatus allows measurements of the shear relaxation modulus at wide ranges of pressures and temperatures.

The pressure is generated by the pressurizing system using silicone oil. The pressure vessel is contained within the thermal bath, where silicone oil is used for precise temperature control. Specimens can be simultaneously subjected to pressures of up to 600 MPa with a precision of ±0.1 MPa, and to temperatures ranging from −50 to +120°C with a precision of ±0.01°C. The apparatus utilizes different measuring inserts, which can be inserted into the pressure vessel. Results presented here were obtained with the relaxometer insert, shown in Fig. 7. The relaxometer measures the shear relaxation modulus by applying a constant

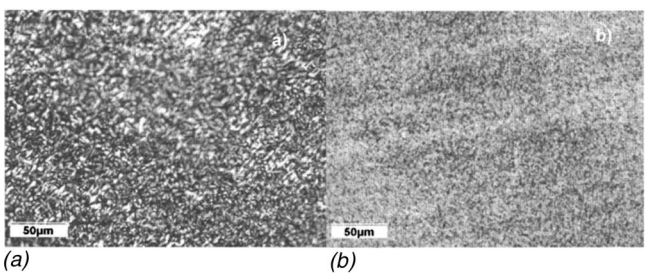


Fig. 4 Comparison of the morphology of the two materials, using polarized microscopy: (*a*) BS400N and (*b*) bimodal PA6

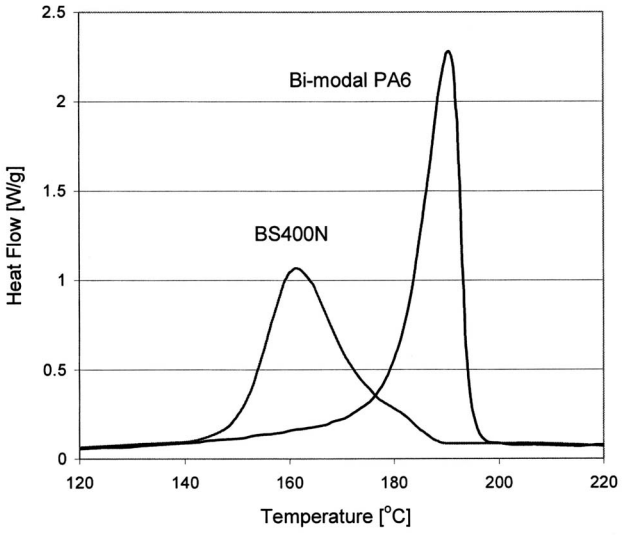


Fig. 5 Comparison of the melting temperature of BS400N, and bimodal PA6 material

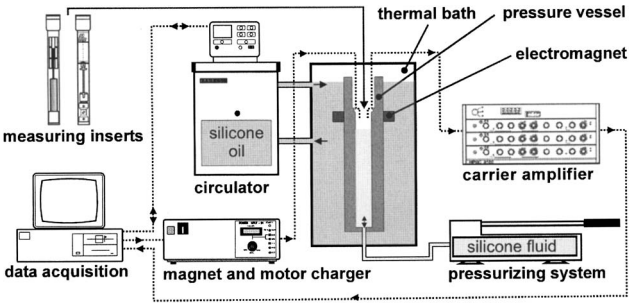


Fig. 6 Schematic of the CEM measuring system

torsional strain to a cylindrical specimen, and by monitoring the induced moment as a function of time. The specimen diameter can range from 2 up to 10 mm, while its length may vary from 52 up to 58 mm. Detailed description of the apparatus is presented elsewhere [24,25].

The temperature and pressure history for the $G(t)$ measurements are schematically shown in Fig. 8. The measurement process begins by first heating the specimen from room temperature (25°C) to 95°C and holding it there for 6 h. Then over a period of 7 h the specimen is slowly cooled down to 20°C, where it is equilibrated for another hour. This concludes the annealing process. Following this the first 630 s long $G(t)$ measurement is performed. After the measurement, the temperature is quickly raised to the next temperature of 30°C, and equilibrated for 2 h prior to taking the $G(t)$ segments measurement again. This procedure is repeated several times, until $G(t)$ at the highest temperature, 95°C, is measured. During the entire experiment, the pressure is kept constant at atmospheric pressure.

According to this experimental procedure the $G(t)$ segments for mono- and bimodal material were measured at seven different temperatures, which are indicated at the top of each diagram. The results are shown in Figs. 9 and 10.

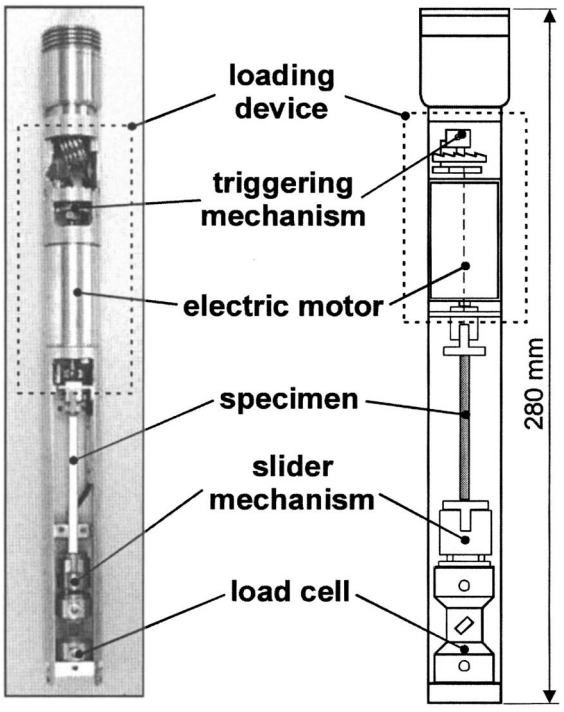


Fig. 7 Schematic of the relaxometer inserts

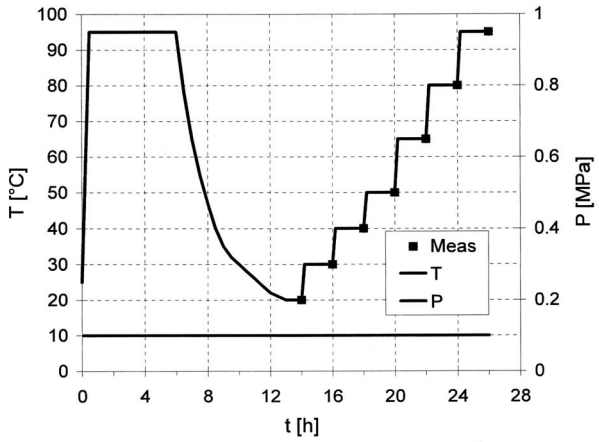


Fig. 8 Temperature and pressure history of $G(t)$

5 Results and Discussion

Taking the $G(t)$ values from the isothermal segments, shown in Figs. 9 and 10, at a selected times one can construct isochronal diagrams showing the shear relaxation modulus as a function of

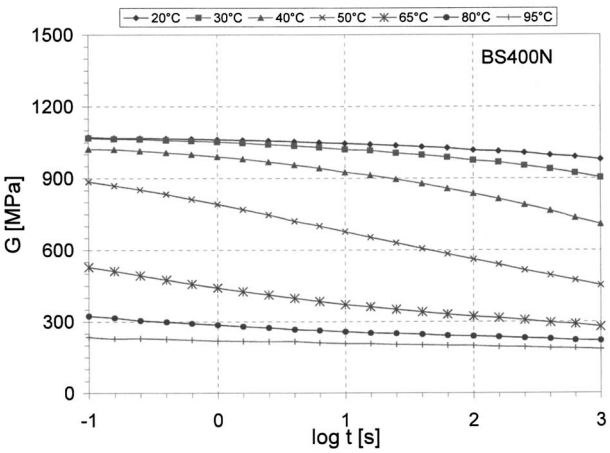


Fig. 9 $G(t)$ segments of BS400N measured at various temperatures

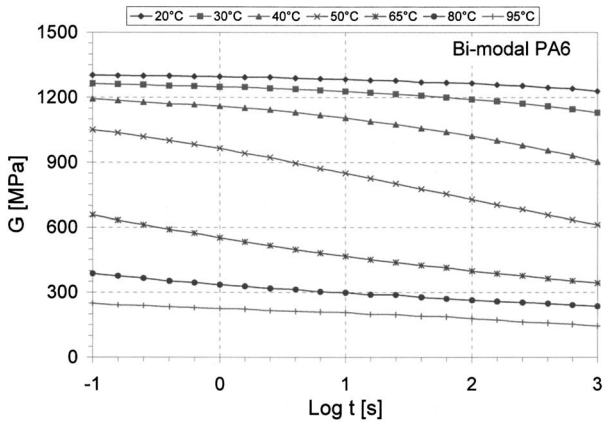


Fig. 10 $G(t)$ segments of bimodal PA material measured at various temperatures

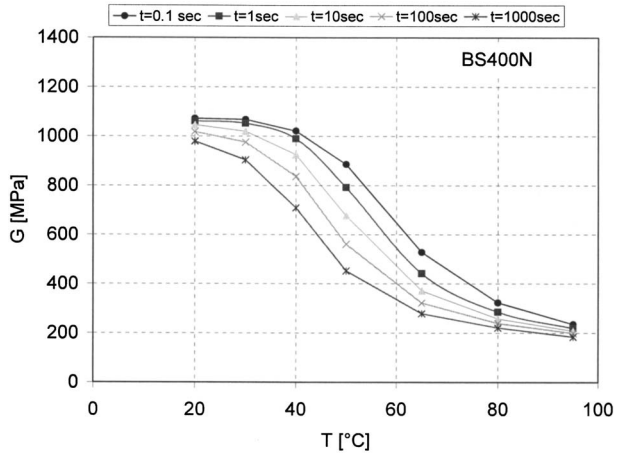


Fig. 11 Shear relaxation modulus of monomodal PA6 as a function of temperature, $G(T)$, at a selected times

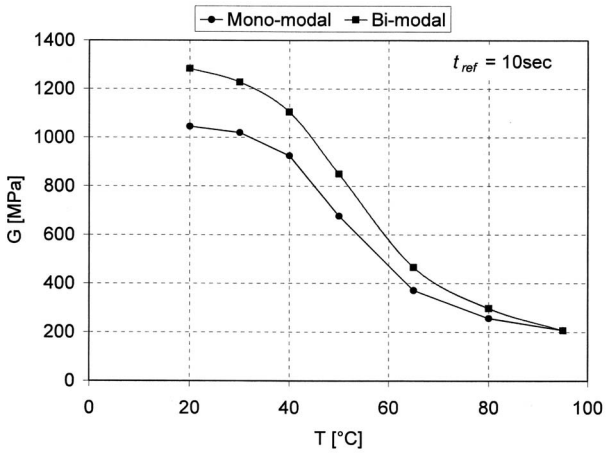


Fig. 13 Shear relaxation modulus, $G(T)$, of mono- and bimodal PA6 as a function of temperature, at $t=10$ s

temperature, $G(T)$. These isochronal relations for five selected times are shown for monomodal material on Fig. 11, and for bimodal material on Fig. 12.

The two isochronal diagrams clearly demonstrate that bi-modal PA6 exhibits different temperature dependence than the monomodal PA6, as shown in Fig. 13.

From the measured segments of $G(t)$, shown in Figs. 11 and 12, we can also construct the corresponding master curves by using the time-temperature superposition principle [22]. As a reference we selected the segment of the relaxation curve measured at 50°C (reference temperature). The master curves for mono- and bimodal PA6 are shown in Fig. 14. The individual segments superpose well, indicating that the $t-T$ superposition principle seems to hold well.

Comparison of the two relaxation curves reveals the difference of almost two decades in time-dependency between the mono- and the bimodal PA6 materials. Thus, modifying the molecular weight distribution affects the time scale of molecular rearrangements in the solid state, which is the result of altered process of molecular interactions during the process of solidification.

This modification of the material time dependency is revealed in the mechanical spectrum, as shown in Fig. 15. The spectrum is slightly truncated, which is due to the truncation error in solving the inverse problem [23]. The black curve represents the spectrum of the monomodal BS400N, while the red curve the spectrum of the bimodal PA6. The difference is obvious. The spectrum of the bimodal material clearly demonstrates presence of the intensified molecular mobility at the short times, which is a result of the increased number of short chains in the bimodal material, see Fig. 6. Similar increased molecular activity is observed at the long response times, which is related to the increased quantity of long chains in the bimodal material. Modification of the shape of the

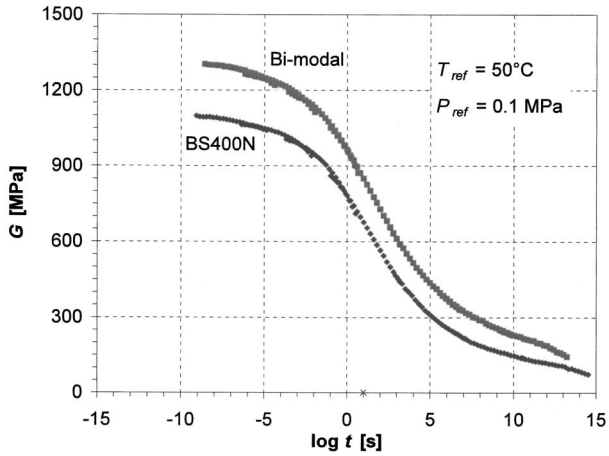


Fig. 14 Shear relaxation properties of mono- and bimodal PA

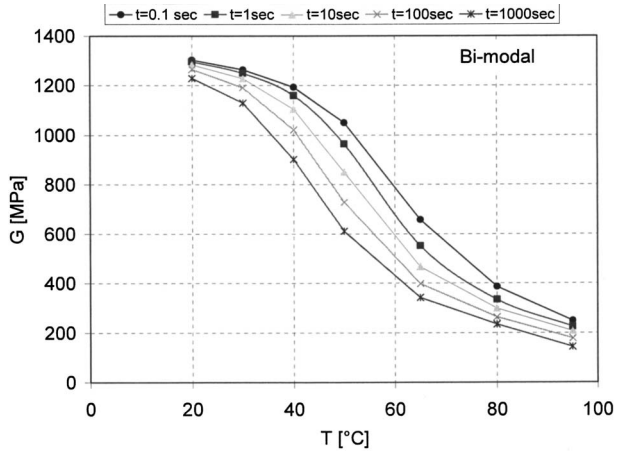


Fig. 12 Shear relaxation modulus of bimodal PA6 as a function of temperature, $G(T)$, at a selected times

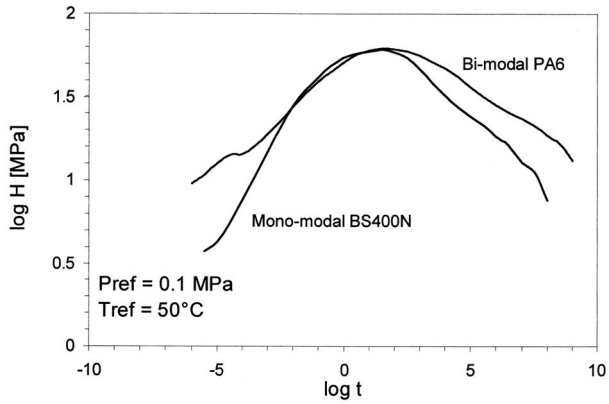


Fig. 15 Comparison of the mechanical spectra of the mono- and bimodal PA6

spectrum indicates that changes of the molecular weight distribution causes nonlinear changes in the material structure. Thus, modification of the molecular weight distribution must affect the time scale of material structure formation. The effect seems to be similar to that of adding to the polymer a small amount of nanoparticles [26].

6 Conclusions

The macroscopic properties of polyamides depend on their inherent structure which is being formed in the process of solidification. The structure is formed at atomistic as well as molecular level. The latter may be strongly affected by the modification of the molecular mass distribution, which was prepared with a novel patented procedure. The investigation showed that macroscopic properties of polyamides are very sensitive to modifications in molecular weight distribution. The observed effects are:

- The multimodal material yields much finer structure than the standard PA6 material, when exposed to the same boundary (processing) conditions in the process of solidification.
- Modification of the molecular weight distribution affects the time scale of molecular rearrangements in the solid state, which is a result of altered process of molecular interactions in the process of solidification.
- Modification of the molecular weight distribution seems to affect the time scale of material structure formation resulting in the modified final properties.
- The melting temperature of the bimodal material is shifted towards higher temperatures for about 30°C.
- Modification of the molecular mass distribution modifies the temperature dependence of the mechanical properties.
- Modification of the molecular weight distribution modifies the time dependency of PA in solid state of almost two decades.

The results shown indicate that understanding the effect of the molecular weight distribution on the process of material structure formation may well become a new approach in development of the new generation functional polymeric materials.

Acknowledgment

The authors would like to acknowledge the assistance Robert Cvelbar in preparing the figures, Klavdija Kitek in performing the microscopy and DSC measurements, and Marko Gerkman in executing the $G(t)$-measurements.

References

[1] Bridgman, P. W., 1946, "General Survey of Certain Results in the Field of High-Pressure Physics," Nobel Lecture, December 11, 1946.

[2] Smalley, R. E., 1996, "Discovering the Fullerenes," Nobel Lecture, December 7, 1996.

[3] Iijima, S., 1991, "Helical Microtubules of Graphitic Carbon," Nature (London), **354**, pp. 56–58.

[4] Fréchet, J. M. J., 1994, "Functional Polymers and Dendrimers: Reactivity, Molecular Architecture, and Interfacial Energy," Science, **263**, pp. 1710–1715.

[5] Aharoni, S. M., 1997, *n-Nylons: Their Synthesis, Structure and Properties*, Wiley, New York, and references therein.

[6] Doi, M., and Edwards, S. F., 1986, *The Theory of Polymer Dynamics*, Oxford University Press, Oxford.

[7] Balabaev, N. K., Darinskii, A. A., Neelov, I. M., Lukasheva, N. V., and Emri, I., 2002, "Molecular Dynamics Simulation of a Two-dimensional Polymer Melt," Polym. Sci., Ser. A Ser. B, **44**, pp. 781–790.

[8] Salem, D. R., 2001, *Structure Formation in Polymer Fibers*, Carl Hanser Verlag, pp. 521–552, and references therein.

[9] Hearle, J. W. S., 2001. "Fiber Formation and the Science of Complexity," Salem, D. R., *Structure Formation in Polymer Fibers*, Carl Hanser Verlag, pp. 521–552.

[10] Swartjes, F. H. M., Peters, G. W. M., Rastogi, S., and Meijer, H. E. H., 2003, "Stress Induced Crystallization in Elongational Flow," Int. Polym. Process., **13**, pp. 53–66; http://www.mate.tue.nl/mate/pdfs/2723.pdf

[11] Swartjes, F. H. M., 2001, "Flow Induced Crystallization in Elongation Flow," Ph.D. thesis, Eindhoven University of Technology, and references therein; http://www.mate.tue.nl

[12] Kornfield, J. A., Kumaraswamy, G., and Verma, R. K., 1999, "Novel Flow Apparatus for Investigating Shear-Enhanced Crystallization and Structure Development in Semicrystalline Polymers," Rev. Sci. Instrum., **70**, pp. 2097–2104.

[13] Eder, G., and Janeschitz-Kriegl, H., 1997, "Structure Development During Processing: Crystallization," Mater. Sci. Technol., **18**, pp. 269–342.

[14] Eder, G., Janeschitz-Kriegl, H., and Liedauer, S., 1990, "Crystallization Processes in Quiescent and Moving Polymer Melts under Heat Transfer Conditions," Prog. Polym. Sci., **15**, pp. 629–714.

[15] Greer, A. L., 2000, "Too Hot to Melt," Nature (London), **404**, pp. 134–137.

[16] BASF Datenblatt Ultramid BS400N, BASF Aktiengesellschaft, Ludwigshafen, Germany.

[17] Emri, I., von Bernstorff, B. S., Rauschenberger, V., and Horn, H. C., 2004, "Polymer Mixture Having Improved Rheological Properties and Improved Shrinking Behavior," Patent Appl. Pub.: US 2004192855, WO 03000796, EP 1401957, DE 10129522, CA 2449895, HU 0401480.

[18] Emri, I., von Bernstorff, B. S., Rauschenberger, V., and Horn, H. C., 2004, "Multimodal Polyamides, Polyesters, and Polyester Amides," Patent Appl. Pub.: US 2004152847, WO 03000772, EP 1401917, DE 10129525, CA 2449893, BR 0210573.

[19] Emri, I., von Bernstorff, B. S., Rauschenberger, V., and Horn, H. C., 2003, "Multi-Modal Polymer Blends," Patent Appl. Pub.: DE 10129524, WO 03006544.

[20] Emri, I., von Bernstorff, B. S., Rauschenberger, V., and Horn, H. C., 2003, "Thermoplastic Polymer Mixtures," Patent Appl. Pub.: DE 10129523, WO 03000786.

[21] Emri, I., von Bernstorff, B. S., Rauschenberger, V., Horn, H. C., Kissel, J., and Molinar, F., 2003, "Thermoplastic Polymer Mixtures," Patent Appl. Pub.: DE 10129532, WO 03000767.

[22] Tschoegl, N. W., 1989, *The Phenomenological Theory of Linear Viscoelastic Behavior*, Springer-Verlag, Berlin.

[23] Selected papers related to this subject: Cost, T. L., and Becker, E. B., 1970, "A Multidata Method of Approximate Laplace Transform Inversion," Int. J. Numer. Methods Eng., **2**, pp. 207–219; Honerkamp, J., 1989, "Ill-Posed Problems in Rheology," Rheol. Acta, **28**, pp. 363–371; Baumgaertl, M., and Winter, H. H., Rheol. Acta, **28**, pp. 511–519; Emri, I., and Tschoegl, N. W., Rheol. Acta, **32**, pp. 311–321; Tschoegl, N. W., and Emri, I., Rheol. Acta, **32**, pp. 322–327; Tschoegl, N. W., and Emri, I., 1992, "Generating Line Spectra From Experimental Responses. 3. Interconversion Between Relaxation and Retardation Behavior," Int. J. Polym. Mater., **18**, pp. 117–127; Emri, I., and Tschoegl, N. W., 1994, "Generating Line Spectra From Experimental Responses. 4. Application To Experimental-Data," Rheol. Acta, **33**, pp. 60–70; Emri, I., and Tschoegl, N. W., Rheol. Acta **36**, pp. 303–306; 1998, "Conversion of Line Spectra Into Continuous Spectral Functions," Polimeri (Zagreb), **19**, pp. 79–85; Winter, H. H., 1997, "Analysis of Dynamic Mechanical Data: Inversion Into a Relaxation Time Spectrum and Consistency Check," J. Non-Newtonian Fluid Mech., **68**, pp. 225–239.

[24] Kralj, A., Prodan, T., and Emri, I., 2001, "An Apparatus for Measuring the Effect of Pressure on the Time-Dependent Properties of Polymers," J. Rheol., **45**, pp. 929–943.

[25] Emri, I., and Prodan, T., "A Measuring System for Characterization of Polymers at Elevated Pressures and Different Temperatures," Exp. Mech. (submitted).

[26] Ulčnik-Krump, M., and Emri, I., 2004, "Structural Morphology and Relaxation Spectra of Polymer Nano-Composites," *Proceedings of the 2004 SEM Conference*.

A Hybrid Numerical-Analytical Method for Modeling the Viscoelastic Properties of Polymer Nanocomposites

Hua Liu
Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208

L. Catherine Brinson[1]
Department of Materials Science
and Engineering,
Northwestern University,
Evanston, IL 60208
e-mail: cbrinson@northwestern.edu

In this paper, we present a novel hybrid numerical-analytical modeling method that is capable of predicting viscoelastic behavior of multiphase polymer nanocomposites, in which the nanoscopic fillers can assume complex configurations. By combining the finite element technique and a micromechanical approach (particularly, the Mori-Tanaka method) with local phase properties, this method operates at low computational cost and effectively accounts for the influence of the interphase as well as in situ nanoparticle morphology. A few examples using this approach to model the viscoelastic response of nanotube and nanoplatelet polymer nanocomposite are presented. This method can also be adapted for modeling other behaviors of polymer nanocomposites, including thermal and electrical properties. It is potentially useful in the prediction of behaviors of other types of nanocomposites, such as metal and ceramic matrix nanocomposites.
[DOI: 10.1115/1.2204961]

1 Introduction

In the past five years, polymer nanocomposites have become one of the most vibrant research areas in nanotechnology development. Synthesized by incorporating nanoscale particles, such as nanotubes and nanoplatelets (mainly clay platelets and graphite nanoplatelet) [1], into polymers, polymer nanocomposites aim at exploiting the exotic physical and chemical properties of these nanoscale structures. Recent experimental results demonstrate that compared to their parent polymer systems, polymer nanocomposites can exhibit significant improvement in mechanical and physical properties, such as dramatically increased strength, modulus, and dimensional stability, decreased permeability to small molecules, increased thermal stability and heat distortion temperature, increased chemical resistance, and increased electrical conductivity and optical clarity [2–20]. These superior properties are achieved at very low loading levels of nanoinclusions, without sacrificing many advantages of the parent polymer, such as their low density and high processibility. Moreover, like conventional polymer composites, the properties of polymer nanocomposites can be tailored by proper control of the morphology [2,21]. These extraordinary behaviors combined make nanocomposites a multifunctional material and open opportunities for their applications in many fields, including the aeronautical and aerospace industry, automobile industry, civil infrastructures, and medical devices.

Substantial effort has been devoted to understanding the underlying mechanisms of morphology control and morphology-property relationship of polymer nanocomposites. Although the nanoinclusions themselves are certainly major contributors to the dramatic changes in behavior of polymer nanocomposites compared to the host polymer, some responses, particularly in the viscoelastic properties, cannot be attributed directly to the nanoinclusions. Investigations of viscoelastic properties (Refs. [2,22], and the references therein) illustrate a few non-material-specific patterns,

- the transition zones of storage and loss modulus curves are broadened or shifted on the time or temperature scale;
- glass transition temperature Tg is either shifted (normally toward higher temperature) or broadened[2].

Since the properties of the nanoparticles are not time/temperature-dependent in the experimental measurement scales, these property enhancements cannot be attributed to the properties of the nanoinclusions.

Recently, many studies have directed interest to the formation and function of an interphase region in the vicinity of the nanoscale fillers [2,22,23]. Because the dimensions of nanoinclusions are comparable to the radius of gyration of the polymer chains, their presence is expected to alter the dynamics of polymer chains in the region directly surrounding them, resulting in a difference of behavior of this region from that of the bulk polymer.

From conventional micromechanics points of view (mainly shear-lag models [24]), the importance of interphase is manifested mainly by its capability of properly transferring stress from the matrix to the fillers. The interphase's capability of stresstransfer in various polymer nanocomposite systems have been carefully examined both experimentally, such as fiber fragmentation tests [25] and single fiber pullout tests [26], and theoretically, i.e., by means of molecular dynamics simulation [27–30] or by analogy to continuum-level micromechanics model, in particular, shear-lag models [31]. Despite a few results on nanotube/epoxy system providing evidence of poor load transfer capability [32,33], the majority of the results reported demonstrated a very robust load transfer capability of the interphase, with a strength in the range of 100–376 MPa [26,29,30] with the possibility of reaching as high as 500 MPa in certain systems (nanotube/polyurenthane nanocomposite), which is at least an order of magnitude higher than the interphase strength in conventional particulate and fiber com-

[1]Author to whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received February 18, 2005; final manuscript received December 5, 2005. Review conducted by K. M. Liechti. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California–Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

[2]Tg is normally designated as the temperature where the tan δ curve peaks. The Tg broadening in the present study refers to the broadening of the tan δ peak.

posites [25]. This highly efficient load transfer capability of the interphase allows significant increase in the reinforcing efficiency of the nanoparticles, pushing the property envelop of polymer nanocomposites.

In conventional short fiber and particulate composites, other than the critical role in load transfer, the interphase itself is not considered to contribute significantly to the overall properties of the composites because, normally, the interphase is only present in a amount considerably smaller than the matrix and the fillers [34,35]. In polymer nanocomposites, however, the surface area to volume/weight ratio of the nanoinclusions is several orders of magnitude higher than that of conventional fillers. Thus, the amount of the interphase generated through the interaction of polymer molecules and the nanoinclusions could be enormous, easily dwarfing the nanoparticle amount. In addition, a strong interaction between the nanoinclusions and the polymer molecules and a sufficient nanoinclusion dosage could result in the alteration of all polymer molecules in the nanocomposite, resulting in disappearance of the signature of bulk polymer in the nanocomposite behavior [2].

Many recent attempts to model the behavior of polymer nanocomposites have extended modeling techniques originally developed for high polymer sciences and conventional fiber and particulate composites. These techniques can be divided into computational chemistry approaches and continuum-theory-based approaches.

Computational chemistry approaches, mainly molecular dynamics (MD) and coarse-grained molecular dynamics (CGMD), capture the structural and dynamical details at the atomic or molecular level and have been utilized to investigate important issues, such as load transfer from matrix to nanoinclusions [27–30] and changes in polymer mobility near a nanoparticle [36–38]. The direct implementation of MD and CGMD onto full-scale modeling of polymer nanocomposites, however, is prohibited due to extremely high computational cost. Typical simulations involve only a few small nanoparticles with 50–100 short polymer chains ($<$100 units long).

On the other hand, continuum-theory-based approaches easily capture full-scale nanocomposite response and allow many aspects of nanocomposites to be modeled and explored. A drawback of this approach is the required continuum assumption of the nanosized inclusions as well as the surrounding polymer, despite the discrete atomistic nature of the nanostructures; the justification and limitation of this assumption is still a subject of debate. These methods, mainly micromechanics models [13,25,39–44] and numerical methods, i.e., finite element analysis [43,45,46], have been employed to model the elastic properties of the polymer nanocomposites and excellent agreement with experimental results has been reported in many cases.

Increasingly more efforts have been devoted to the development of multiscale modeling, particularly, hierarchical multiscale modeling, which refers to the hierarchical coupling of modeling schemes at different length and time scales in order to link the localized behavior to the macroscopic properties. The subject system is "decomposed" into different regions in the sense of length scale and/or time scale. Different modeling techniques are then applied to different regions accordingly and are connected in a sequential way in which the output of one is used directly as input for another up the length/time hierarchy. Many of the recently reported modeling efforts on polymer nanocomposites, including a few that have been classed above into micromechanics approaches, fall into the category of hierarchical multiscale modeling (though it should be noted that in these studies, while the hierarchy of the length scale was maintained, the hierarchy of the time scale was not addressed).

In a recent work by Li and Chou [47], the carbon nanotubes were modeled as beam structures by a molecular structural mechanics approach [48], which is similar to quasicontinuum (QC) method[3] earlier developed to simulate deformation in solids [49]. The polymer matrix was modeled with FEM. The two parts were either connected directly in case of perfect nanotube/polymer bonding or connected by equivalent truss rod representing the van der Waals interaction in case of imperfect nanotube/polymer bonding.

Also worth noting are several works that explicitly or implicitly used the idea of "pseudoparticle" or "effective particle" [39,40,43,44]. The subject systems were treated as polymer reinforced by pseudoparticles, which represent the localized morphology. The effective properties of pseudoparticles were evaluated and subsequently used as the input for the direct implementation of continuum-theory based models on the virtual system consisting of polymer and pseudoparticles. Some of the complexities in the polymer nanocomposites, such as the clay nanoplatelet cluster (intercalated clay particles) in clay polymer nanocomposites and the wavy in situ configuration of nanotubes in nanotube-based polymer nanocomposites, have been considered in the models.

To summarize, the development of modeling methods for polymer nanocomposites is fairly vibrant. The most promising class is probably multiscale modeling where various modeling schemes at different length (and possibly, time) scales are coupled. The investigation of the elastic properties of the polymer nanocomposites using these models has been proved to be successful. However, at the present time none of the existing models have examined the influence of nanoparticles on the viscoelastic nature of the material responses, nor have they taken into consideration the interphase as a significant contributor to the overall performance of the nanocomposites rather than merely a load transfer agent. Given the importance of the interphase demonstrated experimentally, the development of a low computation cost modeling technique that specifically accounts for the interphase is thus needed to provide accurate predictions of the viscoelastic behavior of the polymer nanocomposite.

In the current paper, we introduce a hybrid numerical-analytical modeling method. It has a multiscale nature; that is, it hierarchically couples several modeling techniques and/or even the results of fine scale experimental measurements. Operating at a relatively low computational cost, it can take into consideration not only the interphase but also many other complexities in the polymer nanocomposites, such as the complicated in situ configuration of the nanoparticles and the interphase. In Secs. 2 and 3, the theoretical background of this modeling technique will be presented. A few sample applications will be provided. Results and discussions will follow.

2 Theoretical

The core of this hybrid modeling technique consists of Mori-Tanaka method and the concept of numerical dilute strain concentration tensor. The Mori-Tanaka method [50] is based on Eshelby's analysis of elastic stress field in and around an ellipsoidal particle in an infinite matrix [51,52] and is commonly regarded as one of the most powerful micromechanics models. A brief description of how to apply this method to a composite is presented here (for more detailed discussion, please refer to [53,54]).

Consider a linear elastic inclusion[4] with elastic modulus C^i embedded in an infinite linear elastic body of matrix material with

[3]Quasicontinuum (QC) method was developed based on the Cauchy-Born rule to simulate deformation in solids. The Cauchy-Born rule is a local approximation of the strain energy and gives the strain energy at a given point by using the strain energy associated with a crystal subjected to the same homogeneous deformation as exists at that point. By using this rule, the continuum stress tensor and tangent stiffness can be derived directly from the interatomic potential by differentiating the potential with respect to local deformation gradient once and twice, respectively.

[4]Note that in Eshelby's analysis, an inclusion is defined as a finite domain embedded in the matrix that has the same properties as the surrounding matrix, while an inhomogeneity is defined as a finite domain embedded in the matrix that has different properties to its surrounding matrix. The term inclusion in this article actually refers to the inhomogeneity.

elastic modulus C^m. Perfect bonding at the inclusion/matrix interface is assumed. Initially, the entire system is in a stress-free state, i.e., zero stress everywhere. When the infinite body is subjected to a uniform strain field $\bar{\varepsilon}$, the volumetric average strain inside the inclusion, $\overline{\varepsilon^i}$, which is given by

$$\overline{\varepsilon^i} = \frac{\int_{V^i} \varepsilon^i dV^i}{\int_{V^i} dV^i} \tag{1}$$

where ε^i and V^i represent the strain field and domain of the inclusion, respectively, can be related to $\bar{\varepsilon}$ with the introduction of Hill's concept of strain concentration tensor [55] as

$$\overline{\varepsilon^i} = A^i \bar{\varepsilon} \tag{2}$$

where A^i is the strain concentration tensor for the inclusion. Note that the strain concentration tensor in this dilute[5] setup is usually called the dilute strain concentration tensor and is denoted by $A^{i,\mathrm{dil}}$ in the present discussion.

This dilute setup can be employed to approximate a two-phase composite containing a number of identical inclusions that are aligned, providing the inclusions are well dispersed and the volume fraction is sufficiently low so that an individual inclusion sees the matrix as infinite. In this situation, the interaction between the inclusions is negligible and the elastic modulus of the composite is determined by

$$C = C^m + v^i (C^i - C^m) A^{i,\mathrm{dil}} \tag{3}$$

where v^i is the volume fraction of the inclusion.

To model a composite that has higher volume fraction of uniformly oriented inclusions, the interaction between the inclusions is taken into account by assuming that the individual inclusion feels the average strain in the matrix $\overline{\varepsilon^m}$ as the far field strain so that the dilute strain concentration tensor relates the average strain in the inclusion and the average strain in the matrix as $\overline{\varepsilon^i} = A^{i,\mathrm{dil}} \overline{\varepsilon^m}$. The strain concentration tensor is then given by

$$A^i = A^{i,\mathrm{dil}} [(1 - v^i) I + v^i A^{i,\mathrm{dil}}]^{-1} \tag{4}$$

where I is the identity tensor. Combining Eqs. (4) and (3), recognizing that $A^{i,\mathrm{dil}}$ in (3) must be replaced by A^i for nondilute composites, and we obtain

$$C = C^m + v^i (C^i - C^m) A^{i,\mathrm{dil}} [(1 - v^i) I + v^i A^{i,\mathrm{dil}}]^{-1} \tag{5}$$

In the case of a composite containing N types of inclusions that have unidirectional orientation, the modulus of the composite is given by

$$C = \left[\left(1 - \sum_{j=1}^{N} v_j^i \right) C^m + \sum_{j=1}^{N} v_j^i C_j^i A_j^{i,\mathrm{dil}} \right] \times \left[\left(1 - \sum_{j=1}^{N} v_j^i \right) I + \sum_{j=1}^{N} v_j^i A_j^{i,\mathrm{dil}} \right]^{-1} \tag{6}$$

where C_j^i, v_j^i, and $A_j^{i,\mathrm{dil}}$ are the stiffness, volume fraction, and dilute strain concentration tensor of type j inclusion, respectively. For composites that have multiple types of inclusions that are nonuniformly oriented, the average with respect to the orientation must be taken. Designating the orientational average operation of a tensor $\boldsymbol{t}$ by $\{t\}$, Eq. (6) is correspondingly changed into

[5]Dilute refers to the fact that there is only a single inclusion embedded in the infinite matrix.

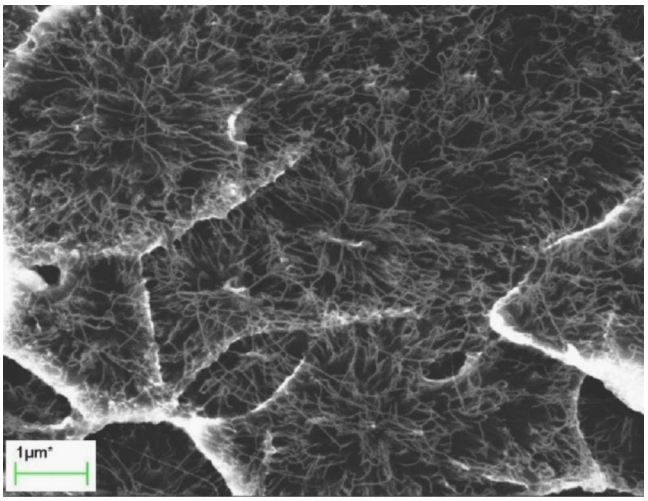


Fig. 1 SEM image of fracture surface of SWNT/PMMA nanocomposites; note that the SWNTs are highly curved in situ [2]

$$C = \left[\left(1 - \sum_{j=1}^{N} v_j^i \right) C^m + \sum_{j=1}^{N} v_j^i \{C_j^i A_j^{i,\mathrm{dil}}\} \right] \times \left[\left(1 - \sum_{j=1}^{N} v_j^i \right) I + \sum_{j=1}^{N} v_j^i \{A_j^{i,\mathrm{dil}}\} \right]^{-1} \tag{7}$$

Although this analysis is derived under the conditions that all phases comprising the composite are linear elastic, this method can also be implemented with correspondence principle [56] to the case where the $n (n \leq N)$ phases of an N-phase composite are linear viscoelastic providing the elastic counterpart has exact solution [57]. Equation (7) is modified as

$$C^* = \left[\left(1 - \sum_{j=1}^{N} v_j^i \right) C^{m^*} + \sum_{j=1}^{N} v_j^i \{C_j^{i^*} A_j^{i,\mathrm{dil}^*}\} \right] \times \left[\left(1 - \sum_{j=1}^{N} v_j^i \right) I + \sum_{j=1}^{N} v_j^i \{A_j^{i,\mathrm{dil}^*}\} \right]^{-1} \tag{8}$$

where the attachment of the superscript asterisk refers to the complex values. It is easily noted that the critical issue of this method to determine effective composite response is the determination of the dilute strain concentration tensors.

In the classical Mori-Tanaka method, the dilute strain concentration tensor is determined analytically via use of the Eshelby solution. However, while the analytical solutions of the dilute strain concentration tensors for ellipsoidal inclusions and some inclusions with regular nonellipsoidal shapes have been derived [58], to obtain analytical solutions for arbitrarily shaped inclusions is quite a challenge. Unfortunately, in most of the polymer nanocomposites, in situ configurations of the nanoparticles are fairly complicated. Scanning electron microscopy (SEM) and TEM images of the fracture surface of nanocomposites revealed that the nanotubes and nanoplatelets are typically highly curved [2,59], as shown in Fig. 1. As the interphase forms around the nanoparticles, the curved double inclusion nature precludes current analytical approaches and the direct implementation of classical Mori-Tanaka method.

In our hybrid modeling scheme, the dilute strain concentration tensors are determined numerically by employing a finite element analysis technique. A similar modeling scheme has been employed to investigate the influence of nanotube waviness in a two-phase elastic (both nanotube and matrix were treated as purely elastic) composite [60]. In our approach, a finite element configuration (unit cell), consisting of a nanoparticle and one or more

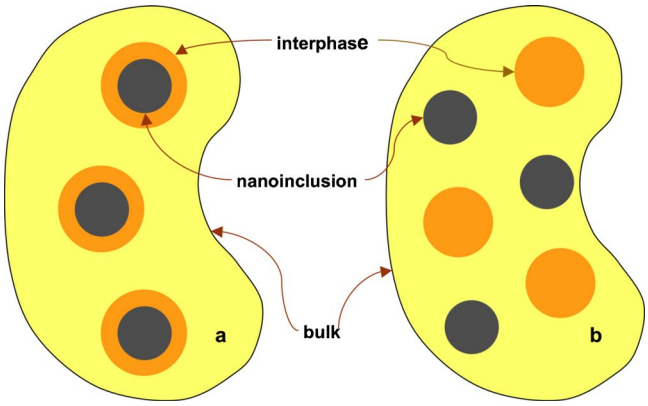


Fig. 2 Schematics of difference between the real three-phase composite (*a*) and the virtual counterpart (*b*) if classical Mori-Tanaka method is applied

layers of surrounding interphase embedded in extensive bulk matrix material, is established. The in situ configurations of the nanoparticle and the interphase are modeled explicitly, and the calculation is done in the viscoelastic domain. The unit cell must be large enough to ensure that the nanoparticle and the interphase see the matrix as infinite so that the dilute solution can be obtained. Appropriate boundary conditions are then applied to unit cell and finite element solutions of the volumetric average strains inside the interphase(s), and the nanoparticle are obtained. By Eq. (2), the numerical dilute strain concentration tensors are calculated for the nanoparticle and the interphase regions. Using these as inputs, a three-phase Mori-Tanaka method is employed to solve for the response of the polymer nanocomposite.

Although the properties of the nanoparticle and the bulk polymer may be readily available, the properties of the interphase are more of a challenge. Ideally, interphase properties can be either directly measured or calculated by implementing molecular dynamics simulation. It should be noted that both modeling and experimental methods at the molecular length scale presently contain significant assumptions. Specifically, MD simulations yield the behavior of polymer molecules in a time period of a few femtoseconds up to a few nanoseconds; in contrast, the continuum calculations require viscoelastic parameters many orders of magnitude longer in time. And limited by computational capability, in order to carry out computations in a reasonable time frame, current MD simulations can only be implemented at a reduced scale in which some structural details of the polymer molecules are often dropped. Thus at the current stage, results of MD simulations cannot directly provide realistic continuum time and length scale properties, but can only provide guidance for trends in continuum properties of the interphase compared to the continuum properties of the bulk polymer. In this paper, the continuum properties of the interphase will be assumed by a particular relationship to the bulk matrix properties, based on published results from MD simulations and experimental work in confined polymers [38,61–63]. It has been noted that attractive interactions can alter polymer dynamics substantially, with Tg changes greater than 10°C, which corresponds to a few decades of change in the characteristic relaxation time. Thus, here we assume that there is only one interphase, whose properties are uniform and can be related to those of the matrix material (bulk polymer) by a shift in the relaxation times. Our ongoing study of molecular dynamics simulation and experimental measurement of the interphase is expected to enable us to examine more accurate assessment in the future.

The ladder hierarchy of this hybrid numerical-analytical modeling technique is thus: the lowest rung is the MD simulation or experimental measurement of the interphase(s) (which, in this paper, is approximated by a relaxation time shift based on existing experimental/modeling data); the middle rung is finite element

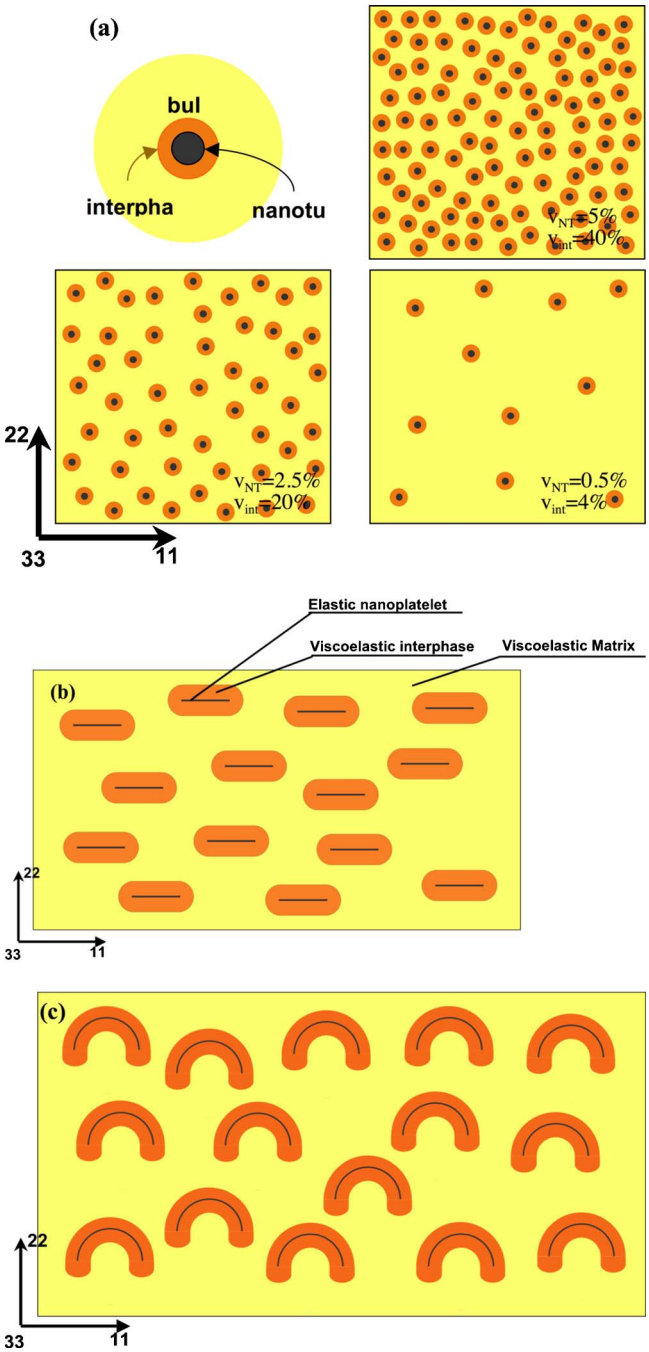


Fig. 3 Schematics of composite configurations studied in this paper: (*a*) nanotube/PC nanocomposite: nanotube fractions (v_{NT}) 2.5% and 5%; nanotube/PMMA nancomposites: v_{NT} 0.5%; interphase fractions (v_{int}) each a factor of 8 larger; (*b*) nanoplatlet/PC nanocomposite when the nanoplatelets are flat; and (*c*) nanoplatlet/PC nanocomposite when the nanoplatelets are curved. For both (*b*) and (*c*), nanoplatelet volume fraction v_{NP} is 0.5% and interphase v_{int} 24%

analysis of the unit cell; the upper rung is Mori-Tanaka method; the upper two rungs are connected with the concept of numerical dilute strain concentration tensor.

Because both the finite element analysis and the classical Mori-Tanaka method are insensitive to the length scale of the composites, this hybrid modeling method can be used in composites with any size of inclusions. However, the specific conditions consid-

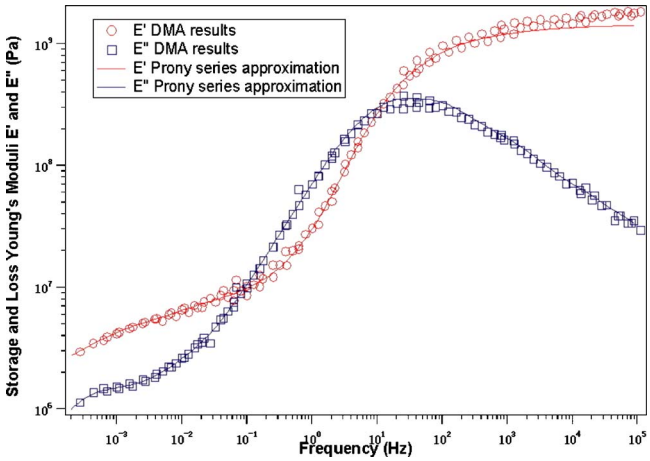


Fig. 4 Experimentally measured complex Young's modulus of PC and its Prony series approximation

ered in this paper, namely, curved particles and large interphase zone, are relevant for nanoparticle reinforced composites for the reasons mentioned earlier.

It should be further noted that the prediction made by our hybrid numerical-analytical method should be more accurate than the prediction made by the classical Mori-Tanaka method for composites with an interphase. In real nanocomposites, such as the three-phased one schematically shown in Fig. 2(*a*), the nanoparticles are encapsulated in the interphase, which is, in turn, embedded in the matrix. While this effect is taken into account in our hybrid numerical-analytical method by the way the dilute strain concentration tensors are obtained (see following section for detailed examples), a direct implementation of the classical Mori-Tanaka method with three phases, however, would correspond to the situation shown in Fig. 2(*b*), where the interphase is a separate inclusion, not encapsulating the nanoparticles. Although recent work [64,65] has provided an analytical solution for the situation of particle encapsulated in one layer (or multilayer) of coating, the particle and coating must be spherical in shape.

3 Sample Applications, Results, and Discussion

A few two-dimensional sample applications of this hybrid modeling technique on polycarbonate (PC) and poly(methyl methacrylate) (PMMA) based polymer nanocomposites are presented in this section. These computations are done in the complex viscoelastic domain in order to provide the frequency- and temperature-dependent viscoelastic properties, which characterize the relaxation of the polymer nanocomposites. Composite con-

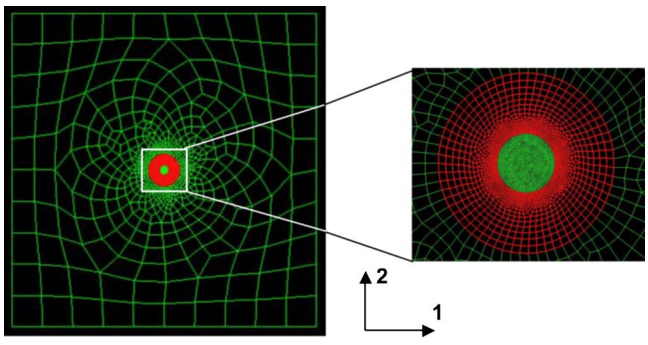


Fig. 5 Two-dimensional finite element model (mesh) for numerical dilute strain concentration tensor calculation in nanotube/PC nanocomposite (left); close up on the nanotube and interphase (right)

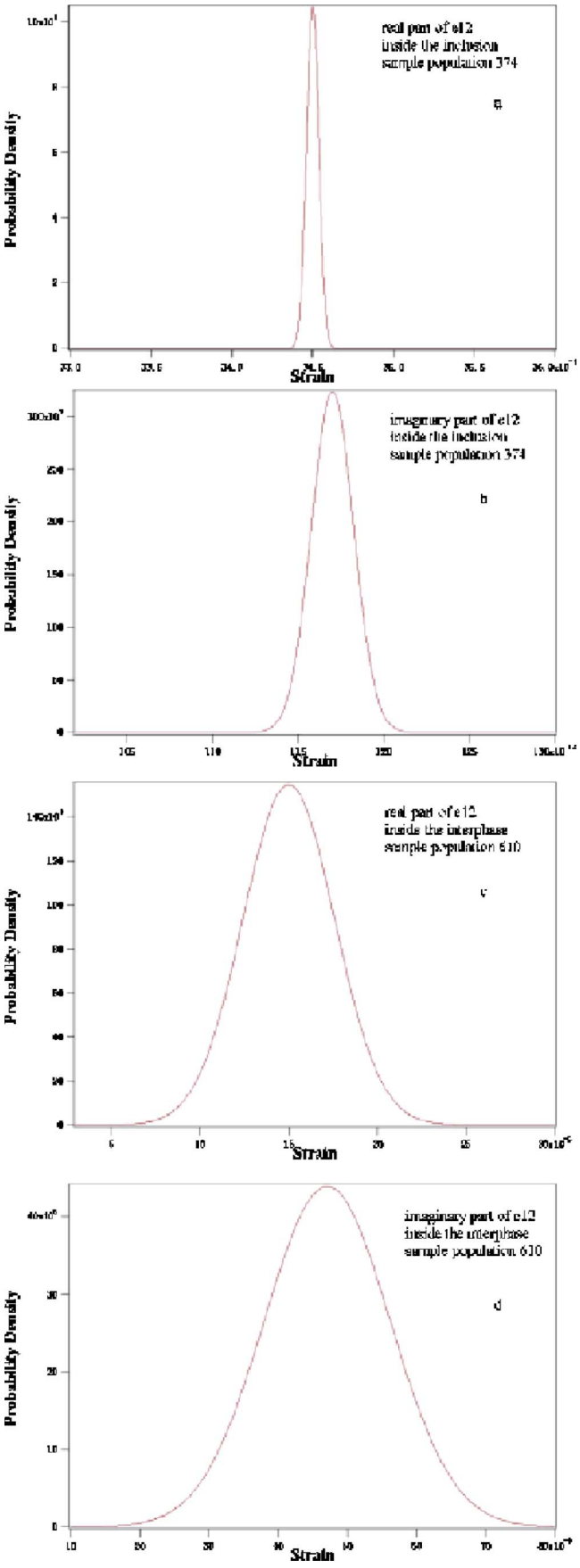


Fig. 6 Normal distribution curves of the complex shear stain fields inside the inclusion (*a*) real part; (*b*) imaginary part; and the interphase (*c*) real part; and (*d*) imaginary part

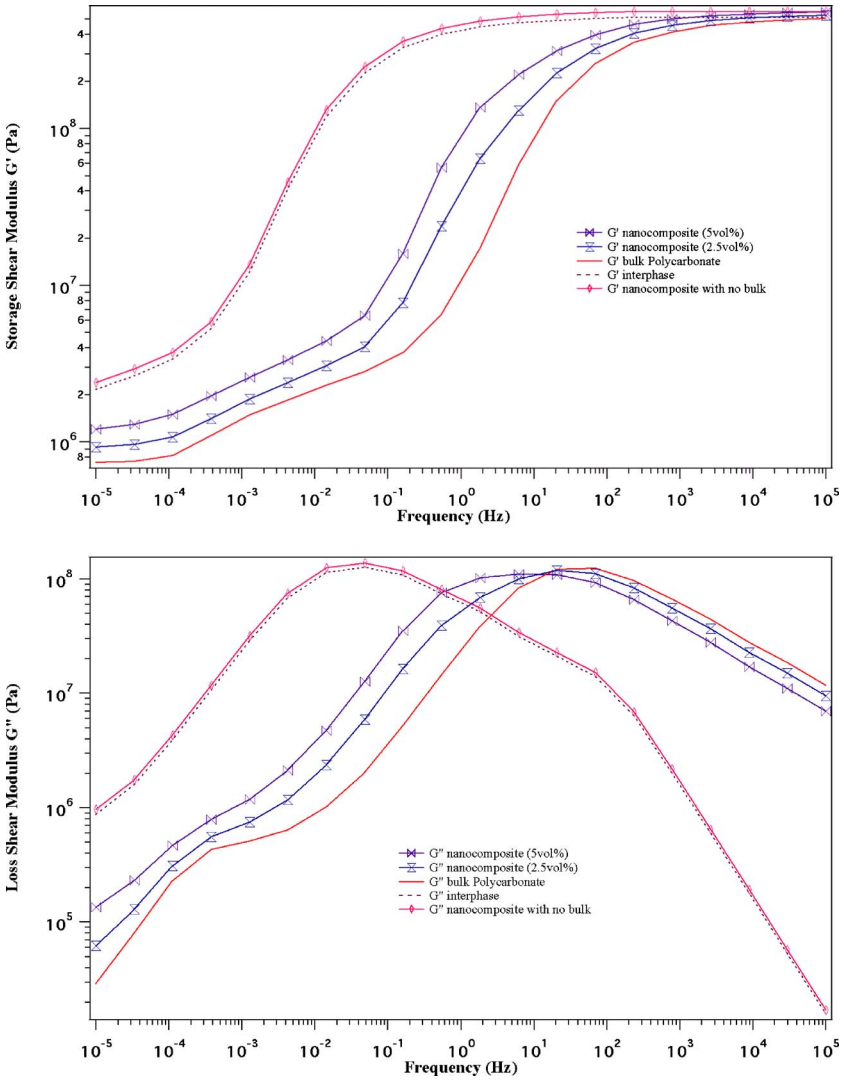


Fig. 7 Predicted storage and loss shear moduli of the nanocomposite (nanotube content 2.5 and 5 vol %) against those of bulk PC and the interphase

figurations studied include several inclusion shapes and interphase volume fractions to highlight the influence of these features. Specifically, we will consider a case of straight nanotube composite with low and moderate interphase volume fractions, straight graphite plate and curved graphite plate composites with moderate interphase volume fractions. These differences are illustrated schematically in Fig. 3.

3.1 Nanotube/PC Nanocomposites. To model nanotube/PC nanocomposites, the nanotubes were assumed to be unidirectionally aligned and ideally dispersed, as schematically shown in Fig. 3(*a*). The transverse (directions 11 and 22) complex moduli of the composite were calculated. The PC used was Lexan from GE plastics, and the frequency domain response of pure PC was obtained by frequency sweep mode dynamic mechanical analysis (DMA) (for details of the experiments, please refer to [23]). The results were manually shifted using time-temperature superposition (TTSP) to obtain the master curves at a reference temperature of 150°C. These curves were properly trimmed and approximated with a 29-term Prony Series using a linear least squares solver DYNAMFIT previously developed in this lab [66]. The comparison between the experimental results and the Prony series approximation is shown in Fig. 4.

The finite element analysis was performed using ABAQUS™, which requires the property definition of viscoelastic materials to be given in complex shear and bulk moduli. DMA results of the pure PMMA were in the format of complex Young's modulus versus frequency. Therefore, conversion from complex Young's modulus to complex shear and bulk moduli was performed: assuming that the instantaneous Poisson's ratio of PC is 0.4 and the instantaneous terms (Young's modulus, shear, and bulk moduli) relate with each other through an elastic relationship, the instantaneous bulk modulus was calculated. Since the variation of the bulk modulus with time/frequency is relatively small, the complex shear modulus at each frequency point was calculated using dynamic correspondence principle assuming a constant bulk modulus

$$G^* = \frac{3K^*E^*}{9K^* - E^*} \tag{9}$$

The calculated values for G^* were then approximated with a 29-term Prony series.

As discussed earlier, the properties of the interphase used were derived from those of the bulk PC by horizontally shifting bulk PC curves in the frequency domain, indicating the difference between the bulk and interphase lies only in the mobility of polymer chains. Such an assumption corresponds to a simple shift in the relaxation times of the polymer in the interphase zone due to interaction with the nanotube. The nanotube was modeled as solid

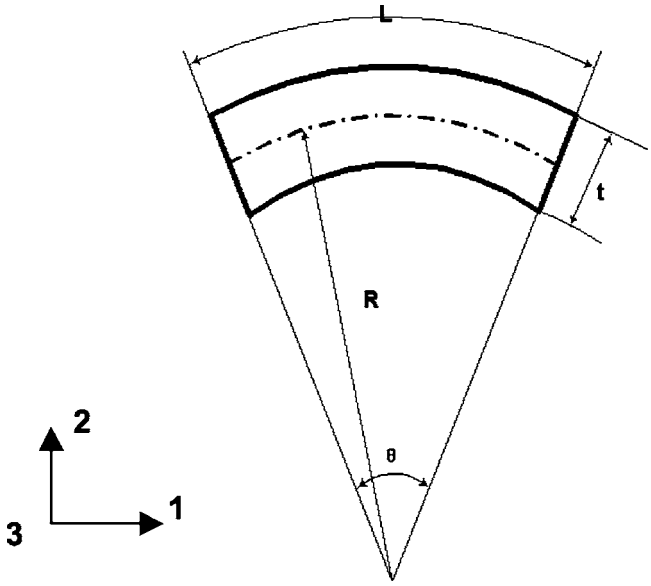


Fig. 8 Schematics of two-dimensional configuration of a nanoplatelet: length *L*, thickness *t*, and radius of curvature *R*

rod with a circular crosssection, despite its hollow nature. It was assumed to be isotropic and linear elastic with Young's modulus of 1 TPa and Poisson's ratio of 0.3.

The finite element model (mesh) that was used to retrieve the numerical dilute strain concentration tensors of the nanotube and the interphase is shown in Fig. 5. The nanotube is represented by the circle region at the center; the interphase is represented by the annulus that surrounds the nanotube with a thickness equal to the diameter of the nanotube, consistent with data in [63]; the bulk PC is represented by the square with a lateral dimension 30 times the diameter of the nanotube. The size of the model was chosen as such so that it is ensured that the matrix appears infinite to both the nanotube and the interphase. Note such a configuration also determined that the ratio of interphase volume fraction to nanotube volume fraction was set to 8. Perfect bonding at the nanotube/interphase and interphase/matrix interfaces was assumed.

Considering the two-dimensional centrosymmetric nature of our finite element models, only two strain cases, namely, one farfield[6] tensional strain ε_{11} (or ε_{22}) and one farfield shear strain ε_{12} (or ε_{21}), were needed to be applied to the models. These farfield strains were applied to the model by specifying the displacement at certain nodes. The strains were restricted to be small so that large deformation would not occur. The finite element solutions for the volumetric average strain inside the inclusions were obtained and subsequently, numerical dilute strain concentration tensors for the nanotube and the interphase were determined. The properties of the composites were then calculated.

An additional step was taken to assess the strain field distribution inside the nanotube and the interphase. The normal distribution curves of the complex strain fields inside the inclusion and the interphase in the case where the interphase was assumed to be three decades less mobile than the bulk PC, and the whole finite element configuration was subject to farfield shear strain with an amplitude of 1.5×10^{-4} and a frequency of 1×10^{5} are shown in Fig. 6. These curves indicate that the strain fields inside both the nanotube and the interphase are relatively homogeneous, which justifies the implementation of Mori-Tanaka method.

The result for the case where the volume fraction of nanotubes

[6]The strains were applied to the matrix. As the inclusions see the matrix as infinite body, they see these strains as farfield strains.

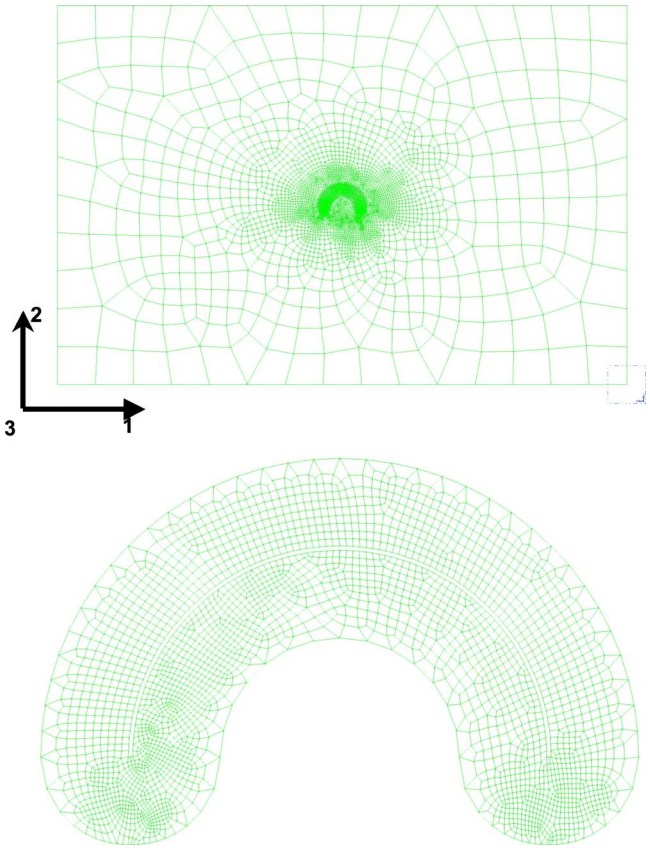


Fig. 9 Two-dimensional finite element model (mesh) for numerical dilute strain concentration tensor calculation in nanoplatelet/PC nanocomposite (top); isolated mesh of the interphase (bottom)

was 2.5 vol % and 5 vol % and the interphase was three decades less mobile than the bulk PC is shown in Fig. 7, where the predicted storage and loss shear moduli of the nanocomposite are plotted against those of the bulk PC and the interphase. It can be clearly seen in Fig. 7 that the nanocomposites show a shift of transition zone towards lower frequency, as compared to the bulk PC. The loss modulus peak, shown in Fig. 7(*b*), is remarkably broadened, and this broadening becomes more obvious as the volume fraction of nanotubes increases. These trends manifest the large influence of the interphase and are consistent with experimental observation [23]. Also included in the plots is the hybrid model prediction for the nanocomposite in which all the bulk PC has been altered to the interphase properties, which potentially happens when the nanotube-PC molecule interaction is strong and nanotube dispersion is excellent, leading to a percolating interphase, which spans through the whole composite. It can be seen that in this case, instead of broadening (of the loss peak), the transition zones are completely shifted corresponding to the interphase. Only a small difference in magnitude distinguishes the curves for the nanocomposite from those of the interphase due to the elastic stiffness of the nanotubes themselves. This prediction provides additional evidence to aid the interpretation of experimental observations of a functionalized nanotube polymer nanocomposite [2], which also demonstrates a complete shift of the loss modulus without broadening.

3.2 Nanoplatelet/PC Nanocomposites. For nanoplatelet/PC nanocomposites, the nanoplatelets were assumed to be fully exfoliated and ideally dispersed in the polymer. The hybrid modeling technique was implemented to assess the influence of nanoplatelet curvature on the frequency domain response of the nanocomposite, the investigation of which has not yet been re-

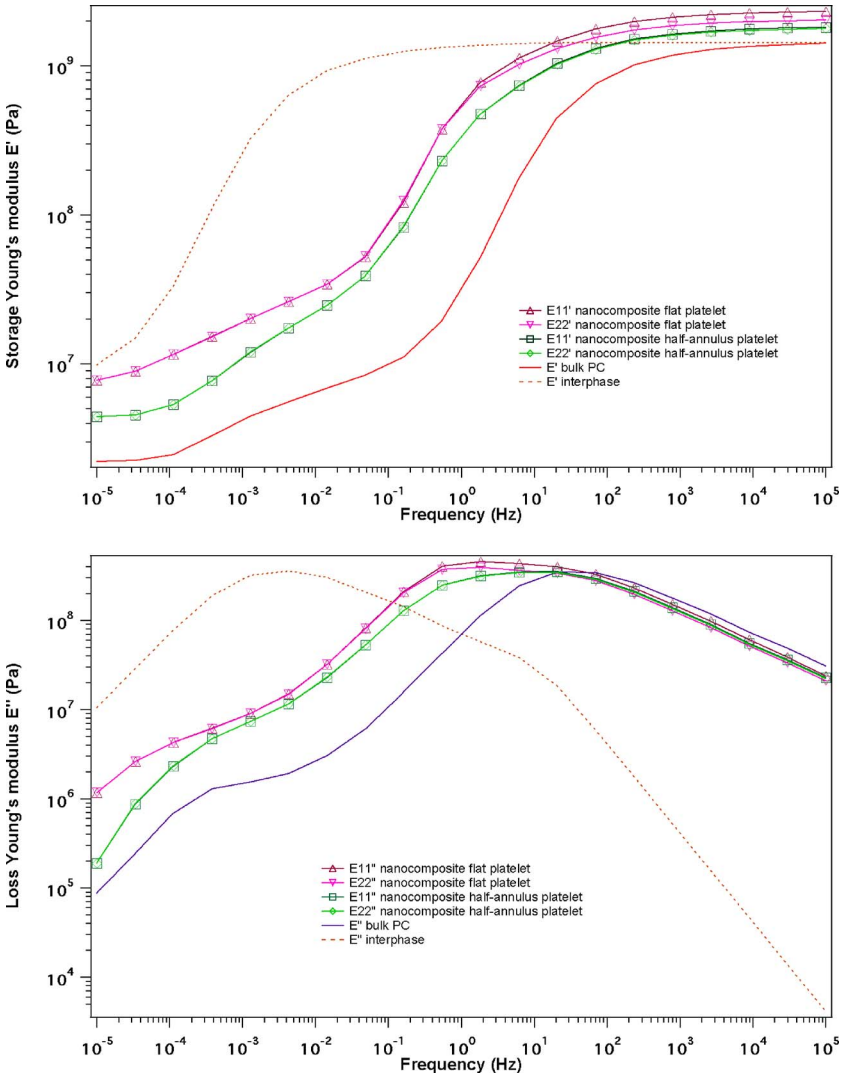


Fig. 10 Storage and loss Young's moduli of nanocomposites with flat/half-annulus nanoplatelets; nanoplatelet content 0.5 vol %

ported in the literature. The two-dimensional configuration of a curved nanoplatelet was mathematically represented by a partial annulus characterized with a length L, a thickness t (and thus a two-dimensional aspect ratio of L/t), and a radius of curvature R, as schematically shown in Fig. 8. The nanoplatelet was assumed to be isotropic and linear elastic with Young's modulus of 450 GPa and Poisson's ratio of 0.3.

The finite element model (mesh) that was used in the case where the nanoplatelet had a two-dimensional aspect ratio of 150 and a radius of curvature $R=L/\pi$, i.e., the nanoplatelet was a half annulus, is shown in Fig. 9. Thus, this calculation represents an extreme case for platelet curvature; note, however, that experimental images of nanoplates that have been exfoliated down to the individual sheet layer reveal high degrees of wrinkling and extreme curvature [59]. The interphase encapsulating the nanoplatelet was set to have a thickness 20 times the thickness of the nanoplatelet, which determined that the volume fraction ratio of interphase to nanotube was ~48. The square, representing the bulk PC, was set to have lateral dimensions 1000 times and 1500 times the thickness of the nanoplatelet, respectively. Three strain cases, including two farfield tensional strains ε_{11} and ε_{22} and one farfield shear strain ε_{12} (or ε_{21}) were applied.

The hybrid modeling method was used to predict the complex Young's moduli in both directions 11 and 22 for the polymer nanocomposite that contains 0.5 vol % curved nanoplatelets unidirectionally aligned in direction 11, as schematically shown in Fig. 3(*c*). Interphase properties four decades less mobile than the bulk PC are used, and the results are plotted in Fig. 10 and compared to predictions for the same composite with flat nanoplate-

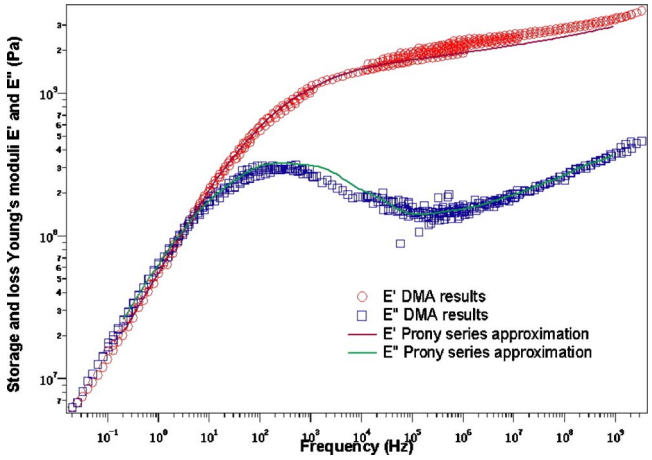


Fig. 11 Experimentally measured complex Young's modulus of PMMA and its Prony series approximation

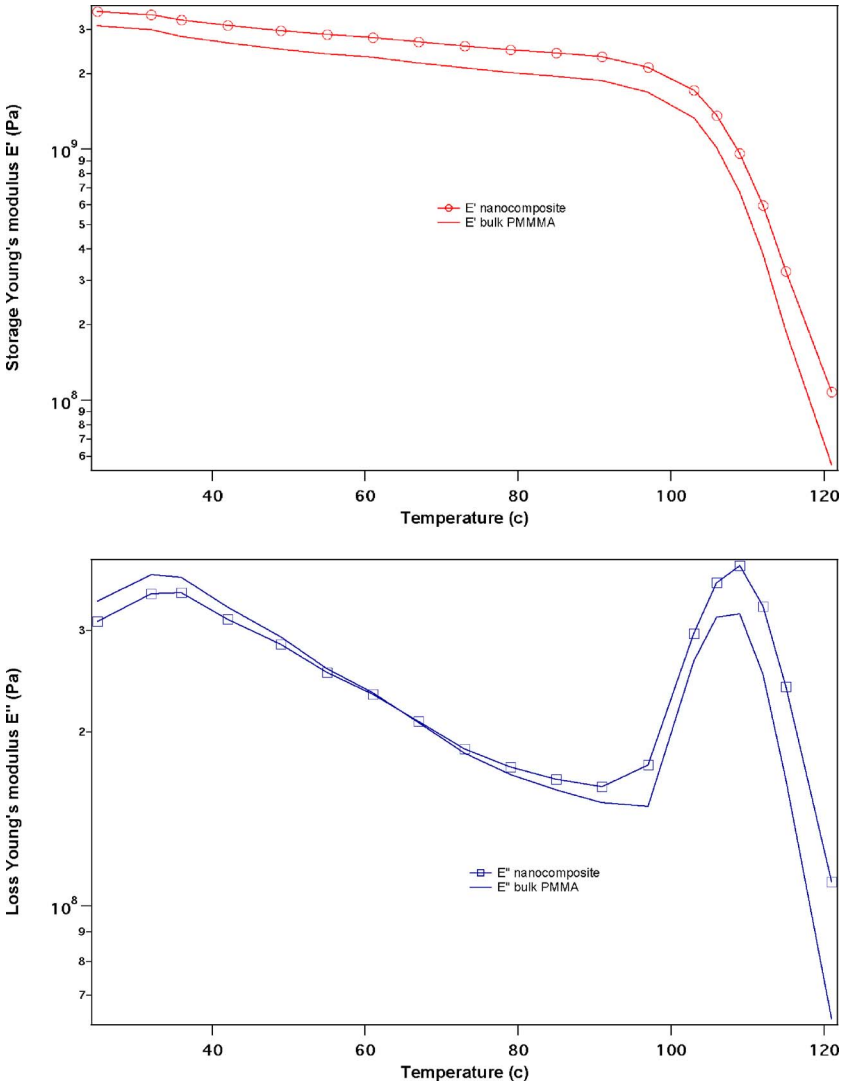


Fig. 12 Temperature domain response at 1 Hz, prediction for the nanotube/PMMA nanocomposite (nanotube content 0.5 vol %) versus the bulk PMMA

lets, schematically shown in Fig. 3(*b*).

As can be seen in Fig. 10, for both directions 11 and 22, the predicted storage and loss moduli of the nanocomposite with flat nanoplatelets are higher than those of the nanocomposite with curved nanoplatelets and the discrepancy is more pronounced in the low-frequency regime. Shown in Fig. 10(*b*), for both the nanocomposite with flat nanoplatelets and the nanocomposite with curved nanoplatelets, the loss Young's modulus peak is remarkably broadened, as compared to the pure PC, due to the large volume fraction of interphase for this case. For the polymer with the flat nanoplatelets, the storage and loss moduli in the 11 and 22 directions differ from each other in the high-frequency regime. Although anisotropy is expected for a composite with aligned platelike inclusions, it should be noted that for the composite here, the volume fraction of the plate itself is quite small, 0.5%. In addition, given the moderate aspect ratio of the more ellipsoidal interphase region (α–0.2) surrounding the plate at 24% volume fraction, the anisotropy between the 11 and 22 directions is small. For the nanocomposite with aligned, curved nanoplatelets, this small anisotropy has completely disappeared, indicating that the waviness of the nanoparticles can reduce the anisotropy of polymer nanocomposites.

3.3 Nanotube/PMMA Nanocomposite. To model nanotube/PMMA nanocomposites, the nanotubes were assumed to be unidirectionally aligned and ideally dispersed. Poly (methyl methacrylate) (PMMA) was provided by Polysciences, Warrington, Pennsylvania, with its frequency domain response obtained by frequency sweep mode dynamic mechanical analysis (DMA) [2]. By time-temperature superposition (TTSP), master curves at a reference temperature of 121°C were obtained and, subsequently, trimmed and approximated with a 29-term Prony series. The comparison between the experimental results and the Prony series approximation is shown in Fig. 11.

In order to predict the temperature domain response of the nanocomposites, the shift factors obtained in TTSP process for frequency domain response master curves were employed during the computational analysis. The frequency domain response of bulk PMMA at a certain temperature was obtained by shifting the master curves at reference temperature by shift factors corresponding to this temperature.

The finite element model used was the same as described in nanotube/PC nanocomposite section. The predicted temperature domain response (complex Young's modulus) at 1 Hz of the nanotube/PMMA nanocomposite with 0.5 vol % nanotubes and 4 vol % interphase that is four decades less mobile than the bulk PMMA is plotted against that of bulk PMMA in Fig. 12. As shown

clearly in the figure, the glassy state (before the transition zone) storage modulus of the nanocomposite is higher than that of pure PMMA. In the loss modulus plot, the location of the loss peak is approximately the same for both the nanocomposite and the pure PMMA, but difference in magnitude is apparent and the width of the loss peak of the nanocomposite is broadened compared to bulk PMMA. These predictions, manifesting the influence of the interphase, are consistent with the experimental observations for unfunctionalized nanotubes [2]. Since the experimental composite has a randomly oriented nanotube morphology, properties have a matrix dominated aspect and as such the results from the 2D transverse simulation here are representative of the influence of the interphase in the experimental system. Because of the 2D nature of our calculation, exact quantitative comparison to experiments is not appropriate. The similarity of the results is, however, striking and provides additional confidence in the interpretation of a relatively small discrete interphase zone in the experimented composite with unfunctionalized nanotubes.

4 Conclusion

A hierarchical numerical-analytical modeling technique was developed to predict the viscoelastic response of polymer nanocomposites and include the influence of the interphase and nanoparticle morphology. The method uses a finite element unit cell analysis to determine the strain concentration tensors of nanoparticle and inclusion for a dilute solution, then couples the numerical result with a traditional micromechanics Mori-Tanaka method to predict overall nanocomposite response. Here, interphase properties were assumed to be related to bulk matrix properties by a simple time shift that corresponds to altered mobility of the polymer in the vicinity of the nanoparticle. Future efforts will address more accurate interphase responses by performing molecular dynamics simulations or experimental measurement of localized behavior (the interphase(s)). This modeling method was implemented two-dimensionally on nanotube- and nanoplatelet-based polymer nanocomposites. Given the experimentally measured frequency domain response of the bulk polymer, the viscoelastic behaviors of the nanocomposites in both frequency and temperature domains can be calculated. Quantitative comparison between the experimental results and the model prediction is not possible at current stage due to uncertainties in nanotube properties, interphase properties, in situ nanotube morphology, and the 2D nature of the current simulations. Nevertheless, the predicted pattern of influence of the interphase on the overall performance of the nanocomposites is consistent with the experimental observation:

- The nanocomposite shows property improvement over the pure polymer.
- The extent of this improvement varies in different frequency or temperature regimes, which can be related to the mobility difference between the interphase and the bulk polymer.
- The transition zones of the nanocomposites are broader than those of the bulk polymer in the cases of small to moderate volume fraction of discrete (nonoverlapping) interphase, which corresponds to data on unfunctionalized nanocomposites.

The modeling method was also shown to have the capability of addressing the effect of in situ configuration of the nanoparticles. Particularly, the model prediction for the nanoplatelet/polycarbonate nanocomposite indicates that the waviness of the nanoplatelets can reduce the anisotropy of the nanocomposites.

The hybrid numerical-analytical method developed here can be adapted to model other behaviors of polymer nanocomposites, including thermal and electrical properties. Further investigation will address this issue, as well as 3D descriptions and the coupling to MD for more accurate assessment of the interphase.

Acknowledgment

The authors gratefully acknowledge the grant support from the NASA University Research Engineering and Technology Institute on Bio Inspired Materials (BIMat) under Grant No. NCC-1-02037.

References

[1] Schadler, L. S., 2003, "*Polymer-Based and Polymer-Filled Nanocomposites*," Nanocomposite Science and Technology, P. M. Ajayan, L. S. Schadler, and P. V. Braun, eds., Wiley, New York, p. 71–153.

[2] Ramanathan, T., Liu, H., and Brinson, L. C., 2005, "Functionalized SWNT Polymer Nanocomposites for Dramatic Property Improvement," J. Polym. Sci., Part B: Polym. Phys., **43**(17), pp. 2269–2279.

[3] Ramanathan, T., Stankovich, S., Dikin, D. A., Liu, H., Shen, H., Nguyen, S. T., and Brinson, L. C., 2006, "An Investigation of Particle Size and Dispersion and Their Influence on Graphite/PMMA Nanocomposites Properties," J. Polym. Sci., Part B: Polym. Phys. (submitted).

[4] Kasimatis, K. G., and Torkelson, J. M., 2004, "Well-Exfoliated, Kinetically Stable Polypropylene-Clay Nanocomposites Made by Solid-State Shear Pulverization," Abstracts of Papers of the American Chemical Society, 228 U429-U429 96-PMSE Part 2.

[5] Curran, S. A., Coleman, J. N., Dalton, A. B., Davey, A. P., Drury, A., and McCarthy, B., 1998, "A Composite From Poly(m-phenylenevinylene-co-2, 5-dioctoxy-p-phenylenevinylene) and Carbon Nanotubes: A Novel Material for Molecular Optoelectronics," Adv. Mater. (Weinheim, Ger.), **10**(14), pp. 1091–1093.

[6] Muzny, C. D., Butler, B. D., Hanley, H. J. M., Tsvetkov, F., and Peiffer, D. G., 1996, "Clay Platelet Dispersion in a Polymer Matrix," Mater. Lett., **28**(4-6), pp. 379–384.

[7] Mack, J. J., Viculus, L. M., Ali, A., Luoh, R., Yang, G., Hahn, H. T., Ko, F. K., and Kaner, R. B., 2005, "Graphite Nanoplatelet Reinforcement of Electrospun Polyacrylonitrile Nanofibers," Adv. Mater. (Weinheim, Ger.), **17**(1), pp. 77–80.

[8] Sandler, J. K. W., Shaffer, M. S. P., Prasse, T., Bauhofer, W. K., and Windle, A. H., 1999, "Development of a Dispersion Process for Carbon Nanotubes in an Epoxy Matrix and the Resulting Electrical Properties," Polymer, **40**(21), pp. 5967–5971.

[9] Sandler, J. K. W., Kirk, J. E., Kinloch, I. A., Shaffer, M. S. P., and Windle, A. H., 2003, "Ultra-Low Electrical Percolation Threshold in Carbon-Nanotube-Epoxy Composites," Polymer, **44**(19), pp. 5893–5899.

[10] Sandler, J. K. W., 2004, "A Comparative Study of Melt Spun Polyamide-12 Fibres Reinforced With Carbon Nanotubes and Nanofibres," Polymer, **45**(6), pp. 2001–2015.

[11] Coleman, J. N., Cadek, M., Blake, R., Nicolosi, V., Ryan, K. P., Belton, C., Fonseca, A., Nagy, J. B., Gun'ko, Y. K., and Blau, W. J., 2004, "High-Performance Nanotube-Reinforced Plastics: Understanding the Mechanism of Strength Increase," Adv. Funct. Mater., **14**(8), pp. 791–798.

[12] Dalton, A. B., Collins, S., Munoz, E., Razal, J. M., Ebron, V. H., Ferraris, J. P., Coleman, J. N., Kim, B. G., and Boughman, R. H., 2003, "Super-Tough Carbon-Nanotube Fibres: These Extraordinary Composite Fibres Can Be Woven Into Electronic Textiles," Nature (London), **423**(6941), p. 703.

[13] Cadek, M., Coleman, J. N., Barron, V., Hedicke, K., and Blau, W. J., 2002, "Morphological and Mechanical Properties of Carbon-Nanotube-Reinforced Semicrystalline and Amorphous Polymer Composites," Appl. Phys. Lett., **81**(27), pp. 5123–5125.

[14] Kilbride, B. E., Coleman, J. N., Fraysse, J., Fournet, P., Cadek, M., Drury, A., Hutzler, S., Roth, S., and Blau, W. J., 2002, "Experimental Observation of Scaling Laws for Alternating Current and Direct Current Conductivity in Polymer-Carbon Nanotube Composite Thin Films," J. Appl. Phys., **92**(7), pp. 4024–4030.

[15] Biercuk, M. J., Llaguno, M. C., Radosavljevic, M., Hyun, J. K., Johnson, A. T., and Fischer, J. E., 2002, "Carbon Nanotube Composites for Thermal Management," Appl. Phys. Lett., **80**(15), pp. 2767–2769.

[16] Wei, C. Y., Shrivastava, D., and Choi, K., 2002, "Thermal Expansion and Diffusion Coefficients of Carbon Nanotube-Polymer Composites," Nano Lett., **2**(6), pp. 647–650.

[17] Lim, S. K., 2002, "Preparation and Interaction Characteristics of Organically Modified Montmorillonite Nanocomposite With Miscible Polymer Blend of Poly(Ethylene Oxide) and Poly(methyl Methacrylate)," Chem. Mater., **14**(5), pp. 1989–1994.

[18] Hyun, Y. H., Lim, S. T., Choi, H. J., and Jhon, M. S., 2001, "Rheology of Poly(ethylene Oxide)/Organoclay Nanocomposites," Macromolecules, **34**(23), pp. 8084–8093.

[19] Messersmith, P. B., and Glannelis, E. P., 1995, "Synthesis and Barrier Properties of Poly(epsilon-Caprolactone)-Layered Silicate Nanocomposites," J. Polym. Sci., Part A: Polym. Chem., **33**(7), pp. 1047–1057.

[20] Messersmith, P. B., and Giannelis, E. P., 1994, "Synthesis and Characterization of Layered Silicate-Epoxy Nanocomposites," Chem. Mater., **6**(10), pp. 1719–1725.

[21] Coleman, J. N., Curran, S., Dalton, A. B., Davey, A. P., McCarthy, B., Blau, W., and Barklie, R. C., 1998, "Percolation-Dominated Conductivity in a Conjugated-Polymer-Carbon-Nanotube Composite," Phys. Rev. B, **58**(12), pp. R7492–R7495.

[22] Putz, K. W., Mitchell, C. A., Krishnamoorti, R., and Green, P. F., 2004, "Elas-

tic Modulus of Single-Walled Carbon Nanotube/Poly(Methyl Methacrylate) Nanocomposites," J. Polym. Sci., Part A: Polym. Chem., **42**(12), pp. 2286–2293.
[23] Fisher, F. T., Eitan, A., Andrews, R., Schadler, L. S., and Brinson, L. C., 2004, "Spectral Response and Effective Viscoelastic Properties of MWNT-Reinforced Polycarbonate," Adv. Compos. Lett., **13**(2), pp. 105–111.
[24] Cox, H. L., 1952, "The Elasticity and Strength of Paper and Other Fibrous Materials," Br. J. Appl. Phys., **3**, pp. 72–78.
[25] Wagner, H. D., Lourie, O., Feldman, Y., and Tenne, R., 1998, "Stress-Induced Fragmentation of Multiwall Carbon Nanotubes in a Polymer Matrix," Appl. Phys. Lett., **72**(2), pp. 188–190.
[26] Cooper, C. A., Cohen, S. R., Barber, A. H., and Wagner, H. D., 2002, "Detachment of Nanotubes From a Polymer Matrix," Appl. Phys. Lett., **81**(20), pp. 3873–3875.
[27] Frankland, S. J. V., Caglar, A., Brenner, D. W., and Griebel, M., 2002, "Molecular Simulation of the Influence of Chemical Cross-Links on the Shear Strength of Carbon Nanotube-Polymer Interfaces," J. Phys. Chem. B, **106**(12), pp. 3046–3048.
[28] Frankland, S. J. V., and Harik, V. M., 2003, "Analysis of Carbon Nanotube Pull-Out From a Polymer Matrix," Surf. Sci., **525**(1–3), pp. L103–L108.
[29] Liao, K., and Li, S., 2001, "Interfacial Characteristics of a Carbon Nanotube-Polystyrene Composite System," Appl. Phys. Lett., **79**(25), pp. 4225–4227.
[30] Wong, M., Paramsothy, M., Xu, X. J., Ren, Y., Li, S., and Liao, K., 2003, "Physical Interactions at Carbon Nanotube-Polymer Interface," Polymer, **44**(25), pp. 7757–7764.
[31] Wagner, H. D., 2002, "Nanotube-Polymer Adhesion: A Mechanics Approach," Chem. Phys. Lett., **361**(1–2), pp. 57–61.
[32] Schadler, L. S., Giannaris, S. C., and Ajayan, P. M., 1998, "Load Transfer in Carbon Nanotube Epoxy Composites," Appl. Phys. Lett., **73**(26), pp. 3842–3844.
[33] Ajayan, P. M., Schadler, L. S., Giannaris, C., and Rubio, A., 2000, "Single-Walled Carbon Nanotube-Polymer Composites: Strength and Weakness," Adv. Mater. (Weinheim, Ger.), **12**(10), pp. 750–753.
[34] Thomason, J. L., 1995, "The Interface Region in Glass Fibre-Reinforced Epoxy Resin Composites: 3. Characterization of Fibre Surface Coatings and the Interphase," Composites, **26**, pp. 487–498.
[35] Mai, K., Mader, E., and NMuhle, M., 1998, "Interphase Characterization in Composites With New Non-Destructive Methods," Composites, Part A, **29A**, pp. 1111–1119.
[36] Starr, F. W., Schroder, T. B., and Glotzer, S. C., 2002, "Molecular Dynamics Simulation of a Polymer Melt With a Nanoscopic Particle," Macromolecules, **35**(11), 4481–4492.
[37] Starr, F. W., Schroder, T. B., and Glotzer, S. C., 2001, "Effects of a Nanoscopic Filler on the Structure and Dynamics of a Simulated Polymer Melt and the Relationship to Ultrathin Films," Phys. Rev. E, **64**(2), p. 021802.
[38] Smith, G. D., Bedrov, D., Li, L. W., and Byutner, O., 2002, "A Molecular Dynamics Simulation Study of the Viscoelastic Properties of Polymer Nanocomposites," J. Chem. Phys., **117**(20), pp. 9478–9489.
[39] Brune, D. A., and Bicerano, J., 2002, "Micromechanics of Nanocomposites: Comparison of Tensile and Compressive Elastic Moduli, and Prediction of Effects of Incomplete Exfoliation and Imperfect Alignment on Modulus," Polymer, **43**(2), pp. 369–387.
[40] Sheng, N., Boyce, M. C., Parks, D. M., Rutledge, G. C., Abes, J. I., and Cohen, R. E., 2004, "Multiscale Micromechanical Modeling of Polymer/Clay Nanocomposites and the Effective Clay Particle," Polymer, **45**(2), pp. 487–506.
[41] Frogley, M. D., Ravich, D., and Wagner, H. D., 2003. "Mechanical Properties of Carbon Nanoparticle-Reinforced Elastomers," Compos. Sci. Technol., **63**(11), pp. 1647–1654.
[42] Shia, D., Hui, C. Y., Burnside, S. D., and Giannelis, E. P., 1998, "An Interface Model for the Prediction of Young's Modulus of Layered Silicate-Elastomer Nanocomposites," Polym. Compos., **19**(5), pp. 608–617.
[43] Fisher, F. T., Bradshaw, R. D., and Brinson, L. C., 2003, "Fiber Waviness in Nanotube-Reinforced Polymer Composites-1: Modulus Predictions Using Effective Nanotube Properties," Compos. Sci. Technol., **63**(11), pp. 1689–1703.
[44] Luo, J. J., and Daniel, I. M., 2003, "Characterization and Modeling of Mechanical Behavior of Polymer/Clay Nanocomposites," Compos. Sci. Technol., **63**(11), pp. 1607–1616.
[45] Liu, Y. J., and Chen, X. L., 2003, "Evaluations of the Effective Material Properties of Carbon Nanotube-Based Composites Using a Nanoscale Representative Volume Element," Mech. Mater., **35**(1-2), pp. 69–81.
[46] Chen, X. L., and Liu, Y. J., 2004, "Square Representative Volume Elements for Evaluating the Effective Material Properties of Carbon Nanotube-Based Composites," Comput. Mater. Sci., **29**(1), pp. 1–11.
[47] Li, C. Y., and Chou, T. W., 2003, "Multiscale Modeling of Carbon Nanotube Reinforced Polymer Composites," J. Nanosci. Nanotechnol., **3**(5), pp. 423–430.
[48] Li, C. Y., and Chou, T. W., 2004, "Modeling of Elastic Buckling of Carbon Nanotubes by Molecular Structural Mechanics Approach," Mech. Mater., **36**(11), pp. 1047–1055.
[49] Miller, R. E., and Tadmor, E. B., 2002, "The Quasicontinuum Method: Overview, Applications and Current Directions," J. Comput.-Aided Mater. Des., **9**(3), pp. 203–239.
[50] Mori, T., and Tanaka, K., 1973, "Average Stress in Matrix and Average Elastic Energy of Materials With Misfitting Inclusions," Acta Metall., **21**, pp. 571–574.
[51] Eshelby, J. D., 1957, "The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems," Proc. Phys. Soc., London, Sect. A, **241**(1226), pp. 376–396.
[52] Eshelby, J. D., 1961, "Elastic Inclusions and Inhomogeneities," *Progress in Solid Mechanics*, Vol. 2, I. N. Sneddon and R. Hills, eds., North-Holland, Amsterdam, pp. 89–140.
[53] Benveniste, Y., 1987, "A New Approach to the Application of Mori-Tanaka Theory in Composite-Materials," Mech. Mater., **6**(2), pp. 147–157.
[54] Tandon, G. P., and Weng, G. J., 1986, "Average Stress in the Matrix and Effective Moduli of Randomly Orientated Composites," Compos. Sci. Technol., **27**, pp. 111–132.
[55] Hill, R., 1963, "Elastic Properties of Reinforced Solid: Some Theoretical Principles," J. Mech. Phys. Solids, **11**(5), pp. 357–372.
[56] Hashin, Z., 1965, "Viscoelastic Behavior of Heterogeneous Media," ASME J. Appl. Mech., **32**(3), pp. 630–636.
[57] Wang, Y. M., and Weng, G. J., 1992, "The Influence of Inclusion Shape on the Overall Viscoelastic Behavior of Composites," ASME J. Appl. Mech., **59**(3), pp. 510–518.
[58] Mura, T., 1982, *Micromechanics of Defects in Solids*, Kluwer, Dordrecht.
[59] Yasmin, A., Abot, J. L., and Daniel, I. M., 2003, "Processing of Clay/Epoxy Nanocomposites by Shear Mixing," Scr. Mater., **49**(1), pp. 81–86.
[60] Bradshaw, R. D., Fisher, F. T., and Brinson, L. C., 2003, "Fiber Waviness in Nanotube-Reinforced Polymer Composites-II: Modeling via Numerical Approximation of the Dilute Strain Concentration Tensor." Compos. Sci. Technol., **63**(11), pp. 1705–1722.
[61] Park, C. H., Kim, J. H., Ree, M., Sohn, B. H., Jung, J. C., and Zin, W. C., 2004, "Thickness and Composition Dependence of the Glass Transition Temperature in Thin Random Copolymer Films." Polymer, **45**(13), pp. 4507–4513.
[62] Ellison, C. J., and Torkelson, J. M., 2003, "The Distribution of Glass-Transition Temperatures in Nanoscopically Confined Glass Formers," Nat. Mater., **2**(10), pp. 695–700.
[63] Ding, W., Eitan, A., Fisher, F. T., Chen, X., Dikin, D. A., Andrews, R., Brinson, L. C., Schadler, L. S., and Ruoff, R. S., 2003, "Direct Observation of Polymer Sheathing in Carbon Nanotube-Polycarbonate Composites," Nano Lett., **3**(11), pp. 1593–1597.
[64] Shodja, H. M., and Sarvestani, A. S., 2001, "Elastic Fields in Double Inhomogeneity by the Equivalent Inclusion Method," ASME J. Appl. Mech., **68**(1), pp. 3–10.
[65] Sarvestani, A. S., 2003, "On the Overall Elastic Moduli of Composites With Spherical Coated Fillers," Int. J. Solids Struct., **40**(26), pp. 7553–7566.
[66] Bradshaw, R. D., and Brinson, L. C., 1997, "A Sign Control Method for Fitting and Interconverting Material Functions for Linearly Viscoelastic Materials," Mech. Time-Depend. Mater., **1**, pp. 85–108.

Mingji Wang

Kenneth M. Liechti

Center for the Mechanics of Solids, Structures and Materials,
Department of Aerospace Engineering and Engineering Mechanics

Vibha Srinivasan
Institute for Theoretical Chemistry

John M. White
Center for Nanomolecular Science and Technology

Peter J. Rossky
Institute for Theoretical Chemistry,
The University of Texas,
Austin, TX 78712

Matthew T. Stone
Exxon Mobil Upstream Research Company,
Houston, TX 77252-2189

A Hybrid Continuum-Molecular Analysis of Interfacial Force Microscope Experiments on a Self-Assembled Monolayer

Nanoindentation experiments were performed on a defect-free, molecular self-assembled monolayer of octadecyltrichlorosilane (OTS) on silicon using an interfacial force microscope (IFM). The IFM provided repeatable and elastic force profiles corresponding to the adhesive and compressive response of these 2.5 nm *thick monolayers. As a first step in the analysis of the force profiles, the OTS was assumed to be linearly elastic and isotropic, and adhesive interactions were accounted for via a cohesive zone model. However, the assumption of linearity gave rise to force profiles that did not match the measurements. As a result, the mechanical behavior of the OTS was extracted from molecular-dynamics simulations and represented as a hypoelastic material, which, when used in finite element analyses of the IFM experiments, was able to fully reproduce the force profiles. This suggests that the continuum representation of the mechanical and adhesive behavior of self-assembled monolayers may be directly obtained from molecular analyses.* [DOI: 10.1115/1.1943435]

1 Introduction

Molecular self-assembled monolayers are films that spontaneously form on certain substrates while immersing the substrate in a solution of an active surfactant in an organic solvent. The organic monolayers formed through self-assembly are usually closely packed and therefore have a highly ordered structure [1]. The functionality of the surface can be controlled by choosing different tail groups or by chemical derivations of tail groups. Multiple layers can also be constructed by functionalizing the end group and then depositing another monolayer on top of the previous one [2,3]. These features allow SAMs to be tailored for studying the fundamental relationship between the interphase chemistry and mechanical properties in interphase regions. They have the potential to form well-defined interphases. A SAM or mixtures of SAMs with different functionalities have been used to create well-defined combinations of strong (specific) and weak (nonspecific) interfacial interactions [4–6]. In the latter study, it was found that fracture occurred in or slightly above the SAM, reinforcing the importance of understanding the nanomechanical behavior of such monolayers. Self-assembled monolayers are used in MEMS devices for the reduction of stiction, friction, and wear [7–10]. This has led to studies of the frictional behavior between scanning probe microscope tips (functionalized [11] and bare [12]) and surfaces coated with self-assembled monolayers. These also require good descriptions of the constitutive behavior of SAMs at molecular and continuum levels.

Nanoindentation has become an accepted experimental technique for measuring the mechanical and adhesive properties of materials, especially when one of the dimensions is small. Interfacial force microscopy (IFM) has been shown to be exceptionally useful in measuring mechanical and adhesive properties of ultra-thin films [13–17]. When an IFM is used as a nanoprobe, the force and displacement are recorded directly. Due to its unique self-balancing sensor and feedback circuit, probing with the IFM allows the entire force-separation response between its tip and the surface to be followed without the mechanical instability that occurs with the atomic force microscope (AFM) [18,19]. The high resolution in both force and displacement measurements allows it to interrogate nanomechanical properties within the interphase region in fiber-reinforced composites. Direct measurement of the elastic modulus in interphase regions has been performed with IFM [13,14]. It was found that the modulus varied continuously with the radial position from the fiber.

Analytical or numerical contact mechanics analysis can be used to extract the mechanical and adhesive properties from the force profile. Hertz solved the contact problem between two spherical elastic bodies by assuming a hemispherical pressure distribution inside the contact area. Johnson, Kendall, and Roberts [20] and Dejaguin, Muller, and Toporov [21] formulated the JKR and DMT theories, which included surface interactions either inside or outside the contact area, respectively. Maugis [22] treated the contact between two spherical bodies as a fracture problem and derived more complex and complete formulas by incorporating Sneddon's solution with a Dugdale cohesive zone model. Maugis' solution bridges the JKR and the DMT theories. All the contact mechanics theories just mentioned are based on, and should usually be applied to, monolithic materials or sufficiently thick films on a substrate [17]. For thin films on substrates, special care must be taken to eliminate substrate effects when the mechanical properties of the film are being extracted. However, in some cases such as when the films are ultra thin, the strong influence of substrates cannot be avoided, so that all the above-mentioned analytical theories fail and numerical analyses are required [17].

Molecular simulations provide another technique for studying the structural and mechanical behavior of self-assembled monolayers on the atomic or molecular level. Siepmann and McDonald [23] used the Monte Carlo technique to study the mechanical relaxation of a $CH_3(CH_2)_{15}SH/Au(111)$ SAM subject to compression with a flat surface. Tupper et al. [24,25] performed molecular dynamics analyses to simulate a SAM of *n*-hexadecanethiol

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received May 24, 2004; final manuscript received November 24, 2004. Editor: R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California–Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

chemisorbed on a gold (111) surface and compressed uniformly by another Au (111) surface. Henda et al. [26] used the static energy minimization technique to study the effects of force fields and monolayer structures on the calculated elastic properties. These models were quite simplified, and the duration of the loading was very short compared to real indentation time scales. Nonetheless, because the force levels matched those of the experiments [24,25], these simulations usefully explored the macroscopic properties of self-assembled monolayers from atomistic levels. All the studies revealed that the variation in mechanical response was related to stress-induced structural rearrangements.

The objective of the present study is to provide a nanomechanical analysis of IFM force profiles produced by a tungsten tip probing a self-assembled monolayer. An octadecyltrichlorosilane (OTS) self-assembled monolayer was deposited on a silicon (100) surface covered with an oxide layer. This configuration was first numerically analyzed based on the assumption that the OTS was isotropic and linearly elastic. Surface interactions were also included in the model. An isotropic, nonlinear elastic constitutive model for the OTS was then developed that was based on molecular dynamics simulations.

2 Fabrication of the Self-Assembled Monolayer

The substrate that was used for the OTS deposition was a polished boron P-type single crystal silicon wafer (100) (MEMC Electronic Materials, Inc). Silicon chips with dimensions 10 $\times 10$ mm^2 were cut from a polished single crystal silicon wafer. Monolayers were deposited from an anhydrous millimolar solution of OTS (CH_3-$(CH_2)_{17}$-Si-Cl_3, 90+ %) in dicyclohexyl ($C_{12}H_{22}$, 99%) (Sigma-Aldrich, Inc). A new deposition process was developed [27] that emphasized the formation of a new, heavily hydroxylated oxide layer on the silicon, anhydrous deposition of the OTS, and the removal of any physisorbed OTS clusters on top of the OTS monolayer. This ensured that a true monolayer was covalently attached to the SiO_2 layer without any cross-linking between OTS chains.

The successful deposition of OTS monolayers on silicon was confirmed by several different techniques, including contact angle, AFM, and ellipsometry measurements [27]. The water contact angle was 110 deg, which was very close to the expected value for a surface terminated with methyl (CH_3) groups. Tapping mode AFM imaging of the OTS monolayer indicated that the surface was uniform with an RMS roughness less than 0.1 nm, indicating that it was free of clusters. The thickness of the monolayer was determined by ellipsometry. The measured thickness of the new oxide layer was 1.9±0.1 nm. The average total thickness of OTS monolayer and the oxide layer was 4.4±0.1 nm. Subtracting out the oxide layer thickness yielded an average OTS film thickness of 2.5±0.1 nm, which is the expected value for a 17 deg tilt angle (see Sec. 6.2, Molecular Dynamics Simulation).

3 Nanoindentation With IFM

The interfacial force microscope (IFM) employs a self-balancing differential capacitance force sensor [18,19]. The sensor sits on a piezoelectric scanner. An electrochemically etched tungsten tip is attached to the sensor as a probe. The geometry of the tungsten tip was measured with a scanning electron microscope (SEM). Tips that had a parabolic shape and radii ranging from 100 to 500 nm were recorded and kept for future use. The overall yield of the electrochemical etching process was about 50%. A special rig was designed to hold the sensor with tip mounted on it so that the shape of the tips could be checked in a scanning electron microscope after each experiment. Only data obtained from experiments without tip damage are reported.

The IFM was used to probe the self-assembled monolayers of OTS deposited on silicon (Fig. 1(*a*)). The IFM that was used in this study is capable of measuring normal forces from 1 nN to 400 μN, and normal displacements with a resolution of 0.01 nm. This level of resolution allows the IFM to probe the samples at low force levels in such a way that no permanent damage occurs to the sample or the tip.

Nanoindentation experiments on OTS monolayers were performed under displacement control at 0.26 nm/s under ambient temperature and humidity conditions. A typical IFM force profile of tungsten tip on OTS is shown in Fig. 1(*b*). As the probe approached the surface, there was a jump to a small tensile load, which was probably due to configurational changes in the OTS. This was followed by slight increase in the tensile load, which then decreased as the probe came into contact with the OTS layer. The specimen was unloaded at about 2.5 μN and the unloading path followed the loading one without the adhesive jump that had occurred during loading. The lack of hysteresis in the compressive portion of the response indicates that the OTS behaved elastically. Figure 1(*c*) is a comparison of the loading portion of the force profiles from bare and coated samples. The obvious differences between the two demonstrates that the IFM is capable of detecting the presence of the OTS monolayer. The OTS is clearly more compliant than the silicon, and hydrophobic nature of the OTS surface resulted in lower adhesion. The adhesive jump that occurred for the bare silicon sample is probably due to the presence of adventitious water.

4 Finite Element Analysis of Adhesive Contact

The reduced elastic modulus $E^*=E/(1-\nu^2)$ of a monolithic sample with a Young's modulus E and Poisson's ratio ν can be extracted from force profiles using contact mechanics analyses. The same analyses can be used for layered samples when the contact radius is less than 10% of the film thickness [17]. However, since the thickness (2.5 nm) of OTS monolayers is much smaller than the indenter radius (115 nm) and the 10% rule cannot be satisfied experimentally, numerical methods have to be employed. In this study, finite element analyses that incorporated user-defined surface interaction elements [17] were used. The IFM indentation was modeled as a spherical tungsten tip probing a two-layered film on a silicon substrate (Fig. 1(*a*)). The geometry of the cohesive zone model and a triangular normal adhesive interaction law are shown in Fig. 2.

The finite element analysis was performed with the commercial finite element package ABAQUS. An axisymmetric analysis was used due to the spherical contact. All the materials were considered to be linearly elastic and isotropic at this stage. Geometrically nonlinear analysis was used throughout the simulation due to large deformations in the contact region (see Table 1). Quadrilateral linear axisymmetric elements with reduced integration were used for both the indenter and specimen. The meshes were biased toward the contact region, where the largest deformation was expected. In order to simulate the contact area accurately, very fine meshes were used over the top layers of the potential contact surfaces. Infinite elements were used for the far-field region ($>$20 times the maximum contact radius) of the half-space so that no boundary conditions needed to be enforced on the unbounded domain. A suitable level of mesh refinement was established on the basis of the Hertz problem. There were 7450 nodes and 6889 elements in the model.

The compressive regime of the nanoindentation was modeled as frictionless contact using standard ABAQUS surface contact interactions. The default contact pressure-clearance relationship in ABAQUS is referred to as the "hard" contact model. It only transmits pressure once the surfaces come into contact. This interaction model is not sufficient for analyzing indentation experiments where the adhesive effect is large. As a result, user-defined elements were incorporated in the above-mentioned mesh when adhesive interactions were required. The user element code was programmed in FORTRAN language and added to the ABAQUS input file. Figure 2(*b*) shows the force-separation interaction with a

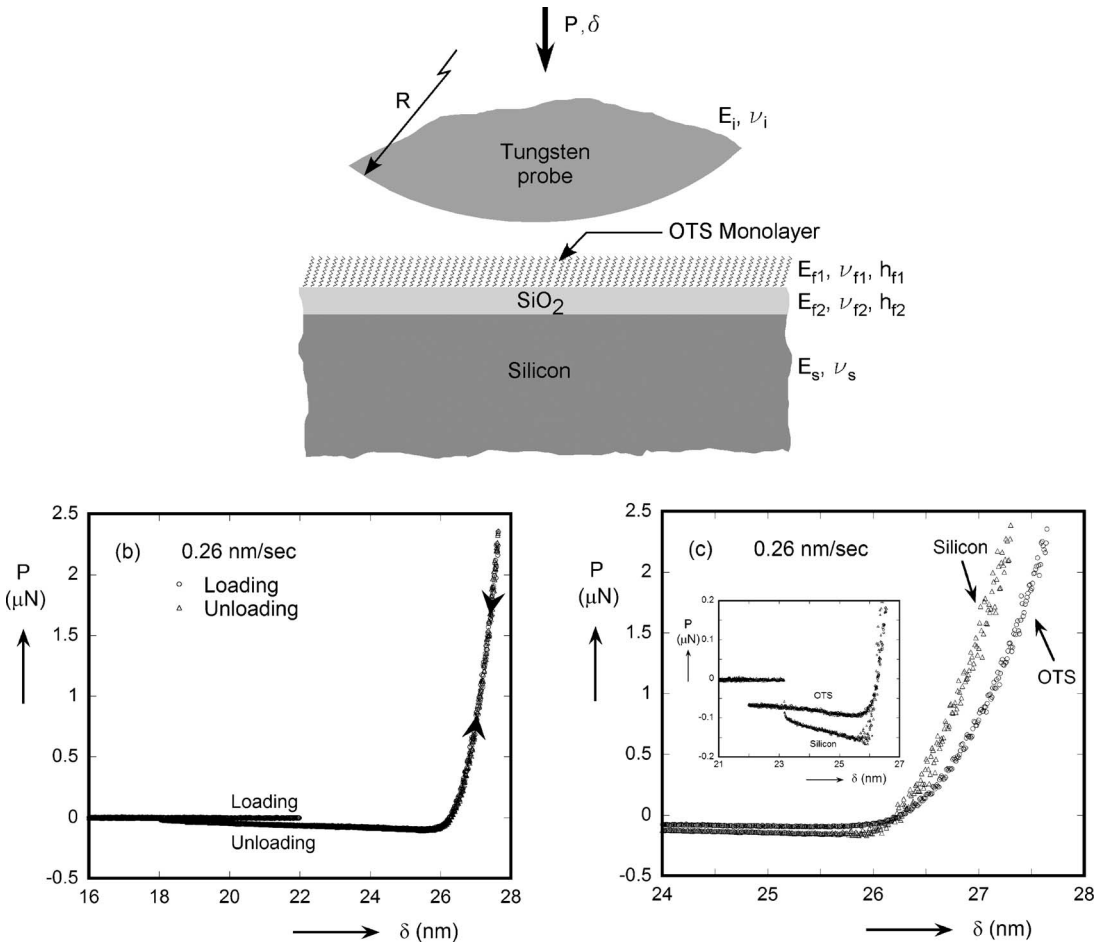


Fig. 1 (*a*) **Nanoindentation of an OTS monolayer,** (*b*) **a typical loading/unloading force profile for OTS on silicon, and** (*c*) **a comparison of the loading force profiles for bare silicon and OTS on silicon**

user-defined triangular traction-separation law. There are three parameters that define it: the maximum adhesive traction σ_0, the corresponding displacement δ_0, and the cutoff displacement δ_t. The surface energy can then be expressed as the area underneath the traction-separation curve, i.e.,

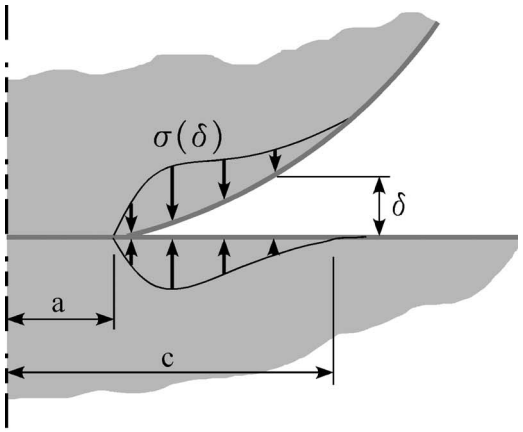
$$\omega = \frac{1}{2}\sigma_0\delta_t \tag{1}$$

5 Molecular Dynamics Simulation

Molecular dynamics was used to simulate a self-assembled OTS monolayer on a SiO_2 substrate and the compression of the monolayer by a flat tungsten plate. The simulation was performed with the molecular dynamics simulation package DL_POLY package 2.0.[1] Integration was based on the simple Verlet leapfrog scheme. The simulation was carried out in the NVT ensemble, which has a fixed number of particles and constant total volume and temperature, using a Hoover-Nose thermostat. Room temperature (298.15 K) was maintained during the calculation. To make the simulation simple, the top layer of the SiO_2 substrate and the tungsten plate were modeled as rigid surfaces with one layer of atoms. The hexagonal arrangement of the SiO_2 surface was modeled as described by Stevens [29], who accounted for surface corrugation (Fig. 3(*a*) and 3(*b*)). The body-centered cubic structure of tungsten crystals was modeled as a flat atomic surface in the (100) direction. Both the SiO_2 and tungsten were treated as rigid because they were expected to be much stiffer than the OTS monolayer. The OTS molecular chains were modeled using united atoms (pseudoatoms), where the CH_3 and CH_2 groups were represented as single effective pseudoatoms. Slab periodic boundary conditions, which were periodic in the x and y directions, were used to represent the infinite extent in the two in-plane directions. The repeating cell had lateral dimensions of 26×27 Å^2 containing 30 OTS molecules. Figure 3(*c*) shows the molecular dynamics simulation cell for OTS on silica. All the OTS molecular chains were initially arranged perpendicular to the SiO_2 surface. This choice of cell size is comparable to earlier studies [25]. A more complete study, of the effect of cell size was beyond the scope of this study, but would be a valuable exercise.

Atomic molecular modeling via empirical potential functions is highly developed [30,31]. The forces are defined by a set of empirical potential functions with forms having a basis in chemical physics, and these are parameterized to give reasonable energy and force levels. For organic molecules, there are two classes of forces: intramolecular and intermolecular interactions. In this study, harmonic bond, harmonic valence angle, and cosine dihedral angle potentials were used for intramolecular interactions. These potentials describe the explicit force between atoms due to chemical bonds, the bond bending, and the interaction arising from torsion forces in molecules, respectively. The total configuration energy of the molecular system was the sum of all the above interaction potentials and could be written as

[1]DL_POLY 2.0 package was developed at Daresbury Laboratory, U.K. and is available free, under license, to academic institutions on a worldwide basis 28.



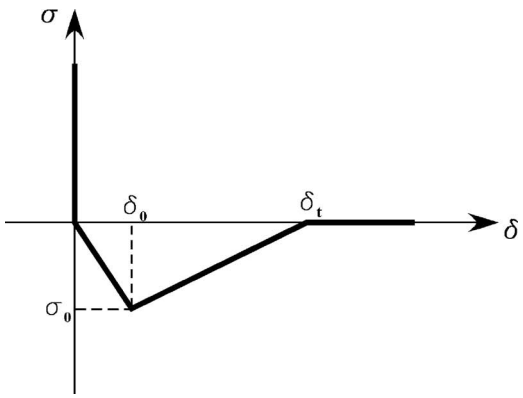


Fig. 2 (*a*) Cohesive zone model geometry showing the contact radius *a*, cohesive zone size (*c*−*a*) and the normal traction and separation σ and δ between the surfaces and (*b*) a triangular traction-separation law specifying $\sigma(\delta)$

$$U(\underline{r}_1,\underline{r}_2,\ldots,\underline{r}_N) = \sum_{i_{\text{bond}}=1}^{N_{\text{bond}}} U_{\text{bond}}(i_{\text{bond}},\underline{r}_a,\underline{r}_b) + \sum_{i_{\text{angle}}=1}^{N_{\text{angle}}} U_{\text{angle}}(i_{\text{angle}},\underline{r}_a,\underline{r}_b,\underline{r}_c) + \sum_{i_{\text{diahed}}=1}^{N_{\text{diahed}}} U_{\text{diahed}}(i_{\text{diahed}},\underline{r}_a,\underline{r}_b,\underline{r}_c,\underline{r}_d) + \sum_{i=1}^{N-1}\sum_{j>i}^{N} U_{\text{pair}}(i,j,|\underline{r}_i-\underline{r}_j|). \quad (2)$$

The intramolecular interactions were specified for all bonds, valence angles, and dihedrals. The potentials were

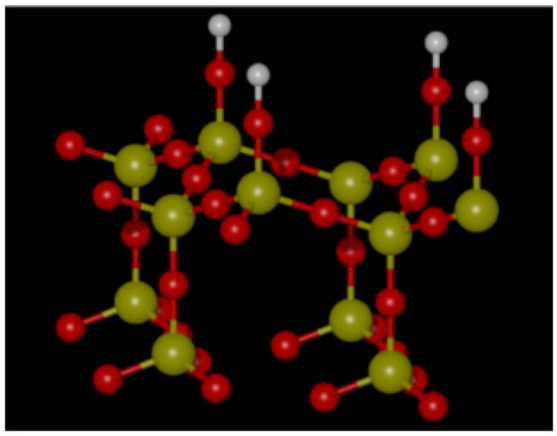
1. Harmonic bond potential

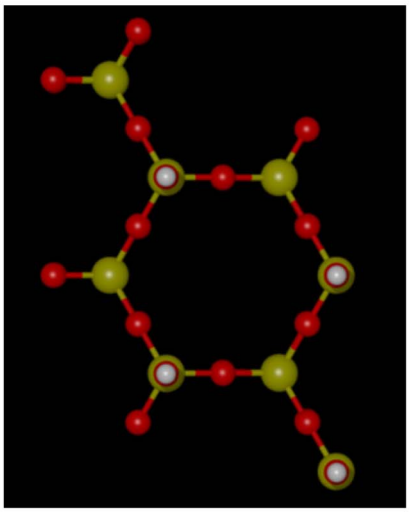
$$U(r_{ij}) = \tfrac{1}{2}k(r_{ij}-r_0)^2 \quad (3)$$

Table 1 Material properties

Material	Elastic modulus E (GPa)	Poisson's ratio ν
Tungsten[a]	392	0.28
SiO_2[b]	73.6	0.17
Silicon[a]	168	0.22

[a]Reference [36]
[b]By Meller Optics, Inc.





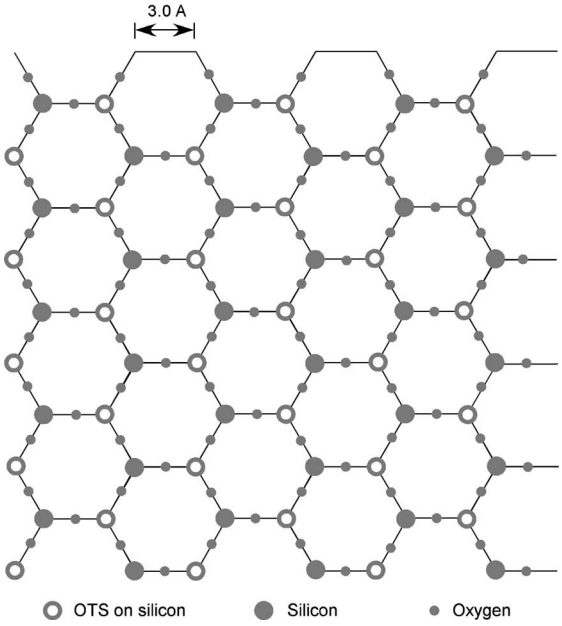


Fig. 3 (*a*) A three-dimensional representation of the atomic structure of silica terminated with hydroxyl groups (O is red, Si is yellow, and H is white.); (*b*) plan view, [29]; and (*c*) molecular dynamics simulation cell for OTS on silica—the small solid circles represent oxygen atoms, large solid circles represent silicon atoms on lower plane, and the remaining open circles represent the silicon atoms with OTS molecules on them

2. Harmonic valence angle potential

$$U(\theta_{jik}) = \frac{k}{2}(\theta_{jik} - \theta_0)^2 \tag{4}$$

3. Cosine dihedral angle potential

$$U(\phi) = \frac{A_1}{2}(1 + \cos\phi) + \frac{A_2}{2}(1 - \cos 2\phi) + \frac{A_3}{2}(1 - \cos 3\phi) \tag{5}$$

All the intermolecular interactions were modeled with (12-6) Lennard-Jones van der Waals potentials. The forcecutoff distance was set to 10 Å. No coulombic forces were considered. The intermolecular interactions were specified by atom types and were represented by the short range (van der Waals) pair-body Lennard-Jones potential as

$$U(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{6}$$

The specific values that were used for the various interactions are listed in the Appendix. The bond angles and bond lengths for the OTS molecule were taken from previous studies [25,29] and correspond to the values from the CFF93 force field. The parameters for the OTS molecule are the values based on the OPLS-UA (optimized potentials for liquid simulations-united atom) force field, and have been determined by Jorgensen et al. [32–34]. The one Si atom in the OTS chain was assumed to behave like a CH_2 group (i.e., the CH_2-CH_2-Si-OH and CH_2-CH_2-CH_2-OH potentials were assumed to be the same), becuase the potentials for Si in alkylsilanes were not available in the literature.

During the simulation, the trajectory of each atom was followed. Macroscopic quantities were averaged on the statistical basis of the configuration history. In particular, the stress tensor was averaged over the volume through

$$\langle \underset{\sim}{\sigma} \rangle = \frac{1}{V}\sum_k \underset{\sim}{f}^k \otimes \underset{\sim}{r}^k \tag{7}$$

with

$$\underset{\sim}{f} = \left[\frac{\partial U(r)}{\partial r}\right]\left(\frac{\underset{\sim}{r}}{r}\right) \tag{8}$$

where $r=|\underset{\sim}{r}|$, and V is the volume of the current configuration.

6 Results and Discussion

The IFM force profiles were analyzed with linear and nonlinear material models. Molecular dynamics simulations were used to develop the nonlinear material constitutive relations that were used.

6.1 Linearly Elastic Analysis of the IFM Profiles. Finite element analyses were used to extract the Young's modulus of the OTS monolayer, assuming that it behaves as an isotropic, linearly elastic material. Adhesive interactions had to be considered in these analyses because their effect is significant at the low force levels that were used in the experiments.

Previous finite element analyses [17] showed that the elastic properties determine the slope of the response from the maximum adhesive force to the maximum indentation force, and the surface energy ω determines the maximum adhesive force itself. The quantities σ_0, δ_t, and δ_0 adjust the shape of the force separation profile from long-range interactions to the maximum adhesive force region. The adhesive portion of the response was best fitted with σ_0=37.5 MPa, δ_t=7 nm, and δ_0=5 nm, corresponding to ω=131.3 mJ/m^2. The Young's modulus of the OTS monolayer was then adjusted to fit the complete force profile. Two different values (0 and 0.44) of Poisson's ratio were considered. The latter value was determined from a molecular dynamics simulation as will be described later. Figure 4 shows comparisons between the analyses and data. The extracted modulus of the OTS monolayer was 15±5 GPa when Poisson's ratio was zero (Fig. 4(*a*)), and 6±1.5 GPa for a Poisson's ratio of 0.44 (Fig. 4(*b*)). These values are higher than those of amorphous γ-APS polymer films of about 3.5 GPa, whose properties were also measured with an IFM [17]. This is not surprising because OTS monolayers are highly ordered. The direction of the indentation force was such that it compressed the covalent carbon bonds and caused the chains to shorten as well as tilt. These types of interactions are expected to be stiffer than those associated with the pure van der Waals interactions associated with amorphous polymers. Tupper et al. [24,25] conducted a molecular dynamics simulation of the compression of an *n*-hexadecane monolayer by gold plates. The Young's modulus that was extracted from the initial linear portion of the force profile was 22 GPa. The molecular chain length of *n*-hexadecane is shorter than that of OTS. Thus, it is reasonable that a monolayer with a shorter chain would be stiffer. Henda et al. [26] conclude that the initial tangent moduli of monolayers of CH_3-$(CH_2)_{17}$-S- had values between 15 and 18 GPa. In both cases, the Young's moduli were extracted based on the assumption that the stress state was uniaxial. However, since periodic boundary conditions had been used to simulate the infinite in-plane extent of the films, the stress state was actually triaxial. Thus, the Young's moduli should be lower by a factor of $(1-\nu-2\nu^2)/(1-\nu)$ (0.3 for a Poisson's ratio of 0.44), or an inferred value of about 5.0±0.5 GPa. Nonetheless, these studies demonstrated that highly ordered organic films are indeed stiffer than amorphous ones.

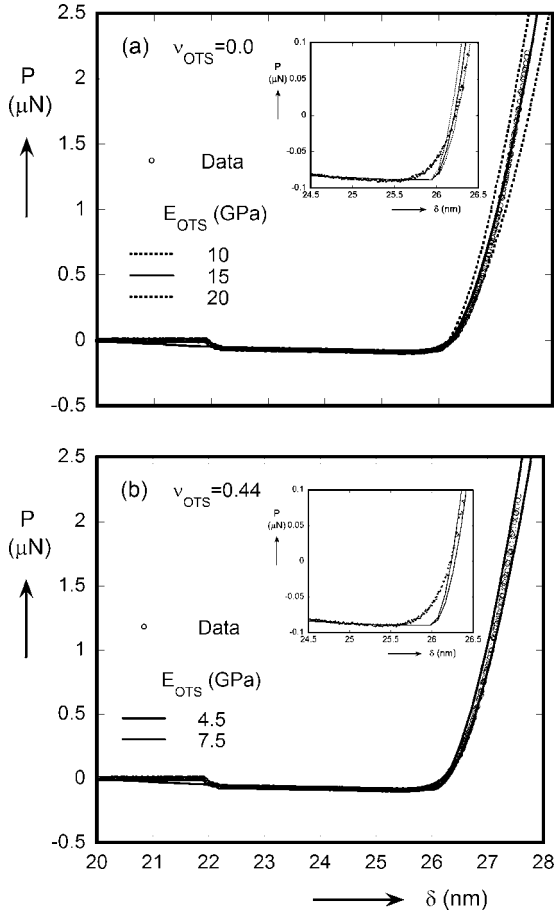


Fig. 4 FEM analysis of IFM experiments on OTS with different Poisson's ratio for OTS: (*a*) ν_{OTS}=0.0 and (*b*) ν_{OTS}=0.44. The insets show the disagreement between measurements and analyses at low force levels.

A closer inspection (insets in Fig. 4) of the response predicted by the linear analyses reveals that the overall fit was poor, particularly at low and high force levels. The linearly elastic analyses could not replicate the whole force profile. This suggested that the mechanical behavior of an OTS monolayer could be quite nonlinear and, therefore, a more complicated constitutive model was required. This was further motivated by the results from the molecular dynamics analyses that we now describe.

6.2 Molecular Dynamics Simulation. Molecular dynamics simulations were performed in this study in two stages. First, the self-assembled configuration of OTS on a silica surface was determined. This configuration was taken as the initial one for the subsequent simulation of compression. The steady-state compression of an OTS monolayer with a tungsten plate was then simulated. In the analysis, the tungsten plane moved down in increments of 0.1 Å. The computation was performed in two stages. The molecular dynamics simulation model was first annealed for 20 ps. This relaxation process allowed the equilibrium state to be achieved. The simulation was then continued for another 100 ps for data collection. The molecular configuration at the end of a step was taken as the initial configuration for the next step and the tungsten plane was moved downward another 0.1 Å. The process was repeated until the desired displacement (30 Å) of the tungsten plane was reached.

It is worth noting that, in the simulation, the effective compression rate is roughly 10^8 times faster than experiment. That is, each 0.1 Å increment in distance is associated with 120 ps of equilibration and data collection. Whether this compression rate can yield reasonably accurate results depends on the system studied. If the important configurations that contribute at each height can be sampled at this time, or those configurations sampled have properties substantially the same as all of those accessed in the experimental, ~1 sec time frame, the results should be valid. In the present case, the system is relatively well ordered, so that this is a favorable case. One does not expect large variations in molecular configuration of the sample in either simulation or experiment. Earlier simulations of compression nanoindentation [23–26] necessarily have the same behavior. One expects glassy materials, which are disordered but do not exhibit substantial structural rearrangement on the experimental time scale, to also be favorable cases, if an appropriate initial structure is used. Nonetheless, a study of rate effects, also beyond the scope of the present study, would be valuable.

The simulation of the self-assembled monolayer without external stress indicated that the molecular chains had a slight tilt angle relative to the normal to the silica surface. The average tilt angle was about 17 deg. Under load, the trajectory of the atoms was followed and the stresses were averaged. Figure 5 shows the computed relationships between the normal stresses and the nominal strain in compression (z direction). The simulation revealed a highly nonlinear response, even at low strain levels. Furthermore, besides the compressive stress in the loading direction, high lateral compressive stresses also existed. The difference between the two lateral stresses was negligibly small. As the deformation increased, these three compressive stresses became nearly equal, indicating an approach to incompressible behavior. Some kinks can also be seen in the response. These kinks may be related to changes in chain conformation. The nonlinear response is quite reminiscent of a rubbery material, and motivated the following continuum representation of the OTS constitutive behavior.

6.3 Nonlinear Elastic Analysis of the IFM Experiments. The nonlinear stress-strain behavior that was obtained from the molecular dynamics simulation was represented by a nonlinear elastic material model. At this stage, the OTS monolayer was still considered to be isotropic. This neglected the expected anisotropy associated with the highly ordered structure. The nonlinearity was accounted for by the hypoelastic material model. Based on the infinite lateral in-plane dimensions represented by the periodic

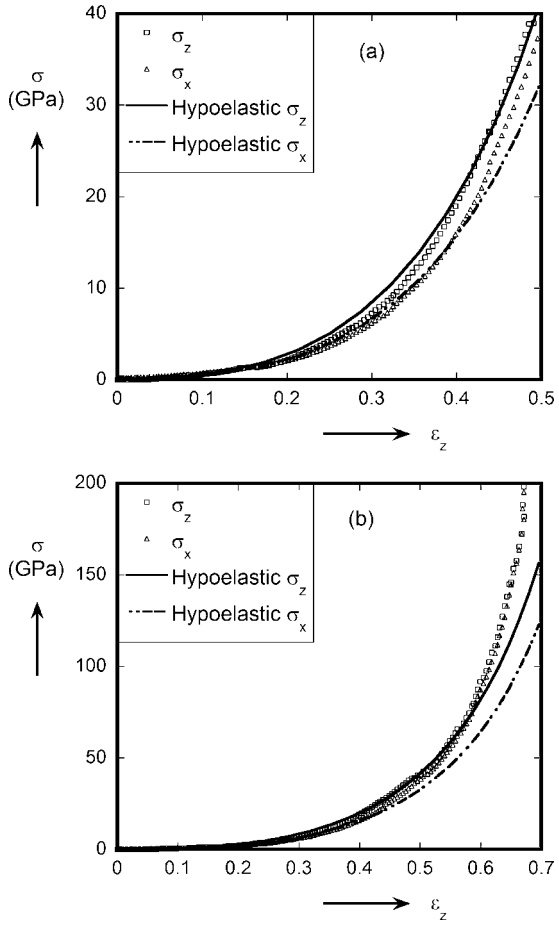


Fig. 5 Comparison of the stress-strain (nominal) behavior of OTS from the molecular dynamics simulation and the hypoelastic model: (*a*) low stress, (*b*) full range

boundary conditions in the molecular dynamics analysis, a uniaxial strain state exits where $\varepsilon_x=\varepsilon_y=0$. The constitutive equation for a hypoelastic material is [35]

$$d\varepsilon_i = \frac{1}{E_t}[d\sigma_i - \nu_t(d\sigma_j + d\sigma_k)] \tag{9}$$

where i, j, and k represent the x, y, and z directions and E_t, ν_t are tangent modulus and Poisson's ratio, respectively. Both E_t and ν_t are generally functions of the strain invariants I_1, I_2, and I_3. For a uniaxial strain, the tangent modulus and Poisson's ratio become

$$E_t = \frac{d\sigma_z}{d\varepsilon_z} - \nu_t\left(\frac{d\sigma_x}{d\varepsilon_z} + \frac{d\sigma_y}{d\varepsilon_z}\right) \tag{10}$$

and

$$\nu_t = \frac{\dfrac{d\sigma_x}{d\varepsilon_z}}{\dfrac{d\sigma_z}{d\varepsilon_z} + \dfrac{d\sigma_y}{d\varepsilon_z}} \tag{11}$$

Figure 6 shows the tangent modulus and Poisson's ratio derived from the results of the molecular dynamics simulation as a function of the logarithmic strain ε_z. The initial value of the tangent modulus (Fig. 6(a)) was about 1.5 GPa, which is close to that of typical polymers. As the deformation increased, the tangent modulus increased to 30 GPa. Figure 6(b) shows that the tangent Poisson's ratio did not change with ε_z. The values of the Poisson's ratio were scattered about an average value of 0.44. The following equations for the tangent modulus E_t were obtained from the fit

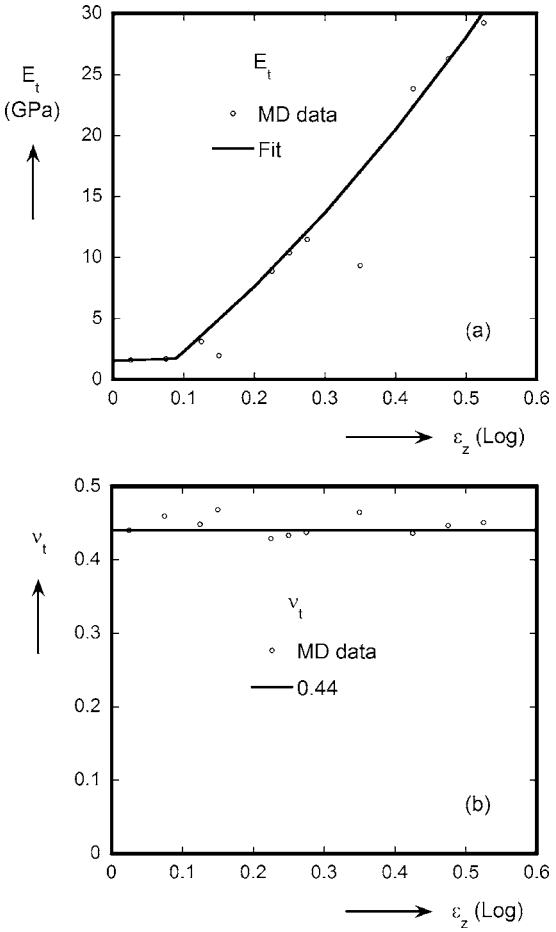


Fig. 6 Derivation of nonlinear material properties based on a uniaxial strain state: (*a*) tangent modulus, (*b*) tangent Poisson's ratio

$$E_t = \begin{cases} 1.5367 + 2.0504\varepsilon_z & |\varepsilon_z| < 0.089 \\ -2.2655 + 41.889\varepsilon_z + 37.625\varepsilon_z^2 & |\varepsilon_z| > 0.089 \end{cases} \quad \text{(GPa)}. \tag{12}$$

Note that some of the points that were obtained from the molecular dynamics simulation were ignored in the fit shown in Fig. 6(*a*). These points corresponded to the kinks in the response (Fig. 5).

Figure 5 also shows a comparison between the stress-strain behavior of OTS obtained from the hypoelastic model and the molecular dynamics simulation. The hypoelastic model fully reconstructed the stress-strain curve for nominal strains $\varepsilon_z < 0.4$, which was sufficient for simulating the IFM experiments.

This hypoelastic constitutive model was then incorporated into the finite element analysis of the IFM experiments. The other materials were still considered to be linearly elastic and isotropic. Figure 7 shows the comparison of this nonlinear elastic analysis with the data from the experiment. This simple nonlinear model provided a much closer match with the experiment than the linear analyses (Fig. 4). This was true at both low (insets of Figs. 4 and 7) and high force levels. The agreement between analysis and experiment in Fig. 7 is remarkable, given that the constitutive law that was used in the continuum model was based directly on molecular dynamics simulations without any adjustment. This result suggests that the behavior of the OTS is indeed simple enough that differences in the time scales of the molecular dynamics analyses and the actual experiments are not important. It appears that a class of problems has been opened up where spatial and temporal scales can be crossed in a relatively simple manner, and molecular dynamics analyses may be used to motivate continuum representations of self-assembled monolayers in a simple but direct manner.

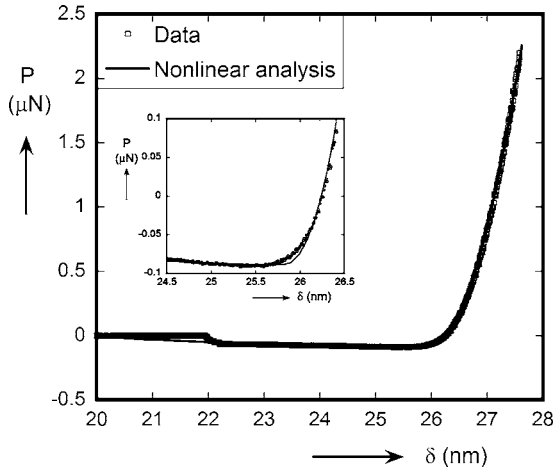


Fig. 7 Nonlinear finite element analysis of the IFM experiment on OTS

This result has a direct bearing on analyses of the frictional behavior of self-assembled monolayers where the contact area is required for the determination of the frictional strength. The usual approach [12] is to use the classical JKR contact mechanics analysis developed for contact of monolithic bodies or layered media with sufficiently thick layers that substrate effects are not important. The load versus contact radius response of the OTS monolayer was obtained from Hertz, DMT, and JKR analyses of the contact using a reduced modulus of 6 GPa and a Poisson's ratio of 0.44 for the OTS and 122.5 $\mathrm{mJ/m^2}$ for the work of adhesion between tungsten and OTS. These were the same parameters that were used in the linear elastic finite element analysis of the IFM experiment. The results of the classical contact mechanics analyses are compared (Fig. 8) with the contact radii that were extracted from the analyses that produced the force profiles in Fig. 7. The JKR contact mechanics analyses underestimated the contact radius by as much as 50%. The DMT and Hertz solutions differed even more from the finite element analyses. This means that frictional strengths derived from JKR analyses are about 50% higher than the actual values.

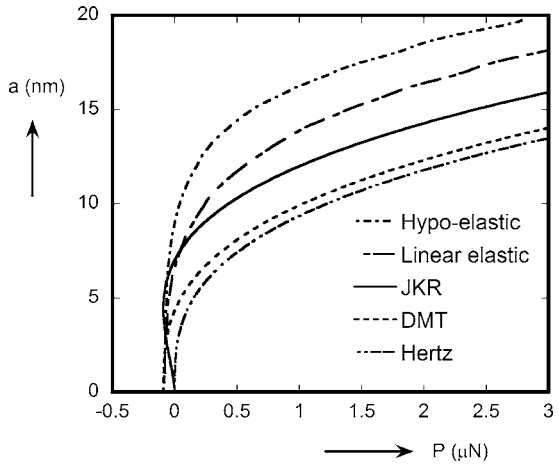


Fig. 8 Contact radii from classical contact mechanics and finite element analyses of an IFM probing of an OTS monolayer

7 Conclusions

The experiments conducted in this study demonstrated that the IFM has sufficiently high resolution in both force and displacement for conducting nanoindentation experiments on ultrathin films whose thickness is on the order of nanometers. The interpretation of the force profiles from such experiments is complicated by substrate effects and interface mismatch. The IFM experiments showed that the OTS monolayers were elastic, even when the indentation depth was relatively large. Linear elastic analyses of the IFM experiments suggested that the Young's modulus of OTS monolayers is relatively high. These high values are not surprising, given the highly ordered structure of the OTS monolayers. However, the linear elastic analyses were unable to fully match the measured force profiles. Molecular dynamics simulations were used to develop a nonlinear constitutive model for OTS, which could be represented as a hypoelastic material. The measured response was fully reproduced when the hypoelastic behavior was incorporated in the finite element analyses of the IFM experiments. This result suggests that representations of the behavior of self-assembled monolayers that are obtained from molecular simulations can be readily incorporated in continuum analyses. Properly accounting for the presence of the OTS increased contact radii by about 50% over classical contact mechanics estimates.

Acknowledgments

This work was supported by the National Science Foundation (Grant No. CMS9978678). P.J. Rossky acknowledges the support of the R.A. Welch Foundation. The authors would like to thank Dr. Jack Houston at Sandia National Laboratory for his generous assistance in setting up and operating the IFM facility at UT Austin. The authors would also like to thank Hibbitt, Karlsson & Sorensen, Inc. for the use of ABAQUS under academic license.

Appendix

Parameters Used in Molecular Dynamics Simulations

Atomic weight

Atoms	O	H	Si	CH_2	CH_3	W
Unit (amu)	15.9994	1.008	28.086	14.027	15.035	183.84

Harmonic bond potentials: $(U(r_{ij})=\frac{1}{2}k(r_{ij}-r_0)^2$ (r_{ij} is the distance between atoms labeled i and j).

i-j	k (kcal/mole)	r_0 (Å)
Si–O	772	1.64
O–H	1106	0.945
$Si-CH_2$	520	1.9
CH_2-CH_2	520	1.53
CH_2-CH_3	520	1.53

Harmonic valence angle potentials: $U(\theta_{jik})=(k/2)(\theta_{jik}-\theta_0)^2$ (θ_{jik} is the angle between bond vectors $\underline{r}_{ij}$ and $\underline{r}_{ik}$).

j-i-k	k (kcal/mole)	θ_0 (Å)
Si–O–H	110	122.9
CH_2-Si-O	160	114.9
O–Si–O	100	110.7
$Si-CH_2-CH_2$	126	120
$CH_2-CH_2-CH_2$	126	112.4
$CH_2-CH_2-CH_3$	126	112.4

Cosine dihedral angle potentials: $U(\phi)=(A_1/2)(1+\cos\phi)+(A_2/2)(1-\cos 2\phi)+(A_3/2)(1-\cos 3\phi)$ [ϕ is the dihedral angle defined by $\phi_{ijkn}=\cos^{-1}\{B(\underline{r}_{ij},\underline{r}_{jk},\underline{r}_{kn})\}$ with $B(\underline{r}_{ij},\underline{r}_{jk},\underline{r}_{kn})=[(\underline{r}_{ij}\times\underline{r}_{jk})(\underline{r}_{jk}\times\underline{r}_{kn})]/(|\underline{r}_{ij}\times\underline{r}_{jk}||\underline{r}_{jk}\times\underline{r}_{kn}|)$].

i-j-k-n	A_1 (kcal/mole)	A_2 (kcal/mole)	A_3 (kcal/mole)
$CH_2-Si-O-H$	0.3	0.0	1.3
$Si-CH_2-CH_2-CH_2$	−3.4	1.25	3.1
O–Si–O–H	0.0	0.0	0.0
H–O–Si–O	0.0	0.0	0.0
$O-Si-CH_2-CH_2$	−2.5	1.25	3.1
CH_2-CH_2-Si-O	−2,5	1.25	3.1
$CH_2-CH_2-CH_2-CH_2$	−3.4	1.25	3.1
$CH_2-CH_2-CH_2-CH_3$	−3.4	1.25	3.1

Intermolecular interactions were represented by the short-range (van der Waals) Lennard-Jones pair potentials: $U(r_{ij})=4\varepsilon[(\sigma/r_{ij})^{12}-(\sigma/r_{ij})^6]$.

i-j	ε (kcal/mole)	σ (Å)
W–Si	0.0707	3.6
W–W	0.05	3.2
W–O	0.092	3.1
$W-CH_2$	0.0768	3.5525
$W-CH_3$	0.0935	3.5525
O–O	0.17	3.0
Si–Si	0.1	4.0
Si–O	0.1303	3.5
CH_2-O	0.1416	3.4525
CH_2-Si	0.1086	3.9525
CH_2-CH_2	0.118	3.905
CH_2-CH_3	0.1437	3.905
CH_3-CH_3	0.175	3.905
CH_3-O	0.1725	3.4525
CH_3-Si	0.1323	3.9525

References

[1] Kojio, K., Ge, S., Takahara, A., and Kajiyama, T., 1998, "Molecular Aggregation State of N-Octadecyltrichlorosilane Monolayer Prepared at an Air/Water Interface," Langmuir, **14**, pp. 971–974.

[2] Ulman, A., 1991, *An Introduction to Ultra Thin Organic Films From Langmuir-Blodgett to Self-Assembly*, Academic, New York.

[3] Ulman, A., 1996, "Formation and Structure of Self-assembled Monolayers," Chem. Rev. (Washington, D.C.), **96**, 1533–1554.

[4] Zhuk, A. V., Evans, A. G., Hutchinson, J. W., and Whitesides, G. M., 1998, "The Adhesion Energy Between Polymer and Self-Assembled Monolayers," J. Mater. Res., **13**, 3555–3564.

[5] Reedy, E. D., Jr., Kent, M. S., and Moody, N. R., 2002, "On the Relationship Between the Molecular Work of Separation and Interfacial Fracture Toughness," *Proc. of 26th Annual Meeting of the Adhesion Society*, Myrtle Beach, SC, pp. 502–504.

[6] Mello, A. W., and Liechti, K. M., 2004, "Controlling Mixed-Mode Interfacial Fracture Toughness With Self-Assembled Monolayers," (submitted to ASME App. Mech.), Engineering Mechanics Research Laboratory Report EMRL #03-10.

[7] Maboudian, R., Ashurst, W. R., and Carraro, C., 2002, "Tribological Challenges in Micromechanical Systems," Tribol. Lett., **12**, 95–100.

[8] B. Bhushan, 1999, *Principles and Applications of Tribology*, Wiley, New York, p. 739.

[9] Mayer, T. M., Elam, J. W., George, S. M., Kotula, P. G., and Goeke, R. S., 2003, "Atomic-Layer Deposition of Wear-Resistant Coatings for Microelectromechanical Devices," Appl. Phys. Lett., **82**, 2883–2885.

[10] Romrig, A. D. Jr., Dugger, M. T., and McWhorter, P. J., 2003, "Materials Issues in Microelectromechanical Devices: Science, Engineering, Manufacturability and Reliability," Acta Mater., **51**, 5837–5866.

[11] Houston, J. E., and Kim, T., 2000, "Separating Mechanical and Chemical Contributions to Molecular-Level Friction," J. Am. Chem. Soc., **122**, 12045–12046.

[12] Carpick, R. W., Agraït, N., Ogletree, D. F., and Salmeron, M., 1996, "Variation of the Interfacial Shear Strength and Adhesion of a Nanometer-Sized Contact," Langmuir, **12**, 3334–3340.

[13] Winter, R. M., and Houston, J. E., 1998(a), "Interphase Mechanical Properties in an Epoxy-Glass Fiber Composites as Measured by Interfacial Microscopy," in *Proc. of SEM Spring Conf. on Experimental and Applied Mechanics*, Houston.

[14] Winter, R. M., and Houston, J. E., 1998(b), "Nanomechanical Properties of the Interphase in Polymer Composites as Measured by Interfacial Force Microscopy," Materials Research Society Spring Meeting, San Francisco.
[15] Thomas, R. C., Houston, J. E., Crooks, R. M., Kim, T., and Michalske, T. A., 1995, "Probing Adhesion Forces at the Molecular Scale," J. Am. Chem. Soc., **117**(13), 3830–3834.
[16] Cabibil, H., Celio, H., Lozano, J., and White, J. M., 2001, "Nanomechanical Properties of Polysiloxane-Oxide Interphases Measured by Interfacial Force Microscopy," Langmuir, **17**, 2160–2166.
[17] Wang, M., Liechti, K. M., and White, J. M., 2004, "Nanoindentation of Polymeric Thin Films With an Interfacial Force Microscope," J. Mech. Phys. Solids, **52**, 2329–2354.
[18] Joyce, S. A., and Houston, J. E., 1991, "A New Force Sensor Incorporating Force-Feedback Control for Interfacial Force Microscopy," Rev. Sci. Instrum., **62**(3), 710–715.
[19] Houston, J. E., and Michalske, T. A., 1992, "The Interfacial-Force Microscopy," Nature (London), **356**, 266–267.
[20] Johnson, K. L., Kendall, K., and Roberts, A. D., 1971, "Surface Energy and the Contact of Elastic Solids," Proc. R. Soc. London, Ser. A, **324**, 301–313.
[21] Derjaguin, B. V., Muller, V. M., and Toporov, Y. P., 1975, "Effect of Contact on the Adhesion of Particles," J. Colloid Interface Sci., **53**, 314–326.
[22] Maugis, D., 1992, "Adhesion of Spheres: The JKR-DMT Transition Using a Dugdale Model," J. Colloid Interface Sci., **150**, 243–269.
[23] Siepman, J. L., and McDonald, I. R., 1993, "Monte Carlo Simulation of the Mechanical Relaxation of a Self-Assembled Monolayer," Phys. Rev. Lett., **70**(4), 453–456.
[24] Tupper, K. J., Colton, R. J., and Brenner, D. W. 1994, "Simulations of Self-Assembled Monolayers Under Compression: Effect of Surface Asperities," Langmuir, **10**, 2041–2043.
[25] Tupper, K. J., and Brenner, D. W., 1994, "Compression-Induced Structural Transition in a Self-Assembled Monolayer," Langmuir, **10**, 2335–2338.
[26] Henda, R., Grunze, M., and Pertsin, A. J., 1998, "Static Energy Calculations of Stress-Strain Behavior of Self-Assembled Monolayers," Tribol. Lett., **5**, 191–195.
[27] Wang, M., Liechti, K. M., Wang, Q., and White, J. M., 2004, "Self-Assembled Silane Monolayers: Fabrication With Nanoscale Uniformity," Langmuir, **21**, pp. 1848–1857.
[28] Smith, W., and Forester, T. R., 1996, "DL_POLY_2.0: A General-Purpose Parallel Molecular Dynamics Simulation Package," J. Mol. Graphics, **14**(3), 136–141.
[29] Stevens, M. J., 1999, "Thoughts on the Structure of Alkylsilane Monolayers," Langmuir, **15**, 2773–2778.
[30] Allen, M. P., and Tildesley, D. J., 2002, *Computer Simulations of Liquids*, Clarendon Press, Oxford.
[31] Frenkel, D., and Smit, B., 2002, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd Ed., Academic Press, San Diego.
[32] Jorgensen, W. L., Madura, J. D., and Swenson, C. J., 1984, "Optimized Intermolecular Potential Functions for Liquid Hydrocarbons," J. Am. Chem. Soc., **106**(22), 6638–6646.
[33] Jorgensen, W. L., and Tirado-Rives, J., 1988, "The OPLS Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin," J. Am. Chem. Soc., **110**(6), 1657–1666.
[34] Jorgensen, W. L., 1986, "Optimized Intermolecular Potential Functions for Liquid Alcohols," J. Phys. Chem., **90**(7), 1276–1284.
[35] *ABAQUS/Standard User's Manual, Version 6.4*, Hibbitt, Karlsson and Sorensen, Inc.
[36] Simmons, G., and Wang, H., 1971, *Single Crystal Elastic Constants and Calculated Aggregate Properties: A Handbook*, 2nd ed., M.I.T. Press, Cambridge, MA.

Influence of Confining Pressure on the Crack Growth Behavior in a Highly Filled Elastomer

C. T. Liu
Materials Technology Center,
Southern Illinois University,
Carbondale, IL 62901

G. Ravichandran
Graduate Aeronautical Laboratories,
California Institute of Technology,
Pasadena, CA 91125

In this study, the effect of confining pressure on the crack growth behavior in a highly filled elastomer was investigated. The material under investigation contains 86% by weight of hard particles embedded in a rubbery matrix. The specimens were tested at a constant strain rate under a confining pressure. Two confining pressures, 3.45 MPa and 6.9 MPa, and two types of loading, monotonically increasing and cyclic, were considered. The experimental data were analyzed, and the effects of confining pressure on the damage process as well as the constitutive and crack growth behavior in the material are discussed. [DOI: 10.1115/1.2189874]

1 Introduction

An important engineering problem in structural design is evaluating its integrity and reliability. It is well known that structural strength may be degraded during its design life due to mechanical or chemical aging, or a combination of these two aging mechanisms. Depending on the structural design, material type, service loading, and environmental condition, the cause and degree of strength degradation due to the different aging mechanisms differs. One of the common causes of strength degradation is the result of damage and crack development in the structure. Therefore, to effectively use the material in structural applications one needs to understand the damage initiation and evolution processes, the effects of damage and crack development on the material's response, and the remaining strength and life of the structures.

In recent years, a considerable amount of work has been done studying crack growth behavior in particulate composite materials under different loading conditions at ambient pressure [1–4]. Experimental findings indicate that power-law relationships exist between the crack growth rate, da/dt, and the mode I stress intensity factor K_I. These experimental findings support the theory developed by Knauss [5] and Schapery [6] in their studies of crack growth behavior in linear viscoelastic materials. It is known that classical fracture mechanics principles, especially linear elastic fracture mechanics, are well established for single-phase materials. Experimental data indicate that linear fracture mechanics theories can be applied to the particulate composite materials with varying degrees of success. However, there has been relatively little effort in understanding the crack growth behavior in such materials under confining pressure condition.

In this study, uniaxial specimens with and without precrack were used to investigate the effects of confining pressure on the constitutive and crack growth behavior in a highly filled elastomer. The material contains 86% by weight of hard particles embedded in a rubbery matrix. The specimens were tested under confining pressure 3.45 MPa and 6.9 MPa. The specimens without precrack were tested at 16.6 cm/cm/min and 0.073 cm/cm/min for monotonically increasing loading and cyclic loading, respectively. The specimens with precrack were tested under monotonically increasing loading at a strain rate of 16.7 cm/cm/min only. The experimental data were analyzed, and the effects of confining pressure on the damage process as well as the constitutive and crack growth behavior in the material are discussed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received January 18, 2005; final manuscript received January 20, 2006. Review conducted by K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Experiments

In this study, a series of experiments were conducted on uniaxial specimens with and without precrack under a confining pressure. Two different confining pressures, 3.45 MPa and 6.9 MPa, were considered. The specimens were made of a highly filled polymeric material, containing 86% by weight of hard particles embedded in a rubbery matrix. The dimensions of the specimen were 7.6 cm long, 2.54 cm wide, and 5.08 cm thick (Fig. 1). For the precracked specimens, a 0.762 cm crack was cut in the middle at one of the long edges of the specimen with a razor blade. Prior to conducting the tests, the specimen was loaded in a materials testing machine inside a pressure chamber. When the pressure inside the pressure chamber reached the predetermined confining pressure, the specimen was deformed at a constant strain rate of 16.7 cm/cm/min until the specimen fractured. In addition, two cyclic loading tests were also conducted to investigate the effect of confining pressure on the damage state and the constitutive behavior of the material. For the cyclic loading tests, the specimen was strained to 10% strain level, and then it was unloaded to zero stress. After waiting for 15 min at zero stress, the specimen was reloaded until fractured. For the cyclic loading tests, the specimens were deformed at a constant strain rate of 0.073 cm/cm/min. For one of the two cyclic loading tests, the specimen was tested under 6.9 MPa confining pressure, whereas, for the other one, the specimen was loaded and unloaded under 6.9 MPa confining pressure but reloaded under ambient pressure. During the deformation of the specimen, a high-speed camera was used to monitor the test. In addition, the load and time were also recorded. The raw data were used to determine the stress, the strain, mode I stress intensity factor, and crack growth rate.

To determine the stress intensity factor at the crack tip, it is necessary to relate the load on the specimen to crack and specimen dimension. A three-dimensional finite element computer code TEXGAP-3D was used to determine the stress intensity as a function of the crack lengths for a given load P applied at the boundary of the specimen. The calculated mode I stress intensity factor K_I was normalized with respect to the applied load P. A linear regression analysis was conducted to determine the relationship between the normalized mode I stress intensity factor K_I/P to the crack length a. This relationship can be used to determine the mode I stress intensity factor for elastic and viscoelastic materials. In calculat-

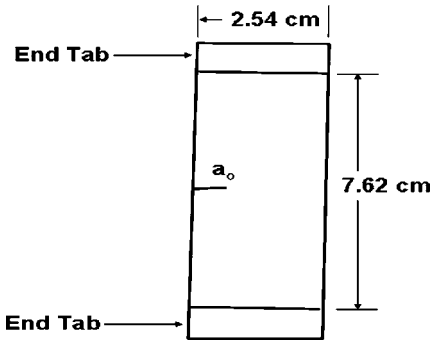


Fig. 1 Specimen geometry

ing K_I from experimental data, the recorded data, a and p for a given time, and the regression equation relating K_I/P and a were used.

To determine the crack growth rate da/dt, the polynomial method was used. This method involved fitting an nth order polynomial to a set of crack length versus time data. The coefficients of the polynomial function were estimated by the method of least squares. The crack growth rate was calculated by taking derivative of the polynomial function at a given time.

3 Results and Discussion

It is well known that, on the microscopic scale, a highly filled polymeric material can be considered to be an inhomogeneous material. When these materials are stretched, the different sizes and distribution of filled particles, the different cross-linking density of polymeric chains, and the variation in bond strength between the particles and the binder can produce highly inhomogeneous local stress and strength fields. Also, the material may contain randomly distributed microvoids, incipient damage sites, and microcracks with statistically distributed size and orientation. Therefore, the local strength in the material varies in a random fashion, so the failure sites in the material do not necessarily coincide with the maximum stress location. Hence, the failure location as well as the degree of damage induced in the material will also vary in a random fashion. Depending on the magnitude of the local stress and the local strength, damage can be developed in the material, especially near the crack tip region. The damage developed in the material may be in the form of microvoids or microcracks in the binder or dewetting between the binder and the filler particles. When the particle is dewetted, the local stress will be redistributed. With time, additional binder/particle separation and vacuole formation take place. The damage growth in the material may take place as material tearing or as successive nucleation and coalescence of the microcracks. These damage initiation and evolution processes are time dependent and are the main factor responsible for the time-dependent constitutive and fracture behavior of the material. It should be pointed that the aforementioned damage initiation and evolution processes are commonly observed in the particulate composite material studied when the material is subjected to a monotonically increasing load under low confining pressure. However, when the confining pressure is very high, the damage mechanisms may change, which is a subject of study. Having discussed the damage mechanisms in the particulate composite material, one can now discuss how the damage state affects the constitutive behavior of the material.

A typical plot of stress-strain curves, obtained from tests on specimens without precrack at a constant strain rate of 16.7 cm/cm/min under ambient, 3.45 MPa, and 6.90 MPa confining pressures, are shown in Fig. 2. From Fig. 2, it is seen that the maximum load increase significantly when the pressure is increased from ambient to 6.90 MPa. However, the magnitude of the confining pressure has no significant effect on effective modulus. Theoretically, the magnitude of the confining pressure should have no effect on the Young's modulus due to the incompressibility of the material. The slight variation of the Young's modulus is due to the scatter of the test data, material variability, and the reason that the material is not truly incompressible. It is interesting to note that under the high confining pressure condition, the critical strain ε_c for the transition of the linear response to the nonlinear response of the material is increased. Since the material's response is closely related to the damage state in the material, the change of the aforementioned parameters is believed due to the suppression, or the delay, of the development of damage under the high confining pressure.

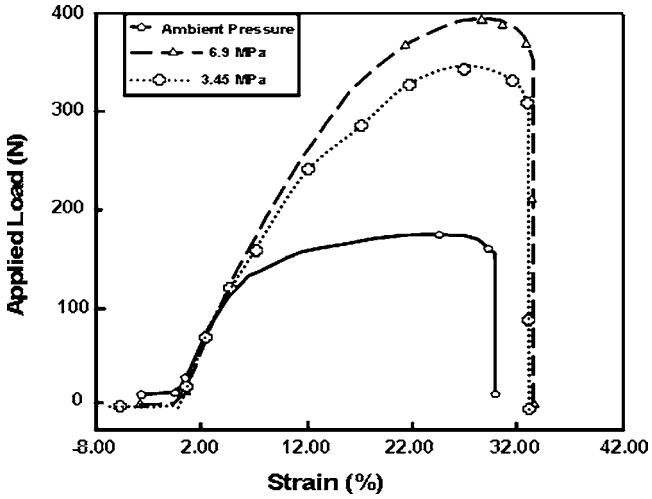


Fig. 2 Applied load versus applied strain under different confining pressure

Throughout the loading history, the progressive development and interaction of various damage modes change the state of the material or the mechanical response of the material. In general, when the highly filled elastomer is deformed under constant strain-rate condition, the initial linear portion of the stress-strain curve is associated with stretching undamaged material, when the filler particles are bonded to the binder. As the external load is continuously increased, at a certain critical stress/strain level, dewetting occurs. When the density of the dewetted particles reaches a critical value, the rigidity of the material is thereby reduced, and usually this critical dewetting state coincides with the transition from linear response to nonlinear behavior. As the specimen is continuously stretched, the number of dewetted particles is increased and the voids that are formed begin to grow and coalesce. This damage process is related, primarily, to the nonlinear response of the material and can be characterized by bulk volume change during stretching. The bulk volume change during straining is usually known as the volume dilatation, which is partially caused by the nucleation of new voids and partially caused by the growth of the existing voids. The extent of the volume dilatation depends on the nature of the binder/particle system, the testing temperature, and the strain rate. The volume dilatation is used as a damage parameter and has been incorporated into a constitutive model [7], which, in turn, has been implemented in a deformation-based finite element simulation. The critical volume dilatation is used as a failure criterion in this phenomenological model for predicting the crack growth behavior in filled elastomers [8].

A typical plot of volume dilatation, measured by dilatometer, versus applied strain under different confining pressures is shown in Fig. 3. It is noted that for a given applied strain, the volume dilatation decreases as the confining pressure is increased. At 6.9 MPa confini ng pressure, the dilatation is very small. In order to see whether or not damage can develop in the material under

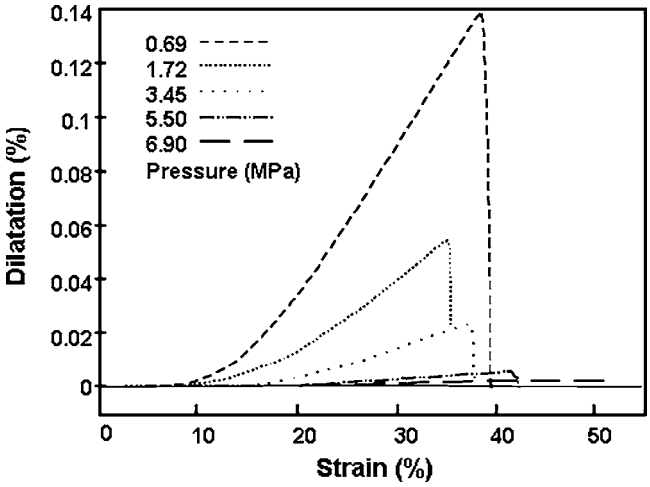


Fig. 3 Volume dilatation versus applied strain under different confining pressure

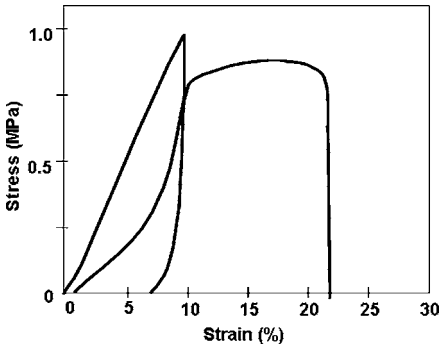


Fig. 5 Stress-strain curves under cyclic loading (loading and unloading under 6.90 MPa confining pressure, reloading under ambient pressure)

6.9 MPa confining pressure, cyclic loading experiments were conducted under 6.9 MPa confining pressure, and the results are shown in Figs. 4 and 5. In addition to plotting the cyclic stress-strain curve, the stress-strain curve under a monotonic increasing loading condition with the same strain rate as the cyclic loading test is also plotted. From Fig. 4, it is seen that the slope of the initial part of the reloading curve is smaller than that of the loading curve corresponding to that of the cyclic loading. In other wards, after the cyclic loading, the modulus of the material is decreased, implying that, under 6.9 MPa confining pressure, the material can be damaged by the cyclic load. It is interesting to note that the reloading stress-strain curve rejoins the constant strain rate stress-strain curve, indicating that the cyclic loading considered in this study has no significant effect on the tensile strength of the material. The development of damage under confining pressure is further illustrated in Fig. 5, which shows the loading and unloading stress-strain curves under 6.9 MPa confining pressure and the reloading stress-strain curve under ambient pressure. It is noted that (similar to Fig. 4) the modulus of the material, after the specimen is subjected to the cyclic load under 6.9 MPa confining pressure, is decreased. However, the maximum stress is lower than that under the previous cyclic load. Based on the experimental findings, it is conjectured that under confining pressure, microcracks are developed near the particle surface where high triaxial tensile stresses are developed due to the high constraint developed near the particle surface. Under the high confining pressure condition, the microcracks cannot open up to form voids, resulting in very small volume dilatation. When the specimen is reloaded under ambient pressure, the microcracks open up and grow. The growth of the microcracks results in the fracture of the specimen and a lower maximum stress.

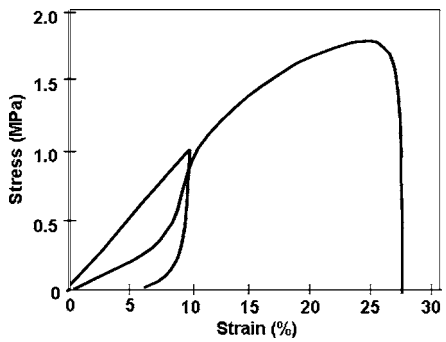


Fig. 4 Stress-strain curves under cyclic loading (loading, unloading, and reloading under 6.90 MPa confining pressure)

The above paragraphs discussed the effect of confining pressure on the constitutive behavior and damage mechanisms in the highly filled elastomer. In the following paragraphs, we will discuss the effect of confining pressure on the crack growth behavior in the material

Experimental results indicate that crack tip blunting takes place both before and after crack growth. The material at the tip of the crack undergoes very large elongation, and the crack tip is nearly straight, indicating extensive blunting. The highly strained or damage zones extends ahead of the crack tip, appearing as an equilateral triangle with the crack tip as its base. This damage zone is known as the failure process zone, which is a key parameter in viscoelastic fracture mechanics under mode I type loading. When the local strain reaches a critical value, small voids are generated in the failure process zone. Because of the random nature of the microstructure, the first void is not restricted to the surface where the maximum normal strain occurs. Since the tendency of the filler particle to separate from the binder under a triaxial loading condition is high, it is expected that voids, or a damage zone, will also be generated in the specimen's interior. Consequently, there are a large number of strands, essentially made of the rubbery binder material, which separate the voids that form inside the failure process zone. Under this condition, the transverse constraint is minimized, approaching a state of plane stress condition. As the applied strain increases with time, material fracture occurs at the blunted end of the crack tip. This will always be the location of the maximum local strain. The failure of the material between the void and the crack tip causes the crack to grow into the failure process zone. This kind of crack growth mechanism continues until the main crack tip reaches the front of the failure process zone. When this occurs, the crack tip resharpens temporarily prior to accumulative of damage and blunting

The damage and crack growth mechanisms discussed above are the basic mechanisms observed in this material under both ambient and 6.9 MPa pressures. The effect of pressure is to suppress the damage and evolution processes and delay the onset of crack growth.

A plot of crack length versus time is shown in Fig. 6, in which the maximum applied load is also marked on the crack growth curve. It is interesting to point out that under confining pressure, the crack grows stably not only beyond the maximum applied stress but also near the break of the specimen. This phenomenon is significantly different from the crack growth behavior of the same material under ambient pressure. Under ambient pressure, crack growth becomes unstable when the applied load exceeds the maximum applied load.

The determination of the crack growth rate requires an analysis of discrete data relating the instantaneous time t to the corre-

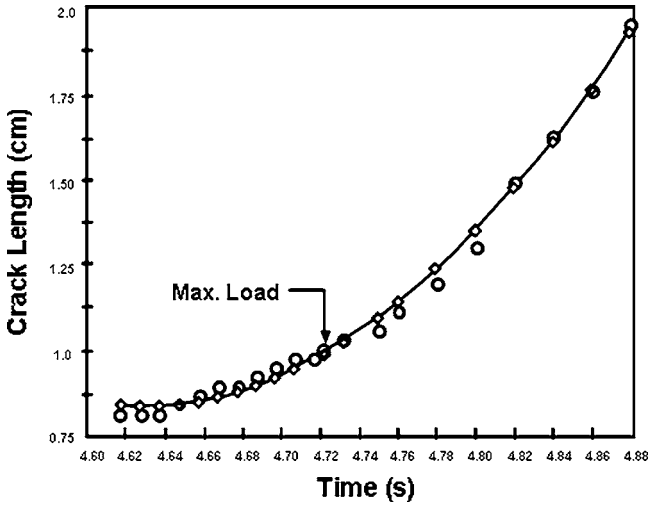


Fig. 6 Crack length versus time curve

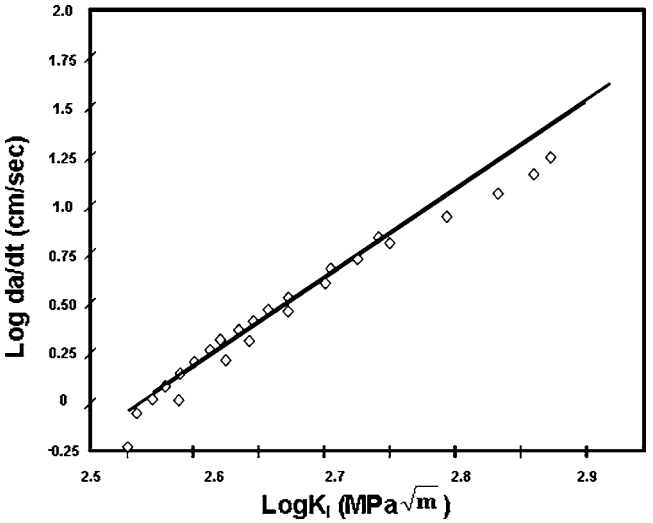


Fig. 7 Crack growth rate versus mode I stress intensity factor (pressure=3.45 MPa)

sponding crack length a. Because of the inhomogeneous nature of the highly filled elastomer the measured data showed a considerable scatter. Therefore, it was anticipated that a smooth and steadily increasing relationship between the crack growth rate and time was difficult to obtain, and different methods of calculating da/dt calculation might result in different quantities. As mentioned earlier, the polynomial method was used to calculate the crack growth rate in this study. The smooth action introduced by the polynomial method results in a continuous smooth crack velocity curve. There is another method, known as the secant method [9], that can be used to calculate the crack growth rate. In the secant method, the crack growth rate was computed by calculating the slope of a straight line connecting two adjacent a versus t data points and assigning the average crack growth rate to the middle point between each pair of data points. Based on our experience, the secant method introduces a pronounced fluctuation in the crack growth rate da/dt. In other words, the crack growth process consists of a slow-fast-slow phenomenon. As mentioned earlier, the damage process is a time-dependent process and requires finite time to develop a failure process zone at the crack tip. Thus, the crack growth process consists of blunt-growth-blunt and slow-fast-slow phenomena, which is highly nonlinear. The fluctuation in crack growth rate, da/dt, is consistent with experimental observation. Based on experimental evidence, in general, the crack does not grow in a continuous and smooth manner. During the crack growth process, crack growth rate both accelerates and decelerates. Therefore, the secant method appears to provide the best emulation of the crack growth process and the actual crack growth rate. It is important and interesting to point out that, in general, the crack growth rate calculated by the secant method oscillate approximately around the smooth crack growth rate curve obtained by the polynomial method. Therefore, it is reasonable to expect that the crack growth velocity calculated by the polynomial method may represent the mean crack growth rates with the crack growth rate calculated by the secant method oscillating around it.

When the crack growth rate fluctuates, there will be no one-to-one correspondence between the crack growth velocity and the stress intensity factor. This contradicts current crack growth theory for viscoelastic materials, which requires a unique relationship between the crack growth rate and the stress intensity factor. The lack of uniqueness between the crack growth rate and the stress intensity factor raises two questions: to what extent should the oscillation of the crack growth rate be taken into account and how to present and use these data. How these questions are answered depends on how the gathered data will be used. If one is interested in knowing the central tendency of the crack growth behavior or comparing the average crack growth behavior under different loading conditions, the fluctuation in the crack growth rate seems unimportant and it recommends using the polynomial method to calculate the crack growth rate. However, if one is interested in determining the upper-bound limit on the crack growth rate, the fluctuation in the crack growth rate should be considered in the data analysis and the secant method should be used to calculate the crack growth rate.

Typical plots of the crack growth rate da/dt versus the stress intensity factor K_I for different confining pressures are shown in Figs. 7 and 8. According to Figs. 7 and 8, power-law relationships appear to exist between the crack growth rate and the mode I stress intensity factor, which support the theories developed by Knauss [5] and Schapery [6] in their studies of crack growth behavior in linear viscoelastic materials. The effect of confining pressure on the crack growth behavior can be determined by comparing Fig. 7 to Fig. 8. It is noted that crack grows slower under a higher confining pressure, for reason discussed earlier.

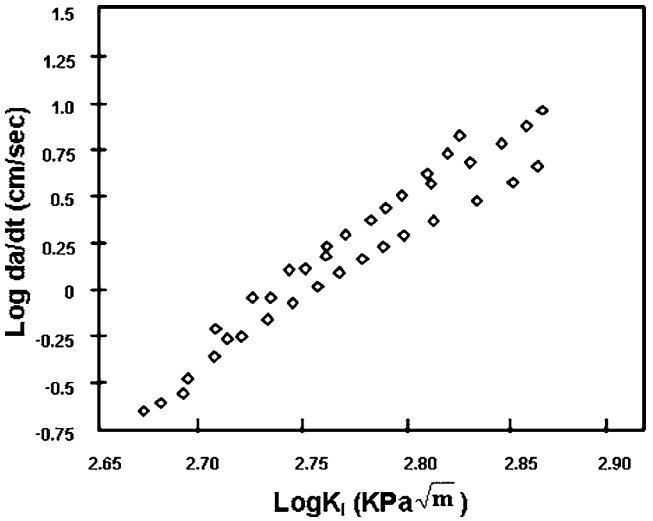


Fig. 8 Crack growth rate versus mode I stress intensity factor (pressure=6.90 MPa)

4 Numerical Modeling

Stress-strain response and crack initiation at room temperature in the highly filled elastomer was analyzed at different confining pressures using a modified version of a phenomenological damage model developed earlier [7] and accounting for time dependence of the homogenized material. Experimental observations, based on dilatation tests conducted under a different program, show that the volume dilatation Θ (damage parameter in RL model [7]) is independent of the loading rate and purely a function of applied strain. In other words, volume dilatation is dependent only on mechanical deformation (strain), and hence, damage is not a time-dependent process. Taking these observations into account, the material time dependence is introduced through a simple Maxwell model for the undamaged material, and it is assumed that both the bulk and shear response of the material are governed by the same relaxation time τ. The undamaged material's bulk (κ_0) and shear (μ_0) moduli are given in terms of its rubbery moduli, κ_∞ and μ_∞ and Prony coefficients κ_1 and μ_1

$$\kappa_0(t) = \kappa_\infty + \kappa_1 \exp(-t/\tau) \text{ and } \mu_0(t) = \mu_\infty + \mu_1 \exp(-t/\tau) \quad (1)$$

Stress components σ_{ij} are evaluated using the convolution integral

$$\sigma_{ij}(t) = \int_0^t \kappa_0(t-t')g(\Theta)\dot{\Theta}\,\delta_{ij}\,dt' + \int_0^t 2\mu_0(t-t')h(\Theta)\dot{\varepsilon}_{ij}\,dt' \quad (2)$$

where g and h are the damage parameters, which are purely functions of the volume dilatation.

The viscoelastic material parameters in Eq. (1) are obtained by fitting uniaxial stress-strain data at two different loading rates. The modified viscoelastic model has been implemented in a special purpose displacement based finite element code FEAP-SP. The damage parameters governing the degradation of the bulk and the shear moduli, g and h, are determined using the uniaxial response and dilatation data as outlined in the original model [7]. The finite element method uses the Newton-Raphson iterative procedure for achieving force equilibrium and Newton's iterative method for enforcing plane stress condition. Constitutive update (stress, dilatation) is performed through integration of Eq. (2) for a finite time step. Time steps are chosen to preserve accuracy in the constitutive update and to ensure quadratic convergence of the residual energy norm.

Crack initiation is simulated in specimens with edge cracks subjected to prescribed displacement rates at the boundaries. Because of the inherent nature of the singular stress fields near the crack tip, the loading rate varies by at least two orders of magnitude from the loading boundary to the crack tip. The relaxation and damage parameters determined from the uniaxial stress-strain curves are used in the finite element calculations to update the local moduli and the damage parameter. Crack initiation is determined by the attainment of critical dilatation at a critical distance from the crack tip and is used as the failure criterion [8]. The finite element model continues check at the end of every time step elements that have attained the specified critical dilatation.

The simulations of crack initiation were carried out in the material at the two different confining pressures, 3.45 MPa and 6.9 MPa. Displacements were specified at the specimen boundaries normal to the crack to maintain a specified loading rate of 16.7 cm/cm/min. The load increased nearly linearly with increasing displacement with the crack blunting monotonically. Upon reaching the critical dilatation in the crack tip vicinity, the crack was initiated. The predicted (measured) critical value for the stress intensity factor (K_{Ic}) for the onset of crack growth that were obtained for the two different confining pressures 3.45 MPa and 6.9 MPa are 0.220 (0.230) and 0.238 (0.243) MPa $m^{1/2}$, respectively. The good correlation between the measured and the predicted K_{Ic} validates the critical damage based simulations on the onset of crack growth.

5 Conclusions

In this study, the effects of confining pressure on the crack growth behavior were investigated. Experimental findings reveal that under confining pressure, cracks grow stably near the fracture of the specimen. They also reveal that, for a given mode I stress intensity factor, the crack growth rate decreases as the confining pressure is increased and a power-law relationship exists between the crack growth rate and the mode I stress intensity factor. In addition, a good correlation exists between the predicted and the measured K_{Ic} for the onset of crack growth. Further investigations concerning the mechanics of the role of pressure on damage initiation and growth in highly filled elastomers is under way.

References

[1] Beckwith, S. W., and Wang, D. T., 1978, "Crack Propagation in Double-Base Propellants," AIAA Paper No. 78–170.

[2] Liu, C. T., 1990, "Crack Growth Behavior in a Composite Propellant With Strain Gradients—Part II," J. Spacecr. Rockets, **27**, pp. 647–659.

[3] Liu, C. T., 1990, "Crack Propagation in a Composite Solid Propellant," *Proceedings of the Society of Experimental Mechanics*, Spring Conference, SEM, Bethel, CT, pp. 614–620.

[4] Liu, C. T., and Smith, C. W., 1996, "Temperature and Rate Effects on Stable Crack Growth in a Particulate Composite Material," Exp. Mech., **36**(3), pp. 290–295.

[5] Knauss, W. G., 1970, "Delayed Failure—The Griffith Problem for Linearly Viscoelastic Materials," Int. J. Fract. Mech., **6**, pp. 7–20.

[6] Schapery, R. A., 1973, "On a Theory of Crack Growth in Viscoelastic Media," Report No. MM 2765–73–1, Texas A&M University, College Station.

[7] Ravichandran, G., and Liu, C. T., 1995, "Modeling Constitutive Behavior of Particulate Composites Undergoing Damage," Int. J. Solids Struct., **32**, pp. 979–990.

[8] Ravichandran, G., and Liu, C. T., 1998, "Crack-Tip Shielding in Particulate Composites Undergoing Damage," Eng. Fract. Mech., **59**, pp. 713–723.

[9] Liu, C. T., 1990, "Critical Analysis of Crack Growth Data," J. Propul. Power, **6**(5), pp. 519–524.

Soma Sekhar V. Kandula[1]
Department of Aerospace Engineering,
University of Illinois at Urbana-Champaign,
306 Talbot Lab,
104 S. Wright Street Urbana, IL 61801
e-mail: skandula@uiuc.edu

Jorge Abanto-Bueno
Department of Mechanical Engineering,
Bradley University,
1501 W. Bradley Avenue Peoria, IL 61625

John Lambros

Philippe H. Geubelle

Department of Aerospace Engineering,
University of Illinois at Urbana-Champaign,
306 Talbot Lab,
104 S. Wright Street Urbana, IL 61801

Cohesive Modeling of Quasi-Static Fracture in Functionally Graded Materials

A spatially varying cohesive failure model is used to simulate quasi-static fracture in functionally graded polymers. A key aspect of this paper is that all mechanical properties and cohesive parameters entering the analysis are derived experimentally from full-scale fracture tests allowing for a fit of only the shape of the cohesive law to experimental data. The paper also summarizes the semi-implicit implementation of the cohesive model into a cohesive-volumetric finite element framework used to predict the quasi-static crack initiation and subsequent propagation in the presence of material gradients.
[DOI: 10.1115/1.2151210]

1 Introduction

Functionally graded materials (FGMs) are a class of nonhomogenous materials characterized by spatially varying material properties. Since their use by Kawasaki and Watanabe [1] to overcome problems associated with direct bonding of ceramic tiles to a metallic substrate, FGMs have been applied to thermal barrier coatings, wear-resistant coatings on gears, ultratough metal-cutting inserts, bearings, fiber optics, and shaving razor blades [2]. Regardless of application, fracture is often a key concern in the design of FGM components and has motivated many contributions in the field of fracture mechanics of FGMs over the past two decades. Theoretical analyses of stress and deformation fields around a crack tip in FGMs have been carried out among others by Erdogan and Wu [3] and Jin and Batra [4]. Jin and Noda [5] concluded that the leading term of the asymptotic crack-tip field in an FGM specimen possessing a continuous or piecewise differentiable material property variation is identical to that in homogeneous materials with properties evaluated at the crack tip, implying that the definition of the stress intensity factor can be extended to FGMs. Delale and Erdogan [6] showed that the effect of Poisson's ratio on the fracture response of a FGM is negligible. Recently, Wang et al. [7] developed an analytical model for fracture mechanics analysis of FGMs with material property variations representing arbitrarily distributed properties.

Fracture experiments in FGMs have been relatively scarce. Difficulty in fabricating metal-ceramic FGM specimens suitable for fracture testing has led many investigators to develop model FGMs, such as a glass-epoxy FGM composite manufactured using a gravity casting technique [8], a cenosphere-polyester FGM obtained using a buoyancy-assisted casting process [9], or a polyester-based FGM with varying amounts of plasticizer [10]. Of particular interest in this research is a model FGM based on polyethylene 1% carbon monoxide co-polymer (ECO), manufactured by selective exposure to ultraviolet (UV) irradiation [11]. The specimens described in [11] have material characteristics mimicking ceramic-metal FGMs, i.e., stiffer and more brittle at one end, becoming gradually less stiff and more ductile at the other. Cracks in these FGM ECO specimens have been analyzed by Li et al. [12] using a hybrid numerical/experimental technique to obtain the crack-resistance curves, and by Abanto-Bueno and Lambros [13] using the digital image correlation technique to extract characteristics of the near-tip fields. These latter experiments form the basis for the validation exercise presented in this paper.

Difficulty in extending the theoretical and experimental analyses to more general FGMs has led many researchers to adopt numerical techniques, including cohesive modeling schemes that have proven to be both accurate and efficient for a variety of quasi-static and dynamic scenarios [14–19]. In the cohesive finite element scheme, the stiffness of the cohesive elements progressively decays following a traction-separation law, simulating the gradual decohesion of the crack faces. The numerical solutions strongly depend on the choice of the two parameters that define the traction-separation law, i.e., the peak separation stress and the local fracture toughness. Analytical studies of cohesive models [20] have shown that numerical results are less sensitive to the shape of the traction-separation law, provided the peak stress and the area under the traction-separation curve (or fracture toughness) are the same. A common way to extract these fracture parameters relies on a calibration-prediction protocol [17,21,22], i.e., by fitting numerical predictions to experimental results. Nevertheless, this process does not provide a direct and visible link to the physics of the failure process. In order to establish this coupling between theory and modeling, Pandya and Williams [23,24] proposed a scheme to determine the traction-separation law experimentally, which they then used in a cohesive finite difference scheme to simulate crack propagation in polyethylene [25]. Using a rate-independent cohesive law, they were able to achieve good agreement between numerical predictions and experimental results for brittle polyethylene. However, ductile polyethylene showed rate dependency and hence necessitated an improvement in the model.

A cohesive modeling approach was also used by Jin et al. [18,26,27] for quasi-static fracture in Ti/TiB graded materials, by

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received December 20, 2004; final manuscript received November 7, 2005. Review conducted by K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

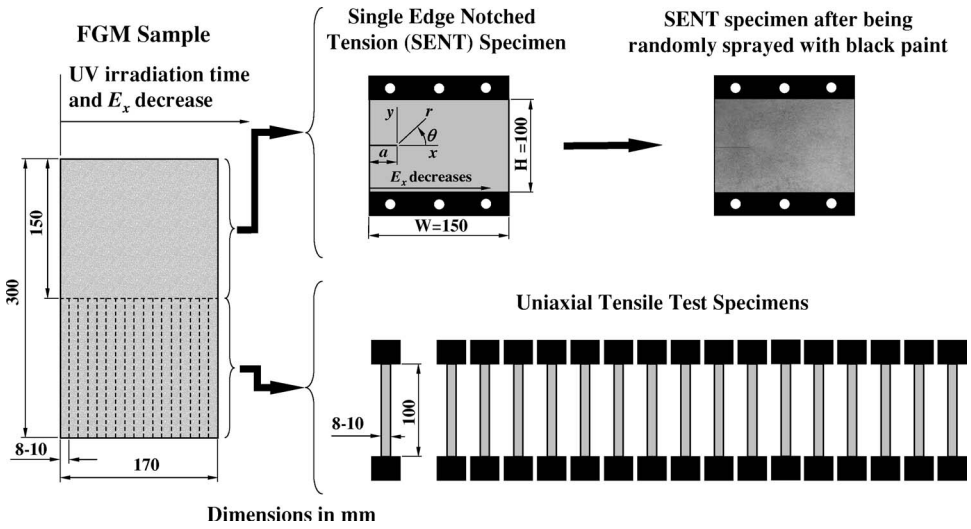


Fig. 1 Details of the ECO SENT FGM specimens used in the experimental study

Tvergaard [19] in the analysis plasticity effects in FGM failure behavior, and by Kandula et al. [21] and Wang and Nakamura [28] to study dynamic fracture in FGMs. In all these publications, the variation of the cohesive law was either assumed or obtained using a calibration-prediction technique. The direct extraction of the failure law from experimental observations is an open area that needs to be explored. This work addresses this issue by fully predicting crack propagation in ECO-based FGMs employing a cohesive methodology in conjunction with detailed experimentation.

The paper is organized as follows: an overview of the quasi-static fracture experiments on ECO model FGM specimens is presented in Sec. 2, followed, in Sec. 3, by a discussion of the numerical scheme. The domain and material properties used in the simulations are presented in Sec. 4. Finally, in Sec. 5, we summarize the results of the quasi-static fracture simulations and compare them to the experimental observations.

2 Material Fabrication and Fracture Experiments

Polymeric model functionally graded materials are manufactured based on selective ultraviolet irradiation of polyethylene co-carbon monoxide (ECO) as developed in [11]. ECO becomes stiffer and more brittle when exposed to increasing UV irradiation. By gradually irradiating a sheet of the material from one end to the other, a sample with continuous in-plane property gradation from stiff and brittle to more compliant and more ductile can be obtained. The ECO samples used are in the form of thin sheets (406 μm thick) so that irradiation is uniform through the thickness. The left-hand side of Fig. 1 shows the dimensions of a typical ECO sheet sample along with the in-plane direction of irradiation.

After irradiation, the sheet is divided into several specimens that are used in a series of uniaxial tension and fracture experiments that provide the bulk and cohesive properties of the material. Each 300 mm $\times$ 170 mm ECO sample is cut into two halves, as shown on the right side of Fig. 1. The lower half is used to generate 8 or 10 mm wide and 100 mm long strips, which are tested to failure in uniaxial tension. These tests provide the Young's modulus E, failure stress σ_f, and failure strain ϵ_f as a function of length along the ECO sheet. Figure 2 shows the spatial variation of these three quantities for two particular FGMs that will be simulated in this work—termed subsequently FGM I and FGM II. Each value has been normalized with its maximum, listed in Table 1 for each case. Piecewise exponential or linear functions have been fitted to the experimental data as shown in Fig. 2. More

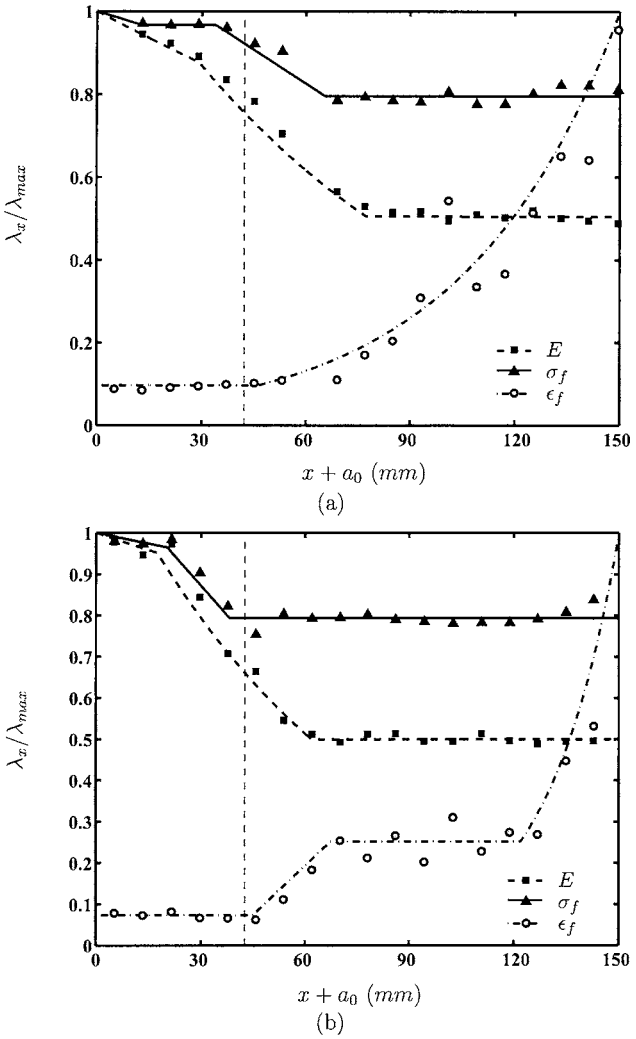


Fig. 2 Spatial variation of normalized material properties in the material gradient direction of ECO SENT FGM I (*a*) and FGM II (*b*) specimens used in the validation study. The maximum values of the material properties, λ_{max}, are given in Table 1. The dashed vertical lines mark the initial crack-tip location. The symbols represent the experimental data, and the curves correspond to their fit using a piecewise exponential or linear functions defined in the Appendix. The width of the specimen is 150 mm, and a_0 is the initial crack length.

Table 1 λ_{max} values for ECO FGM I and FGM II

	FGM I	FGM II
Young's modulus, E (MPa)	413	414
Failure stress, σ_f (MPa)	10.64	10.32
Failure strain, ϵ_f (%)	91	98

details of the experimental methodology, as well as more experimental results from FGMs with different property variations than those discussed here, can be found in [13].

The upper half of the ECO sample has an edge crack of 30 mm introduced parallel to the gradient direction and is used in a single-edge notch tension (SENT) experiment. An MTS RT/30 tabletop universal testing machine is used to load the sample at a constant displacement rate of 0.5 mm/min, and the resulting load is continually recorded. In parallel with the far-field measurement of applied load, the near-tip field is monitored during the entire fracture event using the optical technique of digital image correlation (DIC). This technique extracts the motion of a random speckle pattern painted on the specimen surface by correlating digital images taken before and during deformation (i.e., in the undeformed and deformed configurations). Since this is a well-established technique, it will not be described in detail here and the reader is referred to Sutton et al. [29] and Bruck et al. [30] for more information. The upper right-hand side of Fig. 1 shows a fracture specimen with a speckle pattern painted on it. Digital images of the growing crack are taken in real time at a rate of 0.5 frames per second (fps) and stored on a computer. Figure 3(*a*) shows an image of the undeformed configuration, and Fig. 3(*b*) shows a typical image during crack growth. The correlation is performed for a series of pictures at different crack lengths, yielding full-field in-plane displacements u and v (along the x and y directions respectively) around the crack tip, as shown in the form of a net displacement vector plot in Fig. 3(*c*) for a particular crack length. The displacement field does not appear to have the usual crack-opening form because a rigid motion is also superposed, which is subsequently removed during the data analysis procedure.

Extraction of the stress intensity factor K_I at each crack length is made by fitting, using a least-squares process, the asymptotic expression for displacement near the tip of a crack in an elastic FGM [31]. To second order, this expression for the opening displacement becomes

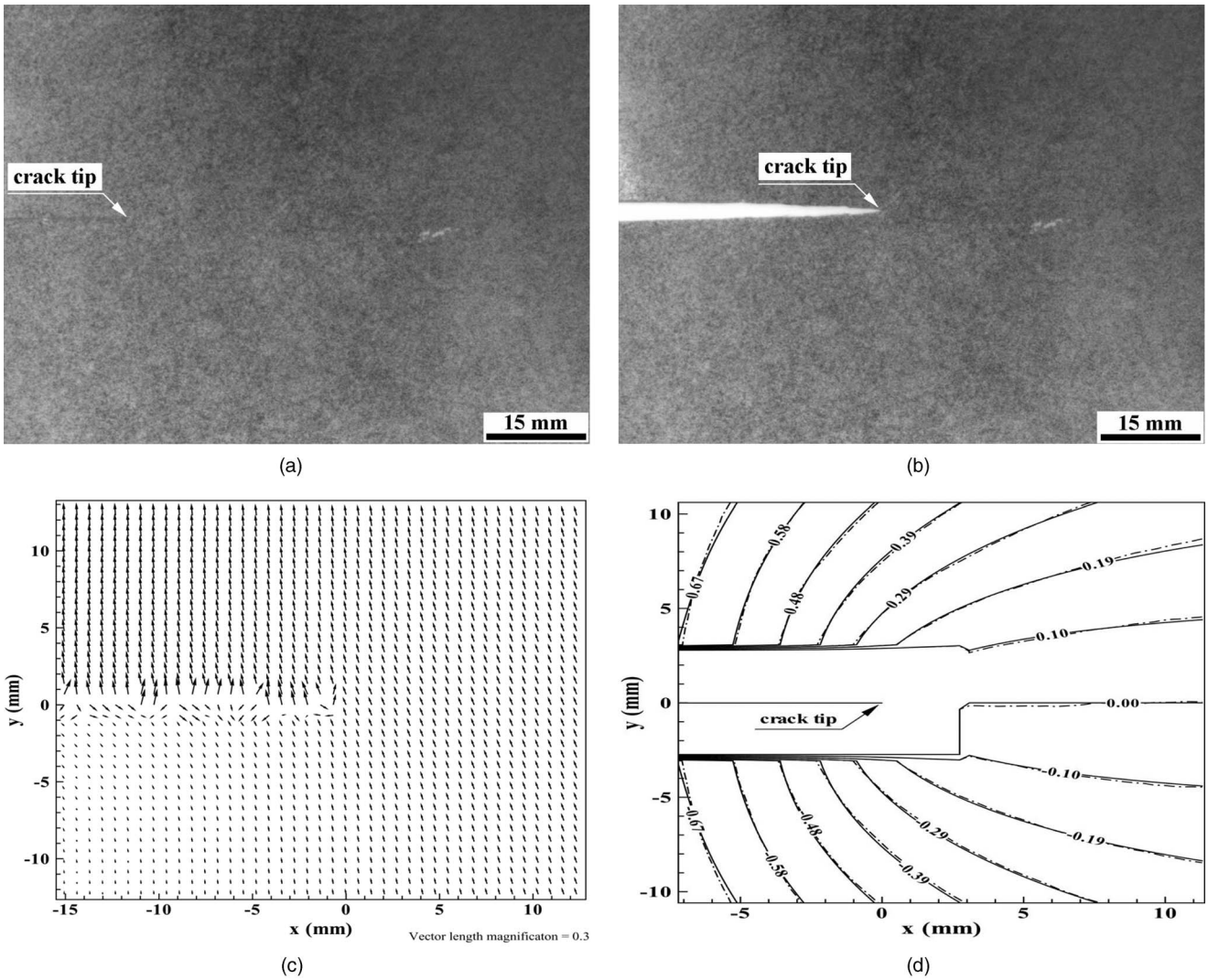


Fig. 3 DIC technique used in the deformation measurements of a SENT specimen: (*a*) undeformed state, (*b*) deformed state, (*c*) vector plot of the measured in-plane displacement field for the instant shown in (*b*), and (*d*) contours of opening displacement comparing measured (dotted-dashed line) and theoretical (solid line) values. In (*c*) and (*d*), the *x*-axis is parallel to the crack line and the crack tip is located at the coordinate origin.

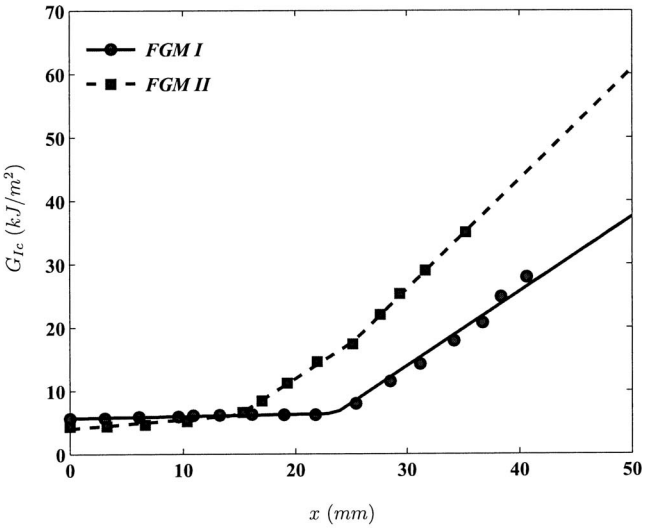


Fig. 4 Fracture resistance (G_{Ic}) curves for ECO FGM I and FGM II along with the experimental data (symbols) versus location *x*. The origin of the coordinate system is defined at the initial crack-tip location.

$$v = \underbrace{\frac{K_I}{2G_{\text{tip}}}\left(\frac{r}{2\pi}\right)^{1/2} \sin\frac{\theta}{2}\left(\frac{3-\nu_{\text{tip}}}{1+\nu_{\text{tip}}} - \cos\theta\right) - \frac{2C_{12}\nu_{\text{tip}}}{G_{\text{tip}}(1+\nu_{\text{tip}})} r\sin\theta}_{\text{Mode I loading}} + \underbrace{A_1 r\cos\theta + v_0}_{\text{Rigid body motion}} + \text{H.O.T.} \tag{1}$$

where K_I is the mode I stress intensity factor, G_{tip} and ν_{tip} are the shear modulus and Poisson's ratio at the crack tip, respectively, C_{12} is a constant proportional to the nonsingular T-stress parallel to the crack plane ($T=4C_{12}$), A_1 defines rigid-body rotation, and v_0 represent rigid-body translation along the y direction. It was shown by Abanto-Bueno and Lambros [13] that the inclusion of the T-stress is critical for the correct measurement of K_I in this configuration. Figure 3(*d*) shows a comparison between the measured v displacement using DIC and the predicted values from the asymptotic field above using the fitted values of K_I and T. Once the mode I stress intensity factor has been obtained, the energy release rate G_{Ic} is then calculated through $G_{Ic}=K_I^2/E_{\text{tip}}$, where E_{tip} is the elastic modulus at the current crack tip location. The process yields an experimental determination of the energy release rate, G_{Ic}, as a function of crack advance. Figure 4 shows this result for FGM I and II, whose material property variation was presented in Figs. 2(*a*) and 2(*b*), respectively. The experimental data terminate at about 40 mm of crack growth since at that point the crack exited the available field of view. Both curves exhibit a "built-in" crack-growth resistance [32], which is one of the primary characteristics of graded materials. In this case, the increasing crack-growth resistance is a consequence of the increasing failure strain with position along the sample, since failure stress is almost constant there (see Fig. 2). More details on the effects of other property gradients are discussed in [13].

The analogy between these model FGM materials and metal-ceramic FGMs can be quantified in terms of stiffness and toughness ratios. For ECO-based FGMs, the stiffness ratio is comparable to that of a metal-ceramic system while the toughness ratio between ductile and brittle phases approaches a value of 10 that is about two orders of magnitude smaller than the toughness ratio observed in metal-ceramic systems. We would therefore expect this built-in crack-growth resistance to be further amplified in the case of metal-ceramic FGMs.

3 Numerical Method

The theory behind cohesive modeling dates back to the early 1960s with the pioneering work of Dugdale [33] and Barenblatt [34] who modeled failure in the process zone ahead of a crack front as a gradual decohesion of two initially congruent surfaces held together by cohesive forces. With the assumption that material softening within the process zone is localized in a thin strip of negligible volume, the failing material continuum there can be substituted with cohesive tractions that resist crack advance. The cohesive tractions act like uniformly distributed nonlinear springs linking the separating failure surfaces, thereby providing some structure to the process zone. The cohesive material continuum can thus be modeled by associating with these cohesive springs a stiffness density, whose magnitude is in accordance with a prescribed cohesive traction-separation law. We can thus divide the continuum into a volumetric part, representing the bulk deformation behavior of the material, and a cohesive part to capture the failure process. Conventional finite element schemes coupled with a cohesive formulation can be used to capture failure in this fashion.

Cohesive elements are introduced between surfaces of the conventional finite element mesh, along lines or in regions where failure is expected. These cohesive elements connect two initially congruent volumetric edges as shown in Fig. 5(*a*) and are characterized by stiffness densities (k_n,k_t). The stiffness densities are damage dependent with amplitudes following the slope of the loading-unloading paths (represented schematically by a dashed line in Fig. 5(*b*)) in the traction-separation law, which relates the cohesive traction vector $\mathbf{T}=\{T_n,T_t\}$ and the displacement jump vector $\Delta=\{\Delta_n,\Delta_t\}$, where the subscripts n and t, respectively denote the normal and tangential components. The progressive decay of the stiffness density leads to complete decohesion, thereby capturing the spontaneous crack initiation and propagation as an eventual outcome of the failure process.

Following the experiments described earlier, we limit our investigation to the quasi-static propagation of mode I cracks in the direction parallel to the material gradient. Accordingly, cohesive elements are placed along the plane of the preexisting crack and the traction-separation law is confined to relate the normal tractions T_n and normal displacement jumps Δ_n. We adopt a simple bilinear relation similar to that used by Geubelle and Baylor [16],

$$T_n = \left(\frac{\mathcal{S}}{1-\mathcal{S}}\frac{\sigma_{\max}}{\mathcal{S}_{\text{initial}}}\frac{1}{\Delta_{nc}}\right)\Delta_n = k_n\Delta_n \tag{2}$$

where $\sigma_{\max}$ and Δ_{nc} respectively denote the local values of the tensile failure strength and critical crack opening displacement of the material. The damage parameter $\mathcal{S}$ is defined by

$$\mathcal{S} = \left\langle 1 - \frac{\Delta_n}{\Delta_{nc}} \right\rangle \tag{3}$$

where $\langle z\rangle = z$ if $z>0$ and $=0$ otherwise. To prevent healing of the cohesive zone in the event of unloading, $\mathcal{S}$ is constrained to be monotonically decreasing from its initial value $\mathcal{S}_{\text{initial}}$, typically chosen close to unity to its final value of zero, at which point complete failure is assumed. The area under the cohesive failure curve corresponds to the mode I fracture toughness G_{Ic}, which, for the present bilinear model, is given by

$$G_{Ic} = \tfrac{1}{2}\Delta_{nc}\sigma_{\max} \tag{4}$$

The problem is analyzed within the framework of the small-strain theory of elasticity with an isotropic linear elastic constitutive law to describe the bulk deformation response of the volumetric elements. In the present study, we employ conventional plane stress three-noded constant strain triangles (CST) as volumetric elements along with four-noded cohesive elements.

The implementation of the cohesive volumetric finite element (CVFE) scheme employs a procedure similar to that outlined by

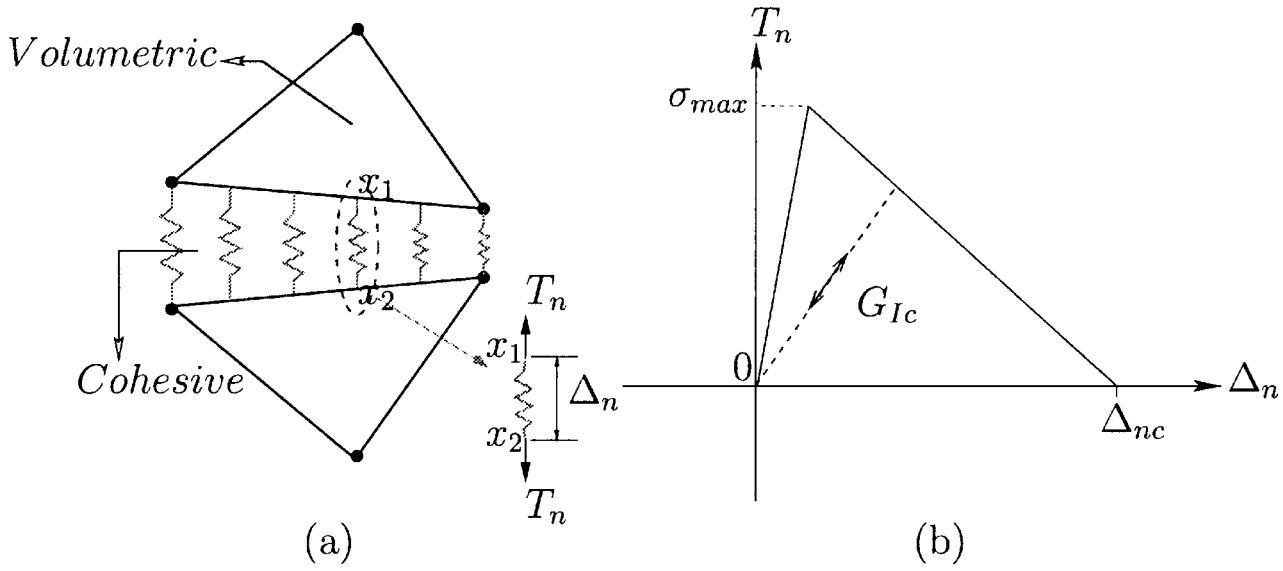


Fig. 5 (*a*) **Schematic of the CVFE concept, showing a cohesive element introduced between two volumetric elements. The cohesive element is shown its deformed state: originally it has no thickness and the adjacent nodes are superposed. It behaves like distributed nonlinear cohesive springs connecting the adjacent volumetric edges.** (*b*) **Schematic of the bilinear cohesive model for mode I. The dashed line shows potential unloading and reloading paths.**

Maiti and Geubelle [35] who modeled fatigue crack propagation in polymeric materials. The governing finite element equations are derived by using the principle of virtual work, which, for this system, is given by

$$\int_{\Omega} \boldsymbol{\sigma}:\delta\boldsymbol{\epsilon}d\Omega - \int_{\Gamma_{ex}} \mathbf{T}_{ex}\,.\,\delta\mathbf{u}d\Gamma_{ex} - \int_{\Gamma_c} T_n\delta\Delta_n d\Gamma_c = 0 \qquad (5)$$

where $\delta\mathbf{u}$ is the virtual displacement vector, $\boldsymbol{\sigma}$ and $\delta\boldsymbol{\epsilon}$, respectively, represent the internal stress and virtual strain tensors, and $\mathbf{T}_{ex}$ denotes the externally applied tractions. The symbols Ω, Γ_c, and Γ_{ex}, respectively, define the volume, cohesive, and external boundaries over which the integrals are evaluated. The principle of virtual work expressed by Eq. (5) is conventional except for the presence of the third integral, which symbolizes the virtual work contribution of the cohesive tractions T_n and the virtual crack opening of $\delta\Delta_n$.

The resulting nonlinear equilibrium equations are solved quasi-implicitly following the procedure outlined by Maiti and Geubelle [35]. To ensure convergence of the numerical results, the CVFE scheme uses an adaptive load-stepping procedure similar to the approach of Roy and Dodds [17] along with the mesh selection criterion suggested by Geubelle and Baylor [16]. The spatial variation of the constitutive and failure response of the FGM is modeled by assigning at each integration point the spatially dependent material properties obtained from the experiments presented in Sec. 2 (Figs. 2 and 4).

4 Domain and Material Description

The two FGM SENT specimens, with the geometry described in the upper right part of Fig. 1, are discretized using three-noded constant strain triangles as shown in Fig. 6. A total of 256 cohesive elements averaging 0.5 mm in length are placed in the path of the mode I crack to capture accurately the failure process in the vicinity of the advancing crack tip. This discretization has been adopted after performing a spatial convergence analysis, with at least 10 elements failing in the cohesive zone at all loading steps during the quasi-static fracture event.

The main inputs of the numerical scheme are bulk properties for the volumetric elements and cohesive properties, i.e., failure strength and fracture toughness, for the cohesive elements. Clearly, both bulk and failure properties are spatially dependent. Since the extraction of the fracture parameters was based on a linearly elastic assumption, the bulk response is assumed to be linearly elastic with the modulus gradation measured in the experiments. The Poisson's ratio, assumed constant, is taken from the literature as 0.45 [36]. To allow for the input of material properties at any location of the specimen, a curve fit to the experimental results has been performed using segments of straight lines or exponentials (Fig. 2).

As described earlier, the spatial dependence of the material toughness has been directly extracted from DIC measurements of the near-tip fields. A continuous curve has been fitted to the experimental DIC data (Fig. 4), and the curve fit is then used to provide the toughness variation G_{Ic} as a function of crack advance along the specimen length. The variation of the cohesive element strength $\sigma_{\max}$ is taken to be that of the uniaxial strength as measured by the 15 tensile tests in each sample, along the lines suggested by Pandya and Williams [24] for polyethylene. Because of the lack of accurate DIC data beyond the range $0 \leqslant x \leqslant 40$ mm as shown in Fig. 4, the toughness values are extrapolated linearly for

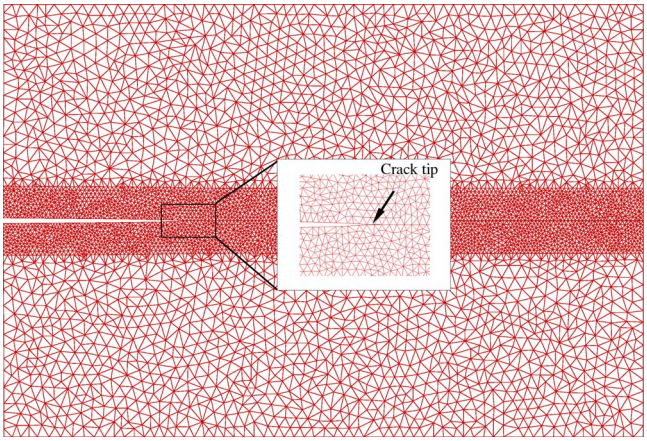


Fig. 6 Details of the finite element mesh used in the validation study with the inset showing a close view of the initial crack tip

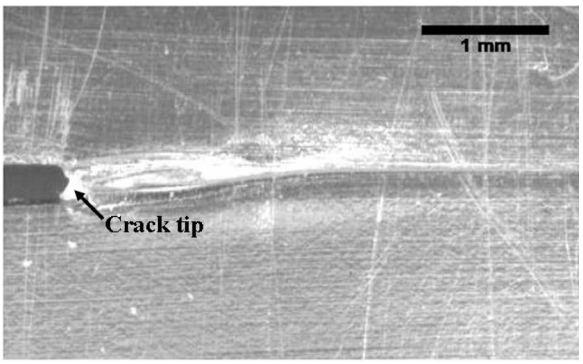


Fig. 7 Snapshot of a 300 h homogeneously UV-irradiated ECO fracture specimen showing the extent of the crazing ahead of the crack tip

the remaining length of the specimens and are used in the simulations. A complete description of the material properties is provided in the Appendix.

Incremental loading is applied along the top edge of the SENT specimen up to the maximum applied displacement U_{app}. Along the bottom edge, we apply a fixed boundary constraint while the vertical edges and the crack faces are left traction free. To eliminate the initial crack-tip blunting effect in these specimens, we simulate and compare the crack propagation only in the region where steady-state conditions are achieved. This is done by shifting the initial crack tip (a_0) further along the gradient direction to the location where steady-state growth is obtained experimentally, shown by the vertical dashed line in Fig. 2. The value of a_0 after the shift is equal to 42.19 and 42.57 mm for FGM I and FGM II, respectively (compared to an experimental a_0 value of 30 mm). The onset of steady-state conditions in each case was determined experimentally from the position where the crack-growth resistance curve reaches a plateau before entering the graded region.

5 Results and Discussion

As mentioned earlier, the spatial variation of the cohesive traction-separation law dictates the fracture behavior of the FGM system under study. The shape of the bilinear cohesive law is determined by the material parameters, σ_{max} and G_{Ic}, whose spatial variation is described in Sec. 2, and the initial value $S_{initial}$ of the damage parameter S, which controls the cohesive law's initial rising part. To simulate crack propagation in systems generally characterized by small and narrow cohesive zones, the value of $S_{initial}$ is typically chosen close to unity ($\sim$0.99) so that the initial value of the cohesive stiffness k_n is very high. Consequently, in the absence of damage, all deformations are concentrated in the surrounding "bulk" material. In the polymeric system of interest, the cohesive zone can often be quite long. Figure 7 shows a craze 3 mm long and 0.5 mm wide extending ahead of a crack in ECO homogeneously irradiated for 300 h. The failure process taking place ahead of the crack tip is quite complex and involves the creation of a large craze. The material drawn into the craze undergoes extensive stretching and chain reorientation. In the cohesive zone model, this damage region is collapsed over a *zero-thickness* line. In order to model the compliance of the material in this damage region, a lower value of $S_{initial}$ needs to be used. To determine a suitable value of $S_{initial}$, we first perform a parametric study on FGM I and closely match the numerical results to experimental observations. The extracted value of $S_{initial}$ is then used to fully predict the mode I crack response in FGM II.

The results of the parametric study performed on FGM I SENT are summarized in Fig. 8 for values of $S_{initial}$: 0.98, 0.8, 0.7, and 0.6. All others properties are kept equal to those described in Sec. 2. Figure 8(*a*) shows the global crack advance Δa and the corresponding reactions $\mathcal{R}$ as function of displacement (U_{app}) applied along the top edge of the SENT specimen, for the four values of $S_{initial}$ mentioned above, compared to their experimental counterparts. As apparent there, the reaction increases linearly with the applied displacement prior to crack initiation. Because of the aforementioned difference in initial crack length values, the numerical stiffness of the specimens are slightly lower than the experimental counterpart. After initiation, the crack propagates into the tougher FGM region and slows down because of an increase in fracture resistance. However, during the initial stages of crack propagation, the slight increase in the fracture toughness (Fig. 4) is not sufficient to slow down the crack advance, thus resulting in the small dip in the global reaction curve apparent in Fig. 8(*a*). As the crack propagates further into the tougher graded region, we again see an increase in the reaction curve, clearly showing the effect of toughening characterizing these FGMs. Prior to the attainment of peak load, the analysis overpredicts the reaction but underestimates the crack advance. As the crack propagates further into the tougher region, the ductile nature of the ECO specimen tends to dominate and our assumption of a linearly elastic material becomes less valid. As mentioned earlier precise crack-tip location as well as the fracture toughness could not be measured be-

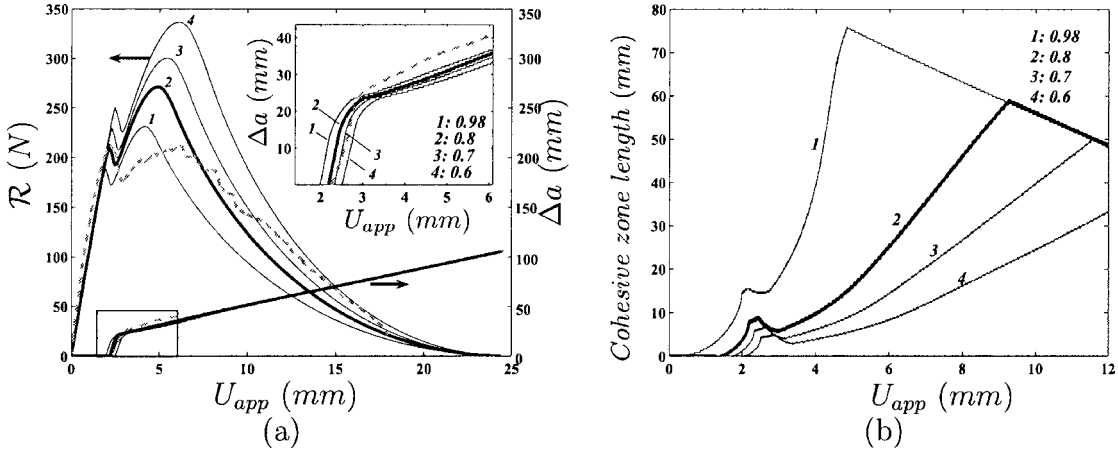


Fig. 8 (*a*) Evolution of the reaction $\mathcal{R}$ and crack length Δa curves of FGM I SENT specimen as a function of the applied displacement U_{app} for four different values of $S_{initial}$: 0.98, 0.8, 0.7, and 0.6. The close-up of the crack propagation curves is plotted in the inset for clarity. The dashed curves correspond to the experimental measurements. (*b*) Evolution of the cohesive zone length with U_{app} for the same four values of $S_{initial}$.

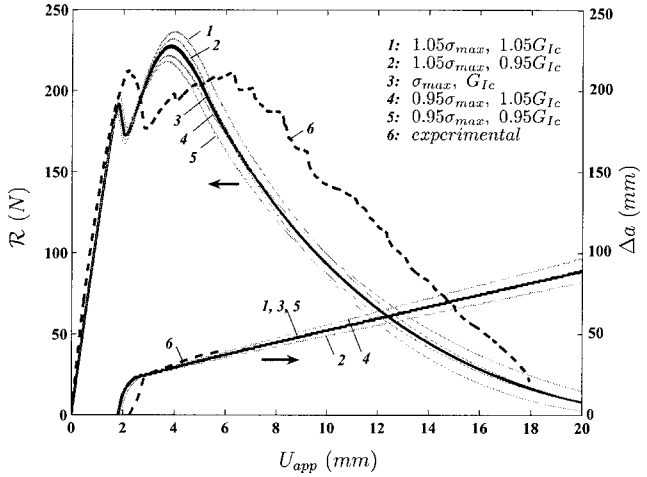


Fig. 9 Comparison between the numerically predicted crack advance (Δa) and corresponding reaction force ($\mathcal{R}$) curves with the experimentally observed counterparts for five different FGMs used in the sensitivity analysis, as a function of applied displacement (U_{app}). A $S_{initial}$ value of 0.98 is used in this study.

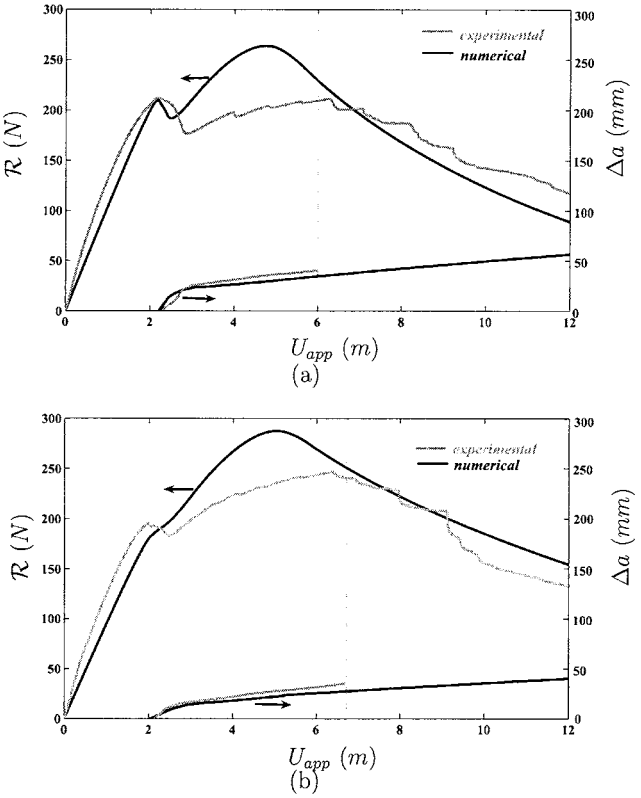


Fig. 10 Comparison between the numerically predicted crack advance (Δa) and corresponding reaction force ($\mathcal{R}$) curves with the experimentally observed counterparts as a function of the applied displacement (U_{app}), assuming a peak separation stress of 0.95 $\sigma_{max}(x)$: (*a*) ECO FGM I SENT specimen and (*b*) ECO FGM II SENT specimen.

yond a certain Δa, when the crack exited the field of view of the camera used. This accounts for the truncation of the experimental crack advance curves.

Decreasing the value of $S_{initial}$ results in an increase of global reaction force but has a very little influence on crack initiation. Interestingly, the crack propagation response is also not effected by the choice of $S_{initial}$ since it is primarily dependent on the value of Δ_{nc} whose variation remains unchanged in these simulations. Nevertheless, the overall shape of the reaction curve prior to crack initiation is preserved. A detailed probe into the failure process revealed that the cohesive zone region, where the failure is assumed concentrated, is highly sensitive to the choice of $S_{initial}$. The cohesive zone length as a function of U_{app} is shown in Fig. 8(*b*) for the same values of $S_{initial}$ considered earlier. Despite the excellent agreement of reaction force and crack length in Fig. 8(*a*) using $S_{initial}$ of 0.98, the result did yield unrealistically large cohesive zone lengths up to 70 mm in the region of interest ($U_{app} \leq 6$ mm). Therefore, a $S_{initial}$ value of 0.98 is deemed unfit for these simulations and a smaller value needs to be adopted.

It is important to note that the entire analysis is carried out assuming that all the plasticity ahead of the crack front is concentrated in the cohesive zone and the surrounding material remains elastic. Also, the critical failure stress (σ_{max}) is determined from the macroscopic tensile test data, which might be accurate for brittle systems, but tends to deviate from the actual value for ductile systems. Both of these assumptions lead to unrealistically high cohesive zone lengths at the later stages of the fracture event. Keeping this in mind and looking at how well the global reaction is compared to the experimental one, we adopt a $S_{initial}$ value of 0.8, as this selection gives a satisfactory comparison with the $\mathcal{R}$ and Δa versus U_{app} curves while yielding physically more realistic values of the cohesive zone lengths.

To account for the experimental scatter associated primarily with the measured values of the failure parameters (σ_{max} and G_{Ic}), which may provide similar effects to the variation of $S_{initial}$ (Fig. 8), we also performed a sensitivity analysis with a ±5% change in the magnitudes of these two parameters. Through these simulations, we also quantified the sensitivity of the numerical results to the choice of failure parameters and estimated the bounds of the numerical predictions. The results of this study are presented in Fig. 9, which represent the evolution of crack advance and corresponding reaction in FGM I as function of applied displacement for five different combinations of the failure parameters: $\{\sigma_{max}, G_{Ic}\}$, $\{1.05\sigma_{max}, 1.05G_{Ic}\}$, $\{0.95\sigma_{max}, 1.05G_{Ic}\}$, $\{0.95\sigma_{max}, 0.95G_{Ic}\}$, and $\{1.05\sigma_{max}, 0.95G_{Ic}\}$. The FGM with the fracture properties $\{1.05\sigma_{max}, 1.05G_{Ic}\}$ represents the strongest and toughest case among all the other FGMs. Note that, for the above combination of the failure parameters, the variation of the critical opening displacements (Δ_{nc}) is exactly the same as that of the FGM I, leading to a similar crack advance curve (see curves labeled 1, 3, and 5). For the FGM combination with $\{0.95\sigma_{max}, 0.95G_{Ic}\}$, a similar argument can be extended for the crack advance and predict low reactions as it has the smallest cohesive strength and toughness. These two combinations thus produce the upper and lower bounds of the numerical load predictions for a particular value of $S_{initial}$. In these simulations, the value of $S_{initial}$ is chosen to be 0.98 and we expect the results to follow a similar trend for any other value of $S_{initial}$. The crack advance curves obtained for the five FGM cases clearly indicate that the crack propagation behavior is highly dominated by Δ_{nc} and that, within the range of values considered, the cohesive parameters influence the crack initiation only weakly. Clearly, the combinations that produce the smallest and largest values of Δ_{nc} [Eq. (4)] bound the crack advance curves.

Having extracted the value of $S_{initial}$, we now use it to *predict* the fracture behavior in ECO FGM II SENT specimen. The spatial variations of $\sigma_{max}(x)$ and $G_{Ic}(x)$ are respectively, chosen as 95% and 100% of those presented in Table 2. Figures 10(*a*) and 10(*b*) show the evolution of the numerical reaction and crack advance curves as a function of applied displacement plotted along with their experimental counterparts for FGM I and FGM II, respectively. The numerical predictions are in good agreement with the experimental observations with respect to the crack initiation point and its subsequent propagation. However, as mentioned be-

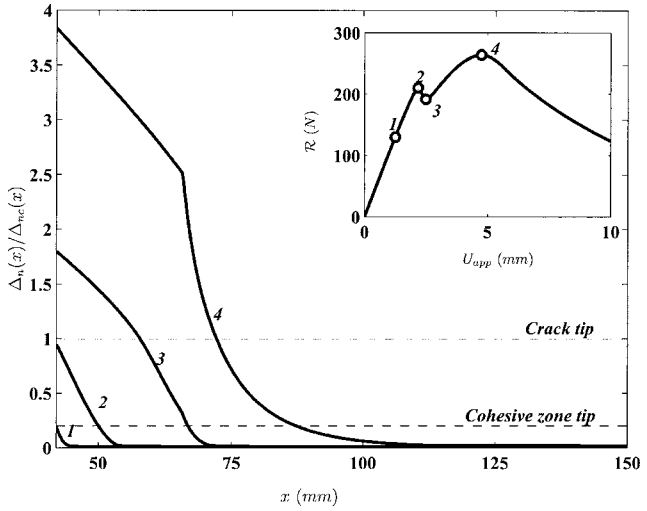


Fig. 11 COD distribution $\Delta_n(x)$ along the fracture plane normalized with the local critical separation $\Delta_{nc}(x)$ values computed at the four points of interest, denoted by symbols in the inset, on the reaction versus applied displacement curve for ECO FGM I and with S_{initial} of 0.8. The intersection of the curves with the top and bottom horizontal dashed lines mark the crack-tip and the cohesive zone tip locations, respectively.

fore, the analysis overpredicts the reaction after crack propagation, implying that the numerical specimens are stiffer than the experimental ones. The agreement between numerical and experimental results for FGM II highlights the predictive capabilities of the numerical simulations. The resulting choice of the cohesive parameters yields cohesive lengths that are realistic. This leads us to conclude that this simplified model that combines linearly elastic constitutive response with a rate independent failure law closely captures the global load response of these FGMs provided the "stiffness" of the craze is properly accounted for through S_{initial}.

Beyond a direct comparison to experimental observations, the CVFE results also provide some insight on the failure process. In Fig. 11, we present snapshots of the spatial distribution of the crack-opening displacement (COD) $\Delta_n(x)$ along the fracture plane normalized by the local value of the critical separation $\Delta_{nc}(x)$, at various observation points marked 1–4 along the reaction curve for FGM I with an S_{initial} of 0.8. The top horizontal dashed line denotes the critical value of the displacement jump beyond which complete failure is assumed, i.e., the crack tip, while the bottom horizontal line marks the onset of damage where the damage parameter S falls below S_{initial}, which we define as the cohesive zone tip. As apparent there, in spite of the stark variation in the local values of $\Delta_{nc}(x)$, we see a gradual decreasing trend in the normalized values as we move away from the crack tip, indicating a smooth and progressive decohesion process.

Table 2 Material property variation of SENT ECO FGM I specimen (x in millimeters)

Young's modulus $E(x)$ (MPa):	
$E(x)=444.09e^{-0.0045(x+18.5)}$	$0<x\leqslant 29.10$
$E(x)=623.71e^{-0.0114(x+18.5)}$	$29.10<x\leqslant 77.57$
$E(x)=208.62$	$77.57<x\leqslant 150.00$
Failure strain $\epsilon_f(x)$ (%):	
$\epsilon_f(x)=8.903$	$0<x\leqslant 46.29$
$\epsilon_f(x)=2.072e^{0.0225(x+18.5)}$	$46.29<x\leqslant 150.00$
Failure stress $\sigma_f(x)$ (MPa):	
$\sigma_f(x)=-0.0277(x+18.5)+11.152$	$0<x\leqslant 12.96$
$\sigma_f(x)=10.281$	$12.96<x\leqslant 34.03$
$\sigma_f(x)=-0.0576(x+18.5)+13.306$	$34.03<x\leqslant 65.59$
$\sigma_f(x)=8.463$	$65.59<x\leqslant 150.00$
Fracture toughness $G_{Ic}(x)$ (kJ/m^2):	
$G_{Ic}(x)=5.6687+0.03(x-42.19)$	$42.19<x\leqslant 65.75$
$G_{Ic}(x)=-21.39+1.1784(x-42.19)$	$65.75<x\leqslant 150.00$

Table 3 Material property variation of SENT ECO FGM II specimen (x in millimeters)

Young's modulus $E(x)$ (MPa):	
$E(x)=435.24e^{-0.0028(x+18)}$	$0<x\leqslant 17.46$
$E(x)=659.02e^{-0.0145(x+18)}$	$17.46<x\leqslant 61.93$
$E(x)=206.794$	$61.93<x\leqslant 150.00$
Failure strain $\epsilon_f(x)$ (%):	
$\epsilon_f(x)=7.262$	$0<x\leqslant 44.89$
$\epsilon_f(x)=0.7874(x+18)-42.254$	$44.89<x\leqslant 67.37$
$\epsilon_f(x)=24.97$	$67.37<x\leqslant 121.94$
$\epsilon_f(x)=0.0259e^{0.0491(x+18)}$	$121.94<x\leqslant 150.00$
Failure stress $\sigma_f(x)$ (MPa):	
$\sigma_f(x)=-0.0176(x+18)+10.638$	$0<x\leqslant 20.29$
$\sigma_f(x)=-0.0987(x+18)+13.743$	$20.29<x\leqslant 38.28$
$\sigma_f(x)=8.188$	$38.28<x\leqslant 150.00$
Fracture toughness $G_{Ic}(x)$ (kJ/m^2):	
$G_{Ic}(x)=3.9992+0.1429(x-42.57)$	$42.57<x\leqslant 57.41$
$G_{Ic}(x)=-10.6071+1.127(x-42.57)$	$57.41<x\leqslant 66.13$
$G_{Ic}(x)=-25.988+1.7362(x-42.57)$	$66.13<x\leqslant 150.00$

Acknowledgment

The authors gratefully acknowledge the National Science Foundation for supporting this research through Grant No. CMS 01-15954.

Appendix

The materials property variations for FGM I and FGM II specimens are presented in Tables 2 and 3, respectively.

References

[1] Kawasaki, A., and Watanabe, R., 1987, "Finite Element Analysis of Thermal Stress of the Metal/Ceramic Multi-Layer Composites With Compositional Gradients," J. Jpn. Inst. Met., **51**, pp. 525–529.
[2] Uemura, S., 2003, "The Activities of FGM on New Application," Mater. Sci. Forum, **423–425**, pp. 1–10.
[3] Erdogan, F., and Wu, B. H., 1997, "The Surface Crack Problem for a Plate With Functionally Graded Properties," ASME J. Appl. Mech., **64**, pp. 449–456.
[4] 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.
[5] Jin, Z. H., and Noda, N., 1994, "Crack-Tip Singular Fields in Nonhomogeneous Materials," ASME J. Appl. Mech., **61**, pp. 738–740.
[6] Delale, F., and Erdogan, F., 1983, "The Crack Problem for a Nonhomogeneous Plane," ASME J. Appl. Mech., **50**, pp. 609–614.
[7] Wang, B. L., Mai, Y.-W., and Noda, N., 2002, "Fracture Mechanics Analysis Model for Functionally Graded Materials With Arbitrarily Distributed Properties," Int. J. Fract., **116**, pp. 161–177.
[8] Butcher, R. J., Rousseau, C. E., and Tippur, H. V., 1999, "A Functionally Graded Particulate Composite: Preparation, Measurements and Failure Analysis," Acta Mater., **47**, pp. 259–268.
[9] Chalivendra, V. B., Shukla, A., Bose, A., and Parameswaran, V., 2003, "Processing and Mechanical Characterization of Lightweight Polyurethane Composites," J. Mater. Sci., **38**, pp. 1631–1643.
[10] Parameswaran, V., and Shukla, A., 1998, "Dynamic Fracture of a Functionally Gradient Material Having Discrete Property Variation," J. Mater. Sci., **33**, pp. 3303–3311.
[11] Lambros, J., Santare, M. H., Li, H., and Sapna III, G., 1999, "A Novel Technique for the Fabrication of Laboratory Scale Functionally Graded Materials," Exp. Mech., **39**, pp. 183–189.
[12] Li, H., Lambros, J., Cheeseman, B. A., and Santare, M. H., 2000, "Experimental Investigation of the Quasi-Static Fracture of Functionally Graded Materials," Int. J. Solids Struct., **37**, pp. 3715–3732.
[13] Abanto-Bueno, J., and Lambros, J., (2005), "Parameters Controlling Fracture

Resistance Curves in Functionally Graded Materials Under Mode I Loading," International Journal of Solids and Structures (in press).
[14] Camacho, G. T., and Ortiz, M., 1996, "Computational Modelling of Impact Damage in Brittle Materials," Int. J. Solids Struct., **33**, pp. 2899–2938.
[15] Xu, X. P., and Needleman, A., 1996, "Numerical Simulations of Dynamic Crack Growth Along an Interface," Int. J. Fract., **74**, pp. 289–324.
[16] Geubelle, P. H., and Baylor, J. S., 1998, "Impact-Induced Delamination of Composites: A 2D Simulation," Composites, Part B, **29B**, pp. 589–602.
[17] Roy, Y. A., and Dodds, 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.
[18] Jin, Z.-H., Paulino, G. H., and Dodds, R. H., 2002, "Finite Element Investigation of Quasi-Static Crack Growth in Functionally Graded Materials Using a Novel Cohesive Zone Fracture Model," ASME J. Appl. Mech., **69**, pp. 370–379.
[19] Tvergaard, V., 2002, "Theoretical Investigation of the Effect of Plasticity on Crack Growth Along a Functionally Graded Region Between Dissimilar Elastic-Plastic Solids," Eng. Fract. Mech., **69**, pp. 1635–1645.
[20] Williams, J. G., and Hadavinia, H., 2002, "Analytical Solutions for Cohesive Zone Models," J. Mech. Phys. Solids, **50**, pp. 809–825.
[21] Kandula, S. S. V., Abanto-Bueno, J., Geubelle, P. H., and Lambros, J., 2005, "Cohesive Modeling of Dynamic Fracture in Functionally Graded Materials," Int. J. Fract., **132**, pp. 275–296.
[22] Bi, X., Li, Z., Geubelle, P. H., and Lambros, J., 2002, "Dynamic Fiber Debonding and Frictional Push-Out in Model Composite Systems: Numerical Simulations," Mech. Mater., **34**, pp. 433–446.
[23] Pandya, K. C., and Williams, J. G., 2000, "Cohesive Modeling of Crack Growth in Polymers: Part 1—Experimental Measurement of the Cohesive Law," Plast. Rubber Compos., **29**, pp. 439–446.
[24] Pandya, K. C., and Williams, J. G., 2000, "Measurement of Cohesive Zone Parameters in Tough Polyethylene," Polym. Eng. Sci., **40**, pp. 1765–1776.
[25] Ivankovic, A., Pandya, K. C., and Williams, J. G., 2004, "Crack Growth Predictions in Polyethylene Using Measured Traction Separation Curves," Eng. Fract. Mech., **71**, pp. 657–668.
[26] Jin, Z.-H., Paulino, G. H., and Dodds, R. H., 2003, "Cohesive Fracture Modeling of Elastic Plastic Crack Growth in Functionally Graded Materials," Eng. Fract. Mech., **70**, pp. 269–283.
[27] Jin, Z. H., and Dodds, R. H., 2004, "Crack Growth Resistance Behavior of a Functionally Graded Material: Computational Studies," Eng. Fract. Mech., **71**, pp. 1651–1672.
[28] Wang, Z., and Nakamura, T., 2004, "Simulations of Crack Propagation in Elastic-Plastic Graded Materials," Mech. Mater., **36**, pp. 601–622.
[29] Sutton, M., Wolters, W., Peters, W., Ranson, W., and McNeill, S., 1988, "Determination of Displacements Using an Improved Digital Image Correlation Method," Image Vis. Comput., **1**, pp. 133–139.
[30] Bruck, H., McNeill, S., Sutton, M., and Peters III, W., 1989, "Digital Image Correlation Using Newton-Raphson Method of Partial-Differential Correction," Exp. Mech., **29**, pp. 261–267.
[31] Eischen, J. W., 1987, "Fracture of Nonhomogeneous Materials," Int. J. Fract., **34**, pp. 3–22.
[32] Erdogan, F., 1995, "Fracture Mechanics of Functionally Graded Materials," Composites Eng., **5**, pp. 753–770.
[33] Dugdale, D. S., 1960, "Yielding of Steel Sheets Containing Slits," J. Mech. Phys. Solids, **8**, pp. 100–104.
[34] Barenblatt, G. I., 1962, "The Mathematical Theory of Equilibrium of Cracks in Brittle Fracture," Adv. Appl. Mech., **7**, pp. 55–129.
[35] Maiti, S., and Geubelle, P. H., 2005, "A Cohesive Model for Fatigue Failure of Polymers," Eng. Fract. Mech., **72**, pp. 691–708.
[36] Domininghaus, H., 1993, *Plastics for Engineers: Materials, Properties and Applications*, Hanser Publishers, Munich.

Eun-Suok Oh
Department of Aerospace Engineering,
Texas A&M University,
College Station, TX 77843-3368

Jay R. Walton
Professor of Mathematics
and Aerospace Engineering
Department of Mathematics,
Texas A&M University,
College Station, TX 77843-3368
e-mail: jwalton@math.tamu.edu

John C. Slattery
Professor of Aerospace and Chemical Engineering
and Mathematics
Texas A&M University,
College Station, TX 77843-3368

A Theory of Fracture Based Upon an Extension of Continuum Mechanics to the Nanoscale

A theory of fracture is presented that is based upon an extension of continuum mechanics to the nanoscale through the incorporation of long-range intermolecular forces which correct bulk material descriptions near interfaces. The surface energy on crack surfaces, which is given in terms of the long-range intermolecular forces, plays an important role in an expression for the stress distribution near the crack tip. It is observed through numerical simulation that the incorporation of these long-range intermolecular forces removes the square-root stress singularity predicted by classical linear elastic fracture mechanics. [DOI: 10.1115/1.2166651]

1 Introduction

Linear elastic fracture mechanics is arguably one of the most successful theories in all of applied continuum mechanics. Its literature is vast and it has been applied to countless practical fracture problems for brittle solids exhibiting small, very limited amounts of inelastic or dissipative fracture processes.[1] However, this record of successful applications has been achieved in spite of the theory's nagging logical inconsistencies and physically impossible predictions. For example, while the theory is set within the context of the linearized equations of elasticity, which is predicated upon the assumption of infinitesimal strains, it predicts strains and stresses that become unbounded in the immediate neighborhood of a crack edge. Even worse, when applied to fractures occurring along an interface between two distinct solid phases, it predicts interpenetration of the two fracture surfaces.

There have been many strategies proposed for rationalizing, resolving, or avoiding these logical inconsistencies. Early on in the development of the theory when these inconsistencies were first encountered, some argued that they could be ignored without harm to the successful application of the theory to practical problems. The rationale for this view was that physically unrealistic magnitudes for stresses and strains predicted by the theory occur only in a region very near to a fracture edge and that, since the theory predicts finite fracture energies, a rational fracture criterion based upon the fracture energy release rate could be implemented in spite of the infinitely large stresses and strains and material interpenetration, at least in the context of quasi-static or equilibrium fracture problems with a small-scale yield zone.

Others argued that the inconsistencies could be avoided by considering elastic fracture mechanics within the context of finite elasticity. However, when the classical boundary conditions used to model a fracture front are imposed upon the equations of finite elasticity, the theory still predicts infinite stresses.

Still others have contended that the inconsistencies are due to the nature of the boundary conditions arising from the classical idealized view of a fracture edge. More specifically, a fracture has been classically modeled as a surface (in the material body's reference configuration, say) across which the displacement vector has a discontinuity. The boundary of this fracture surface is a curve representing an idealized fracture edge or front. The boundary conditions specify the tractions on the fracture surface, resulting in a sharp change in the boundary conditions across the fracture edge. It was then argued by many researchers that one should give a fracture edge or zone a two- (cohesive zone) or three- (process zone) dimensional structure, in order to prevent the development of unbounded stresses and strains. Many such analyses have appeared in the literature in the context of both linear and finite elasticity. Despite its obvious virtues, this approach is not without its own thorny difficulties, both theoretical and practical. For one thing, many implementations of either cohesive or process zone models employ constitutive relations for the cohesive or process zone of dubious physical basis. Moreover, determining constitutive response functions for either a cohesive or process zone model from experimental observation is highly problematical.

Some have argued that, since in real materials fracture occurs via nano- or molecular-scale processes, one should not hope to be able to model it consistently or effectively within the context of continuum mechanics. Consequently, there is also a large and growing literature devoted to modeling and simulating fracture via molecular dynamics or lattice fracture methodologies [2–10]. A major difficulty that one immediately encounters in attempting to attack such problems using a molecular description of the material is the specification of the intermolecular forces [6]. This can be done exceptionally well for dilute gases, but is more problematic for liquids and solids. Alternatively, attempting to apply nonlocal continuum theories of material behavior in these interfacial regions is hampered by the difficulty of measuring local stress-deformation behavior at such small scales.

The continuum approach to fracture modeling is based on bulk descriptions of behavior; for example, bulk stress-deformation behavior. This bulk description of stress deformation behavior assumes that, about each material point, the material extends to "infinity" (perhaps 100 nm) in all directions. However, within the immediate neighborhood of crack surfaces, including the crack tip, the material is subjected to long-range intermolecular forces from the adjoining phases, and the observed material behavior is different from the bulk material behavior.

Here we present an application of an extension of continuum mechanics to the near molecular or nanoscale proposed by Slattery et al. [11]. For an interfacial region, such as the immediate

[1] See the recent treatise by Broberg [1], for example, for an overview of much of these developments.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received December 22, 2004; final manuscript received November 23, 2005. Review conducted by K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

neighborhood of a crack surface, this theory employs known bulk material behavior, such as linear elasticity, corrected for the long-range intermolecular forces from an adjoining phase. The theory is supported by three applications in which the results have been compared with experimental data: supercritical adsorption, contact angles, and surface energy [11,12]. It should be emphasized here that the theoretical predictions in these applications required no adjustable parameters, yet gave excellent agreement with experimental data.

An unusual aspect of this approach is that it proves more convenient to linearize the equations of elasticity about the deformed configuration in which the crack is open. This is in contrast to classical linear elastic fracture mechanics, in which the linearization is relative to a reference configuration with the crack modeled as a virtual, unopened slit. In particular, it is more convenient to discuss the correction to bulk stress deformation in the deformed frame than in the undeformed frame.

Within the context of finite element approximations to continuum models of material behavior in the neighborhood of defects such as fractures and dislocations, a number of methodologies for incorporating atomistic or molecular scale information have been examined in the literature. Perhaps the most extensively studied approach of this type is the quasi-continuum method introduced in 1996 by Tadmor, Philips, and Ortiz [13–15], which is based upon a fundamentally discrete, atomistic view of material behavior [14]. The continuum flavor of the method derives from the use of energy minimization/finite element technology in its implementation. In regions near defects or cracks in which large deformation gradients occur, the finite element discretization can be taken so fine as to incorporate only a single atom per element, whereas in regions corresponding to slowly varying deformation, the discretization can be much coarser, with the strain energy associated with such coarse elements viewed as representing the average energy of a large ensemble of atoms contained within the element, thereby eliminating many atomistic degrees of freedom.

Another recent approach to combining continuum and molecular models of material behavior involves introducing bridging domains between regions modeled either using bulk, continuum descriptions of material behavior or discrete, molecular descriptions [16]. The idea is to average the continuum and molecular Hamiltonians across the bridging domain, creating a continuous Hamiltonian throughout the body. For example, as applied to fracture, one would employ a molecular Hamiltonian in a near-crack-tip region with a bridging region connecting it to a far-field region described through a continuum Hamiltonian.

Both of these approaches utilize averaging of strain energies in order to describe material behavior. Moreover, fitting such models to experimental data requires the selection of physical length scales associated with the subregions over which the required averages are taken.

The approach adopted herein is based upon a continuum theory of material behavior containing a long-range intermolecular force correction in the vicinity of material interfaces that is derived from an atomistic/molecular view of material behavior near interfaces. As such, this theory is not tied to a particular implementation method, such as finite elements. In particular, as demonstrated below, it is quite amenable to singular perturbation techniques from which useful insight can be gleaned.

1.1 An Extension of Continuum Mechanics to the Nanoscale. While *nanoscale* can have several different meanings, here we mean the immediate neighborhood (within 10 nm) of a material interface. Before presenting our theory of fracture, we first summarize briefly the concept for extending continuum mechanics to the nanoscale recently developed by [11] upon which it is based.

Within the context of continuum mechanics, there are six commonly employed descriptions of a material interface, each having its own somewhat different notation.

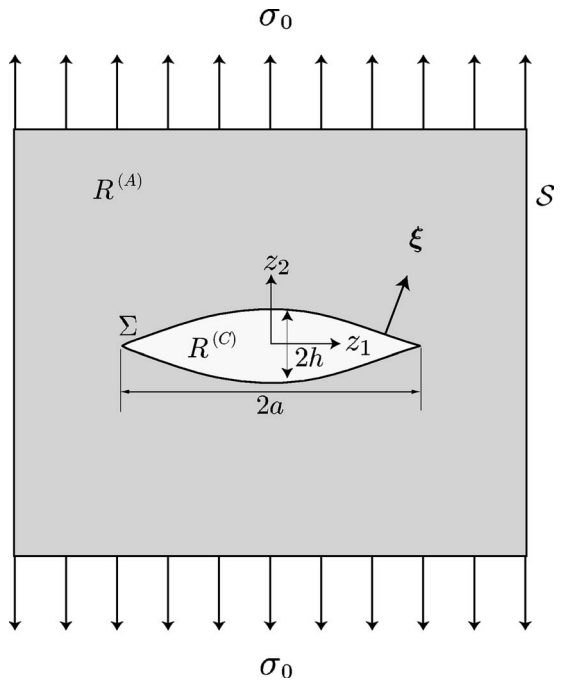


Fig. 1 Schematic of Mode I fracture of phase *C* in phase *A*

(i) The most realistic view of an interface is as a thin, three-dimensional region. There is a smooth transition of the material's density and stress-deformation behavior through this interfacial region from one phase to the other. Because it is so thin, it is extremely difficult to study experimentally the material's behavior in the interfacial region, except as the critical point is approached [17–21]. With this point of view, $\rho^{(I)}$, $\mathbf{v}^{(I)}$, and $\mathbf{T}^{(I)}$ denote the true interfacial density, velocity, and stress tensor, respectively, in the interfacial region.

(ii) In the second view, we recognize that, since we have no way of knowing the true material behavior of the interfacial region, we will use bulk descriptions of material behavior corrected for the effects of long-range intermolecular forces from the adjacent phases. By $\rho^{(I,bulk)}$, $\mathbf{v}^{(I,bulk)}$, and $\mathbf{T}^{(I,bulk)}$, we indicate the density, velocity, and stress tensor respectively, observed with this point of view.

Note that with this point of view, the interface is three-dimensional as in view (i). In contrast with view (i), there is a surface of discontinuity separating the two regions having different bulk behavior. For the sake of simplicity and clarity, we focus on quasi-static systems and ignore effects of gravity. With this point of view, the differential momentum balance reduces to

$$\operatorname{div}\mathbf{T}^{(I,bulk)} + \mathbf{b}^{(corr)} = \operatorname{div}(\mathbf{T}^{(I,bulk)} - \Phi^{(A,corr)}), \qquad =0 \tag{1}$$

where $\mathbf{T}^{(I,bulk)}$ is described by bulk material behavior and $\mathbf{b}^{(corr)}$ is the body force per unit volume introduced to correct for the use of bulk material behavior in the interfacial region. For a body containing a fracture as shown in Fig. 1, we have shown ([11], p. 4629) that

$$\mathbf{b}^{(A,corr)} = \nabla \int_{R^{(C)}} n^{(A)^2} \phi^{(A,A)} d\mathbf{r}, \qquad = -\nabla\Phi^{(A,corr)} \tag{2}$$

in which

$$\Phi^{(A,corr)} \equiv -\int_{R^{(C)}} n^{(A)^2}\phi^{(A,A)}d\mathbf{r}. \quad (3)$$

Here $\phi^{(A,A)}$ is the potential energy between the point at which $\Phi^{(A,corr)}$ is defined and any point in the region $R^{(C)}$, and $n^{(A)}$ is the number density (number of molecules or atoms per unit volume).

(iii) In the third view, the interfacial region is described by a two-dimensional dividing surface. The effects of long-range intermolecular forces are taken into account by introducing *excess quantities*, such as a surface tension or energy γ, in the dividing surface ([22], Secs. 13.5 and 2.1.6). The dividing surface now becomes a two-dimensional representation of the interface.

Again ignoring inertial and gravitational effects, the differential and jump momentum balances become ([22], p. 152, 157)

$$\text{div}\,\mathbf{T} = 0 \quad (4)$$

$$\boldsymbol{\nabla}_{(\sigma)}\gamma + 2H\gamma\boldsymbol{\xi} + [\mathbf{T}\boldsymbol{\xi}] = 0 \quad (5)$$

where $\boldsymbol{\nabla}_{(\sigma)}$ denotes a surface gradient, H is the mean curvature of the dividing surface, and ξ the unit normal to the dividing surface Σ. The boldface brackets denote the jump of the quantity enclosed across the interface between phases A and C:

$$[A\boldsymbol{\xi}] \equiv A^{(A)}\boldsymbol{\xi}^{(A)} + A^{(C)}\boldsymbol{\xi}^{(C)}. \quad (6)$$

By ρ, $\mathbf{v}$, and $\mathbf{T}$ we mean the density, velocity, and stress tensor, respectively, determined with this introduction of a dividing surface and excess properties.

(iv) In the fourth view, one enforces consistency between views (ii) and (iii). Comparing (1) and (4), we conclude that [11]

$$\mathbf{T} = \mathbf{T}^{(I,bulk)} - \Phi^{(corr)}\mathbf{I}. \quad (7)$$

We can also conclude [11] that

$$\gamma = \int_0^{\tau} \Phi^{(A,corr)}d\tau' \quad (8)$$

where τ denotes distance measured along the normal to the dividing surface. It should be noted particularly that $\mathbf{T}^{(I,bulk)}$ rather than $\mathbf{T}$ is represented by bulk material behavior.

(v) The fifth point of view is a variation of view (ii) used only when there are two or more interfaces. In principle, view (iv) could be used for two or more interfaces. In reality, this would be difficult, because of the overlapping intermolecular forces.

We recommend that dividing surfaces be introduced with constant values of surface tension or energy γ^{∞} that corresponds to static, unbounded dividing surfaces. These would be the surface tensions or energies commonly used. The correction potential $\Phi^{(A,corr)}$ must of course be modified to account for these surface tensions or energies.

(vi) The sixth point of view is the one most commonly taken in the literature. It looks similar to view (iii), but no corrections for long-range intermolecular forces are introduced. With this point of view, $\mathbf{T}=\mathbf{T}^{(I,bulk)}$ is represented by bulk material behavior, and excess properties such as γ are determined empirically.

2 Problem Description

Figure 1 shows a static Mode I fracture with an applied stress σ_0. The length of the fracture is specified to be a. Our objective is to determine both the crack configuration and the stress distribution in the solid, particularly within the immediate neighborhood of the fracture tip.

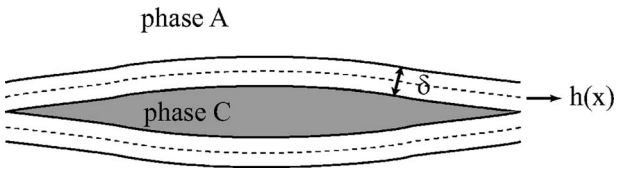


Fig. 2 δ is the distance separating the two phases *A* and *C*, corresponding physically to the sum of the effective radii of the *A* and *C* molecules or the effective distance between molecules of *A* and *C* ([11], p. 4623). The dividing surface $h(x)$ is located halfway between the two phases.

We make several assumptions.

1. The analysis is in the context of linear elasticity and view (iv) above in which each crack face is regarded as a single dividing surface with a surface energy γ that is a function of position.
2. It is assumed that $T^{(I,bulk)}$ can be described by Hooke's law

$$\mathbf{T}^{(I,bulk)} = \lambda(\text{tr}\,\mathbf{e})\mathbf{I} + 2\mu\mathbf{e} \quad (9)$$

in which

$$\mathbf{e} \equiv \frac{1}{2}(\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^{\mathrm{T}}) \quad (10)$$

is the infinitesimal strain tensor considered as a function of position in the deformed configuration, and λ and μ are constants independent of position.

3. We assume that the point-to-point intermolecular force potential can be represented as

$$\phi^{(A,A)} = -\frac{\mathcal{A}}{\pi^2 n^{(A)^2} r^6} \quad (11)$$

where $\mathcal{A}$ is the Hamaker constant [23] computed using *Lifshitz theory* [24]. The Hamaker constants of most condensed phases are found to lie in the range of $(0.4-4)\times10^{-10}$ J [[25], p. 176]. Assuming that the system is two-dimensional, we find from (3) that

$$\Phi^{(A,corr)} = \frac{\mathcal{A}}{\pi^2}\int_{-a}^{a}\int_{-h(x)+\delta/2}^{h(x)-\delta/2}\int_{-\infty}^{\infty} \times\frac{1}{[(z_1-x)^2+(z_2-y)^2+z^2]^3}dzdydx, \quad (12)$$

where δ in Fig. 2 is the distance separating the two phases A and C, corresponding physically to the sum of the effective radii of the A and C molecules or the effective distance between molecules of A and C ([11], p. 4623). The dividing surface $h(x)$ is taken to be located halfway between the two phases.

4. For the case in which gravitational and inertial effects are neglected, our objective is to solve the differential momentum balance (4) consistent with the jump momentum balance (5). In doing so, we utilize (7), (9), (8), and (12). It proves helpful to work in terms of the dimensionless variables

$$\mathbf{u}^{\star} \equiv \frac{\mathbf{u}}{a}, \quad z_i^{\star} \equiv \frac{z_i}{a}, \quad h^{\star} \equiv \frac{h}{a}, \quad \delta^{\star} \equiv \frac{\delta}{a}, \quad \mu^{\star} \equiv \frac{\mu}{E},$$

$$\lambda^{\star} \equiv \frac{\lambda}{E}. \quad (13)$$

Recognizing (7) and (9), the differential momentum balance (4) becomes

$$\mu^\star \operatorname{div}(\nabla \mathbf{u}^\star) + (\lambda^\star + \mu^\star)\, \nabla \operatorname{div} \mathbf{u}^\star - \nabla \Phi^\star = 0 \tag{14}$$

where from (12)

$$\Phi^\star(z_1^\star, z_2^\star) \equiv \frac{\rho\Phi}{E} = \mathcal{A}^\star \int_{-1}^{1} \int_{-h^\star(x)+\delta^\star/2}^{h^\star(x)-\delta^\star/2} \int_{-\infty}^{\infty} \times \frac{1}{[(z_1^\star - x^\star)^2 + (z_2^\star - y^\star)^2 + z^{\star 2}]^3} dz^\star dy^\star dx^\star. \tag{15}$$

Here,

$$\mathcal{A}^\star \equiv \frac{\mathcal{A}}{\pi^2 E a^3} \tag{16}$$

which is normally very small.

The integral expression in (15) depends upon the crack face profile $h(x)$ which is unknown a priori and must be determined as part of the problem's solution. However, two of the three iterated integrals can be evaluated in closed form using elementary techniques. In particular, one readily shows that

$$\Phi^\star(z_1^\star, z_2^\star) = \frac{\mathcal{A}^\star 3\pi}{8} \int_{-1}^{1} \phi^\star[(h(x) - \delta/2), z_1^\star, z_2^\star, x^\star] \frac{dx^\star}{(z_1^\star - x^\star)^4} \tag{17}$$

where

$$\phi^\star(b, z_1, z_2, x) = \frac{b - z_2}{\sqrt{(z_1 - x)^2 + (b - z_2)^2}} + \frac{b + z_2}{\sqrt{(z_1 - x)^2 + (b + z_2)^2}} - \frac{1}{3}\left[\left(\frac{b - z_2}{\sqrt{(z_1 - x)^2 + (b - z_2)^2}}\right)^3 + \left(\frac{b + z_2}{\sqrt{(z_1 - x)^2 + (b + z_2)^2}}\right)^3\right]. \tag{18}$$

To simplify the subsequent analysis, we approximate the integral expression (17) by assuming a priori that $h(x)$ is slowly varying; i.e., that the magnitude of its derivative $|h'(x)|$ is small. This assumption will then be verified a posteriori for consistency. Assuming $h(x)$ to be slowly varying, we then approximate the integral expression (17) by the expression resulting from the assumption that $h(x)$ is locally constant. This integral expression can then be evaluated analytically, after which h is again assumed to be a function of z_1. From our experience with similar approximations in different contexts [11], we expect the effect of this approximation on the final results to be small. When this approximation scheme is applied to (17) and (18), there results the expression for $\Phi^\star$:

$$\begin{aligned}\Phi^\star = \mathcal{A}^\star \Bigg\{ & \frac{\pi}{6(z_2^\star - h^\star + \delta^\star/2)^3}\left[\frac{2(z_1^\star+1)^4 + (z_1^\star+1)^2(z_2^\star - h^\star + \delta^\star/2)^2 + 2(z_2^\star - h^\star + \delta^\star/2)^4}{4(z_1^\star+1)^3\sqrt{(z_1^\star+1)^2 + (z_2^\star - h^\star + \delta^\star/2)^2}}\right. \\ & \left. - \frac{2(z_1^\star-1)^4 + (z_1^\star-1)^2(z_2^\star - h^\star + \delta^\star/2)^2 + 2(z_2^\star - h^\star + \delta^\star/2)^4}{4(z_1^\star-1)^3\sqrt{(z_1^\star-1)^2 + (z_2^\star - h^\star + \delta^\star/2)^2}}\right] \\ & - \frac{\pi}{6(z_2^\star + h^\star - \delta^\star/2)^3}\left[\frac{2(z_1^\star+1)^4 + (z_1^\star+1)^2(z_2^\star + h^\star - \delta^\star/2)^2 + 2(z_2^\star + h^\star - \delta^\star/2)^4}{4(z_1^\star+1)^3\sqrt{(z_1^\star+1)^2 + (z_2^\star + h^\star - \delta^\star/2)^2}}\right. \\ & \left. - \frac{2(z_1^\star-1)^4 + (z_1^\star-1)^2(z_2^\star + h^\star - \delta^\star/2)^2 + 2(z_2^\star + h^\star - \delta^\star/2)^4}{4(z_1^\star-1)^3\sqrt{(z_1^\star-1)^2 + (z_2^\star + h^\star - \delta^\star/2)^2}}\right]\Bigg\}. \end{aligned} \tag{19}$$

Introducing the dimensionless variables

$$H^\star \equiv Ha, \quad \gamma^\star \equiv \frac{\gamma}{Ea} \tag{20}$$

and recognizing (7) and (9), the jump momentum balance (5) (on the crack faces $z_2^\star = h^\star(z_1)$) can be expressed as

$$\text{at } z_2^\star = h^\star: \nabla_{(\sigma)}\gamma^\star + 2H^\star\gamma^\star\boldsymbol{\xi} + \mu^\star(\nabla\mathbf{u}^\star + \nabla\mathbf{u}^{\star\mathrm{T}})\boldsymbol{\xi} + \lambda^\star(\operatorname{div}\mathbf{u}^\star)\boldsymbol{\xi} - \Phi^\star\boldsymbol{\xi} = 0 \tag{21}$$

where (8) now takes the form

$$\gamma^\star = \int_0^{\tau^\star} \Phi^\star d\tau^\star. \tag{22}$$

The integral expression in (22) gives the surface energy at a point on the crack faces $(z_1^\star, h^\star(z_1^\star))$ via integration along a ray emanating from the crack into the solid phase in the direction of the normal to the crack.

3 Singular Perturbation Solution

To gain insight into the behavior of this highly nonlinear problem, we seek a singular perturbation solution using as our perturbation parameter $\mathcal{A}^\star$. When $\mathcal{A}^\star = 0$, our theory demands that there must be no fracture. However, as $\mathcal{A}^\star \to 0$, the solution to the boundary value problem does not approach this limit. For this reason, the perturbation analysis performed below is singular.

3.1 Outer Solution. The outer solution, corresponding to $\mathcal{A}^\star = 0$, is simple uniaxial extension of the uncracked body. Since there is no fracture, the jump momentum balance (5) is not a constraint, and the solution is

$$T_{22}^{(I,bulk)^\star} = \sigma_0^\star \tag{23}$$

$$u_1^\star = -\frac{\lambda^\star(3\lambda^\star + 2\mu^\star)}{4(\lambda^\star + \mu^\star)^2}\sigma_0^\star z_1^\star \tag{24}$$

where

$$T_{22}^{(I,bulk)^\star} \equiv \frac{T_{22}^{(I,bulk)}}{E}, \quad \sigma_0^\star \equiv \frac{\sigma_0}{E}. \tag{25}$$

3.2 Inner Solution. Within the immediate neighborhood of the crack surfaces (inner region), the effects of the intermolecular forces must be preserved. This suggests that we introduce as expanded variables in this region

$$y_2^{\star\star} \equiv \frac{z_2^\star - h^\star}{\mathcal{A}^{\star n}} \tag{26}$$

$$v_2^{\star\star} \equiv \frac{u_2^\star - h^\star}{\mathcal{A}^{\star n}} \tag{27}$$

$$\delta^{\star\star} \equiv \frac{\delta^\star}{2\mathcal{A}^{\star n}} \tag{28}$$

in order not to magnify the crack itself. The exponent n is chosen in (35) so that the effect of the long-range intermolecular forces is retained as $\mathcal{A}^\star \to 0$. Note that, with this value of n,

$$\delta^{\star\star} = \frac{\delta \pi^{2/3} E^{1/3}}{2\mathcal{A}^{1/3}} = \text{constant}. \tag{29}$$

In terms of these scaled variables, the two components of the differential momentum balance (14) become

$$\left\{(\lambda^\star + 2\mu^\star)\frac{\partial^2 u_1^\star}{\partial z_1^{\star 2}} + (\lambda^\star + \mu^\star)\frac{\partial^2 v_2^{\star\star}}{\partial z_1^\star \partial y_2^{\star\star}}\right\} + \frac{1}{\mathcal{A}^{\star 2n}}\left\{(\lambda^\star + 2\mu^\star) \times \left(\frac{dh^\star}{dz_1^\star}\right)^2 + \mu^\star\right\}\frac{\partial^2 u_1^\star}{\partial y_2^{\star\star}} - \frac{1}{\mathcal{A}^{\star n}}\left\{(\lambda^\star + 2\mu^\star)\frac{dh^\star}{dz_1^\star}\frac{\partial^2 u_1^\star}{\partial z_1^\star \partial y_2^{\star\star}} - (\lambda^\star + 2\mu^\star)\frac{d^2 h^\star}{dz_1^{\star 2}}\frac{\partial u_1^\star}{\partial y_2^{\star\star}} - (\lambda^\star + \mu^\star)\frac{dh^\star}{dz_1^\star}\frac{\partial^2 v_2^{\star\star}}{\partial y_2^{\star\star 2}}\right\} = \frac{\partial \Phi^{\star\star}}{\partial z_1^\star} \tag{30}$$

and

$$\mathcal{A}^{\star n}\mu^\star \frac{\partial^2 v_2^{\star\star}}{\partial z_1^{\star 2}} + \left\{\mu^\star \frac{d^2 h^\star}{dz_1^{\star 2}} - \mu^\star \frac{d^2 h^\star}{dz_1^{\star 2}}\frac{\partial v_2^{\star\star}}{\partial y_2^{\star\star}} - 2\mu^\star \frac{dh^\star}{dz_1^\star}\frac{\partial^2 v_2^{\star\star}}{\partial z_1^\star \partial y_2^{\star\star}}\right\} - \frac{1}{\mathcal{A}^{\star 2n}}\left\{(\lambda^\star + \mu^\star)\frac{dh^\star}{dz_1^\star}\frac{\partial^2 u_1^\star}{\partial y_2^{\star\star 2}}\right\}\frac{1}{\mathcal{A}^{\star n}}\left\{(\lambda^\star + 2\mu^\star)\frac{\partial^2 v_2^{\star\star}}{\partial y_2^{\star\star 2}} + \mu^\star \frac{dh^{\star 2}}{dz_1^\star}\frac{\partial^2 v_2^{\star\star}}{\partial y_2^{\star\star 2}} + (\lambda^\star + \mu^\star)\frac{\partial^2 u_1^\star}{\partial z_1^\star \partial y_2^{\star\star}}\right\} = \frac{\partial \Phi^{\star\star}}{\partial y_2^{\star\star}} \tag{31}$$

We have assumed that $h^\star$ is nearly independent of $z_1^\star$, permitting us to neglect both the first and second derivatives. We then have a set of simpler equations

$$\mathcal{A}^{\star -2n}\mu^\star \frac{\partial^2 u_1^\star}{\partial y_2^{\star\star 2}} + (\lambda^\star + 2\mu^\star)\frac{\partial^2 u_1^\star}{\partial z_1^{\star 2}} + (\lambda^\star + \mu^\star)\frac{\partial^2 v_2^{\star\star}}{\partial z_1^\star \partial y_2^{\star\star}} = \frac{\partial \Phi^{\star\star}}{\partial z_1^\star} \tag{32}$$

and

$$\mathcal{A}^{\star 2n}\mu^\star \frac{\partial^2 v_2^{\star\star}}{\partial z_1^{\star 2}} + \mathcal{A}^{\star n}\mu^\star \frac{\partial^2 h^\star}{\partial z_1^{\star 2}} + (\lambda^\star + 2\mu^\star)\frac{\partial^2 v_2^{\star\star}}{\partial y_2^{\star\star 2}} + (\lambda^\star + \mu^\star)\frac{\partial^2 u_1^\star}{\partial z_1^\star \partial y_2^{\star\star}} = \frac{\partial \Phi^{\star\star}}{\partial y_2^{\star\star}}. \tag{33}$$

We have also introduced

$$\Phi^{\star\star}(z_1^\star, y_2^{\star\star}) \equiv \Phi^\star(z_1^\star, z_2^\star) = \mathcal{A}^{\star 1-3n}\left\{\frac{\pi}{6(y_2^{\star\star} + \delta^{\star\star})^3}[\cdots] - \frac{\pi}{6(y_2^{\star\star} + 2h^{\star\star} - \delta^{\star\star})^3}[\cdots]\right\}. \tag{34}$$

If any terms of intermolecular force correction in (32) and (33) are to be the same order of magnitude as the other terms, we must have the identity

$$n \equiv \frac{1}{3}. \tag{35}$$

In view of (35), equations (32) and (33) reduce to

$$\frac{\partial^2 u_1^\star}{\partial y_2^{\star\star 2}} = 0 \tag{36}$$

$$(\lambda^\star + 2\mu^\star)\frac{\partial^2 v_2^{\star\star}}{\partial y_2^{\star\star 2}} + (\lambda^\star + \mu^\star)\frac{\partial^2 u_1^\star}{\partial z_1^\star \partial y_2^{\star\star}} = \frac{\partial \Phi^{\star\star}}{\partial y_2^{\star\star}}. \tag{37}$$

Equations (36) and (37) are satisfied by

$$u_1^\star = f(z_1^\star) y_2^{\star\star} + g(z_1^\star) \tag{38}$$

and

$$(\lambda^\star + 2\mu^\star)v_2^{\star\star} = \int \Phi^{\star\star} dy_2^{\star\star} + m(z_1^\star) y_2^{\star\star} - f'(z_1^\star)\frac{\lambda^\star + \mu^\star}{2} y_2^{\star\star 2} + n(z_1^\star) \tag{39}$$

where a prime indicates a derivative with respect to $z_1^\star$. Functions $f(z_1^\star)$, $g(z_1^\star)$, $m(z_1^\star)$, and $n(z_1^\star)$ can be obtained using the boundary conditions.

On the crack surfaces, the two components of the jump momentum balance (21) are

$$\text{at } y_2^\star = 0\text{: } \frac{\partial \gamma^\star}{\partial z_1^\star}\left\{1 + \left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}^{-1/2} - \frac{\partial^2 h^\star}{\partial z_1^{\star 2}}\left\{1 + \left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}^{-3/2}\gamma^\star\left(\frac{\partial h^\star}{\partial z_1^\star}\right) + \Phi^\star\left(\frac{\partial h^\star}{\partial z_1^\star}\right) + \left[(\lambda^\star + 2\mu^\star)\frac{\partial u_1^\star}{\partial z_1^\star} + \lambda^\star \frac{\partial v_2^\star}{\partial y_2^\star}\right]\left(-\frac{\partial h^\star}{\partial z_1^\star}\right) + \mu^\star\left[\frac{\partial v_2^\star}{\partial z_1^\star} + \frac{\partial h^\star}{\partial z_1^\star} + \frac{\partial u_1^\star}{\partial y_2^\star}\right] = 0 \tag{40}$$

and

$$\text{at } y_2^\star = 0\text{: } \frac{\partial \gamma^\star}{\partial z_1^\star}\left\{1 + \left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}^{-1/2}\left(\frac{\partial h^\star}{\partial z_1^\star}\right) + \frac{\partial^2 h^\star}{\partial z_1^{\star 2}}\left\{1 + \left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}^{-3/2}\gamma^\star - \Phi^\star + \mu^\star\left[\frac{\partial v_2^\star}{\partial z_1^\star} + \frac{\partial h^\star}{\partial z_1^\star} + \frac{\partial u_1^\star}{\partial y_2^\star}\right]\left(-\frac{\partial h^\star}{\partial z_1^\star}\right) + \left[(\lambda^\star + 2\mu^\star)\frac{\partial v_2^\star}{\partial y_2^\star} + \lambda^\star \frac{\partial u_1^\star}{\partial z_1^\star}\right] = 0. \tag{41}$$

In terms of inner variables, these become

$$\text{at } y_2^\star = 0\text{: } \frac{\partial u_1^\star}{\partial y_2^{\star\star}} = 0 \tag{42}$$

and

$$\text{at } y_2^\star = 0\text{: } \frac{\partial \gamma^\star}{\partial z_1^\star}\left\{1 + \left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}\frac{\partial h^\star}{\partial z_1^\star} + \frac{\partial^2 h^\star}{\partial z_1^{\star 2}}\gamma^\star - \Phi^{\star\star}\left\{1 + \left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}^{3/2} + \left\{1 + \left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}^{3/2}\left\{-\mu^\star\left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2 + \left[(\lambda^\star + 2\mu^\star)\frac{\partial v_2^{\star\star}}{\partial y_2^{\star\star}} + \lambda^\star \frac{\partial u_1^\star}{\partial z_1^\star}\right]\right\} = 0. \tag{43}$$

These equations resulting from satisfaction of the jump momentum balance allows us to obtain the configuration of the dividing surface, which in this context is the crack face opening profile.

As $y_2^{\star\star} \to \infty$, we must require that the inner solutions given as (38) and (39) should approach asymptotically the outer solution, (23) and (24):

$$\text{as } y_2^{\star\star} \to \infty\text{: } u_1^\star \to -\frac{\lambda^\star(3\lambda^\star + 2\mu^\star)}{4(\lambda^\star + \mu^\star)^2}\sigma_0^\star z_1^\star \tag{44}$$

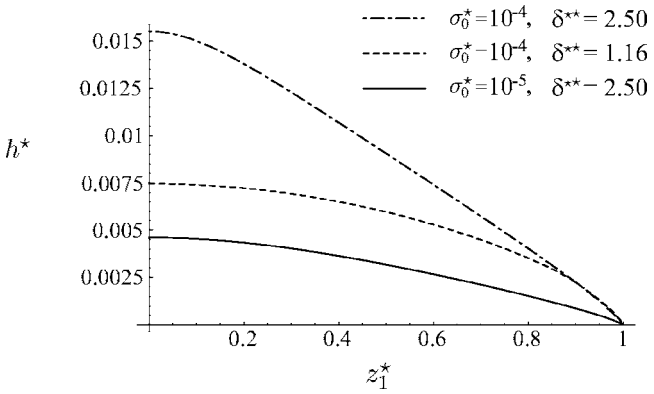


Fig. 3 Dimensionless crack configurations $h^\star$ for different $\sigma_0^\star$ and $\delta^{\star\star}$. Here we have used E=100 GPa, δ=0.2 nm [11,25,31] and a=10 nm.

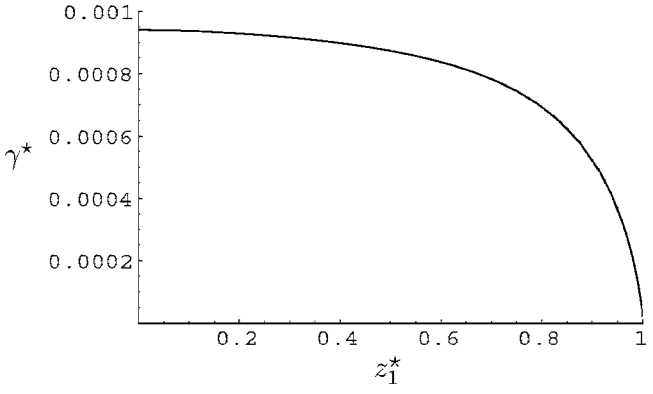


Fig. 4 Dimensionless surface energy $\gamma^\star$ for $\sigma_0^\star=10^{-4}$ and $\delta^{\star\star}=2.50$. Here we have used E=100 GPa, δ=0.2 nm [11,25,31], and a=10 nm.

$$T_{22}^{(I,bulk)^\star} \to \sigma_0^\star. \tag{45}$$

In view of (42), (44), and (45), the inner solutions given as (38) and (39) become

$$u_1^\star = -\frac{\lambda^\star(3\lambda^\star+2\mu^\star)}{4(\lambda^\star+\mu^\star)^2}\sigma_0^\star z_1^\star \tag{46}$$

and

$$(\lambda^\star+2\mu^\star)v_2^{\star\star} = \int \Phi^{\star\star} dy_2^{\star\star} + \sigma_0^\star\left[1+\frac{\lambda^{\star 2}(3\lambda^\star+2\mu^\star)}{4(\lambda^\star+\mu^\star)^2}\right]y_2^{\star\star} + g(z_1^\star). \tag{47}$$

The indefinite integral in (47) becomes fully specified through appropriate specification of the function $g(z_1^\star)$. Using (46) and (47), we express the $T_{22}^\star$ component of stress in the inner region as

$$T_{22}^{(I,bulk)^\star} \equiv (\lambda^\star+2\mu^\star)\frac{\partial v_2^{\star\star}}{\partial y_2^{\star\star}} + \lambda^\star\frac{\partial u_1^\star}{\partial z_1^\star} = \Phi^{\star\star}+\sigma_0^\star. \tag{48}$$

Substituting (46) and (47) into (43), one has

$$\frac{\partial\gamma^\star}{\partial z_1^\star}\left\{1+\left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}\frac{\partial h^\star}{\partial z_1^\star}+\frac{\partial^2 h^\star}{\partial z_1^{\star 2}}\gamma^\star+\left\{1+\left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}^{3/2} \times\left\{\sigma_0^\star-\mu^\star\left(\frac{\partial h^\star}{\partial z_1^\star}\right)^2\right\}=0 \tag{49}$$

where $\gamma^\star$ can be calculated from (22) and (34). From (49) one can obtain the configuration of the fracture surfaces. This is a second-order differential equation to which we impose the two boundary conditions:

$$\text{at } z_1^\star=0\text{: } \frac{\partial h^\star}{\partial z_1^\star}=0 \tag{50}$$

$$\text{at } z_1^\star=1\text{: } h^\star=\delta^\star/2 \tag{51}$$

the first of which reflects symmetry of the crack profile and the second the fact that the crack face separation distance tends to zero at the crack tips. This two-point boundary value problem was solved using the MATHEMATICA software package for a variety of values of the Hamaker constant and loading magnitudes. These results are described in the following section.

3.3 Discussion of Results. Figure 3 shows three dimensionless fracture configurations. Notice that as $\sigma_0^\star$ increases, $h^\star$ increases at the centerline. As the correction for long-range intermolecular forces characterized by $\mathcal{A}^\star$ decreases, i.e., $\delta^{\star\star}$ increases, $h^\star$ also increases at the centerline. Notice also that the predicted crack face profile $h(x)$ is indeed a slowly varying function as assumed a priori. Rigorous verification of this property is demonstrated in a forthcoming paper devoted to a variety of analytical issues associated with the present study.

Traditionally, this problem would be solved in the reference configuration without either a correction for intermolecular forces or surface energy [26–29]. Such an analysis exhibits a blunt fracture tip, whereas the model presented here predicts a crack face opening profile, $h(z_1^\star)$, with finite slope at the fracture tips. Moreover, referring to Fig. 4, we see that, as the fracture tip is approached, the dimensionless surface energy $\gamma^\star\to 0$ and its derivative with respect to $z_1^\star$ significantly increases. It is clear from (49) that both $\gamma^\star$ and $\Phi^{\star\star}$ play important roles in determining the fracture configuration in the neighborhood of the tip.

Finally Fig. 5 shows the corresponding stress on the axis of the fracture as the tip is approached. The most important point to notice is that, as opposed to the traditional view discussed above, the stress remains finite as the fracture tip is approached.

4 Summary

Elaborating on our previous extension of continuum mechanics to the nanoscale [11], we have discussed the role of long-range intermolecular forces in the mechanics of a mode I fracture. We have drawn three conclusions.

1. Contrary to the traditional view [26–30] the fracture tip is sharp rather than blunt.
2. Contrary to the traditional view, there is no stress singularity at the fracture tip.

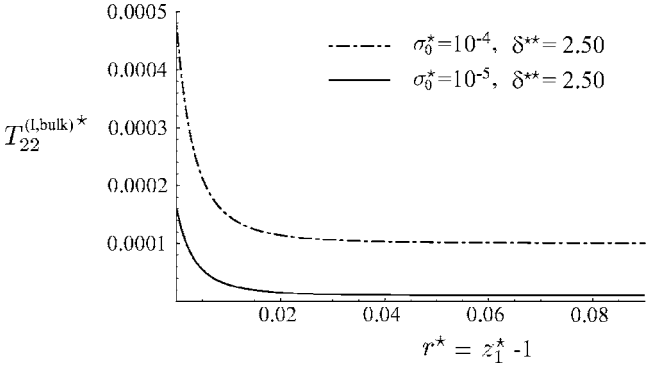


Fig. 5 Dimensionless stress distribution $T_{22}^{(I,bulk)^\star}$ on the fracture axis for $\sigma_0^\star=10^{-4}$ and $\delta^{\star\star}=2.50$. Here we have used E=100 GPa, δ=0.2 nm [11,25,31], and a=10 nm; $r^\star$ indicates the distance from the crack tip.

3. It is important to recognize that the surface energy is nonzero and that it is a function of position in the fracture surface.

Acknowledgment

The authors acknowledge the support provided by the National Science Foundation under grant CMS-030927 and the Air Force of Scientific Research under grant F49620-03-1-0068. We are also grateful for the many extended conversations with Mr. Kaibin Fu.

Nomenclature

a = crack length
$\mathcal{A}$ = Hamaker constant
$\mathbf{b}^{(A,corr)}$ = body force per unit volume at a point in phase A. This is introduced to correct for the use of bulk material behavior in the interfacial region
$\mathbf{e}$ = infinitesimal stress tensor
E = Young's modulus
h = configuration of the A-C interface
H = mean curvature of a dividing surface
$n^{(A)}$ = number density at the specified point in phase A
$R_{(A)}$ = region occupied by phase A
$\mathbf{T}$ = bulk stress tensor using bulk description of material behavior
$\mathbf{T}^{(I)}$ = stress tensor using real description of material behavior
$\mathbf{T}^{(I,bulk)}$ = stress tensor using bulk description of material behavior, corrected for intermolecular forces from the adjoining phase
$\mathbf{v}^{(I)}$ = velocity using real description of material behavior
$\mathbf{v}^{(I,bulk)}$ = velocity using bulk description of material behavior, corrected for intermolecular forces from the adjoining phase
$\mathbf{v}^{(\sigma)}$ = surface velocity, velocity on a dividing surface

Greek Letters

$\boldsymbol{\xi}^{(\alpha)}$ = unit normal to the interface pointing into phase α
γ = surface tension (or energy)
γ^{∞} = surface tension (or energy) that corresponds to static, unbounded diving surfaces
δ = separation distance between the two phases
λ, μ = Lame constant
$\rho^{(I)}$ = bulk density using real description of material behavior
$\rho^{(I,bulk)}$ = bulk density using bulk description of material behavior, corrected for intermolecular forces from the adjoining phase
σ_0 = external stress
τ = distance measured along the normal to the dividing surface
$\phi^{(AC)}$ = potential energy for two molecules A and C separated by a distance
$\Phi^{(A,corr)}$ = net correction for intermolecular potential at a point in phase A

References

[1] Broberg, K. B., 1999, *Cracks and Fracture*, Academic Press, San Diego.
[2] Fineberg, J., Gross, S. P., Marder, M., and Swinney, H. L., 1991, "Instability in Dynamic Fracture," Phys. Rev. Lett., **67**(4), pp. 457–460.
[3] Marder, M., and Gross, S., 1995, "Origin of Crack Tip Instabilities," J. Am. Soc. Mass Spectrom., **43**(1), pp. 1–48.
[4] Abraham, F. F., Brodbeck, D., Rudge, W. E., and Xu, X., 1997, "A Molecular Dynamics Investigation of Rapid Fracture Mechanics," J. Am. Soc. Mass Spectrom., **45**(8), pp. 1595–1619.
[5] Abraham, F. F., Brodbeck, D., Rudge, W. E., Broughton, J. Q., Schneider, D., Land, B., Lifka, D., Gerner, J., Rosenkrantz, M., Skovira, J., and Gao, H., 1998, "*Ab Initio* Dynamics of Rapid Fracture," Modell. Simul. Mater. Sci. Eng., **6**, pp. 639–670.
[6] Holland, D., and Marder, M. P., 1998, "Ideal Brittle Fracture of Silicon Studied With Molecular Dynamics," Phys. Rev. Lett., **80**(4), pp. 746–749.
[7] Slepyan, L. I., Ayzenberg-Stepanenko, M. V., and Dempsey, J. P., 1999, "A Lattice Model for Viscoelastic Fracture," Mech. Time-Depend. Mater., **3**, pp. 159–203.
[8] Abraham, F. F., and Gao, H., 2000, "How Fast Can Cracks Propagate?" Phys. Rev. Lett., **84**(14), pp. 3113–3116.
[9] Abraham, F. F., 2001, "The Atomic Dynamic of Fracture," J. Am. Soc. Mass Spectrom., **49**, pp. 2095–2111.
[10] Swadener, J. G., Baskes, M. I., and Nastasi, M., 2002, "Molecular Dynamics Simulation of Brittle Fracture in Silicon," Phys. Rev. Lett., **89**(8), Ad. No. 085503.
[11] Slattery, J. C., Oh, E.-S., and Fu, K., 2004, "Extension of Continuum Mechanics to the Nanoscale," Chem. Eng. Sci., **59**, pp. 4621–4635.
[12] Fu, K., Robinson, R. L., and Slattery, J. C., 2004, "An Analysis of Supercritical Adsorption in the Context of Continuum Mechanics," Chem. Eng. Sci., **59**, pp. 801–808.
[13] Tadmor, E. B., Philips, R., and Ortiz, M., 1996, "Mixed Atomistic and Continuum Models of Deformation in Solids," Langmuir, **12**, pp. 4529–4534.
[14] Shenoy, V. B., Miller, R., Tadmor, E. B., Philips, R., and Ortiz, M., 1998, "Quasicontinuum Models of Interfacial Structure and Deformation," Phys. Rev. Lett., **80**(4), pp. 742–745.
[15] Knap, J., and Ortiz, M., 2001, "An Analysis of the Quasicontinuum Method," J. Mech. Phys. Solids, **49**(9), pp. 1899–1923.
[16] Xiao, S. P., and Belytschko, T., 2004, "A Bridging Domain Method for Coupling Continua With Molecular Dynamics," Comput. Methods Appl. Mech. Eng., **193**, pp. 1645–1669.
[17] Hein, P., 1914, "Untersuchungen Über den Kritischen Zustand," Z. Phys. Chem., **86**, pp. 385–410.
[18] Winkler, C. A., and Maass, O., 1933, "Density Discontinuities at the Critical Temperature," Can. J. Res., **9**, pp. 613–629.
[19] Maass, O., 1938, "Changes in the Liquid State in the Critical Temperature Region," Chem. Rev. (Washington, D.C.), **23**, pp. 17–28.
[20] McIntosh, R. L., Dacey, J. R., and Maass, O., 1939, "Pressure, Volume, Temperature Relations of Ethylene in the Critical Region II," Can. J. Res. Sect. B, **17**, pp. 241–250.
[21] Palmer, H. B., 1952, "Schlieren Optical Studies of the Critical Region of Pure Substances," Ph.D. thesis, University of Wisconsin, Wisconsin.
[22] Slattery, J. C., 1990, *Interfacial Transport Phenomena*, Springer-Verlag, New York.
[23] Hamaker, H. C., 1937, "The London-Van der Waals Attraction Between Spherical Particles," Physica (Utrecht), **4**, pp. 1058–1072.
[24] Lifshitz, E. M., 1956, "The Theory of Molecular Attractive Forces Between Solids," Sov. Phys. JETP, **2**, pp. 73–83.
[25] Israelachvili, J. N., 1991, *Intermolecular and Surface Forces*, Academic Press, London.
[26] Westergaard, H. M., 1934, "Stresses at a Crack, Size of the Crack and the Bending of Reinforced Concrete," Proc. Am. Concrete Inst., **30**, pp. 93–102.
[27] Westergaard, H. M., 1937, "Bearing Pressures and Cracks," J. Appl. Mech., **6**, pp. A.49–A.53.
[28] Williams, M. L., 1957, "On the Stress Distribution at the Base of a Stationary Crack," J. Appl. Mech., **24**, pp. 109–114.
[29] Williams, M. L., 1959, "The Stresses Around a Fault or Crack in Dissimilar Media," Bull. Seismol. Soc. Am., **49**, pp. 199–204.
[30] Hirschfelder, J. O., Curtiss, C. F., and Bird, R. B., 1954, *Molecular Theory of Gases and Liquids*, Wiley, New York, corrected with notes added, 1964.
[31] Israelachvili, J. N., 1973, "Van der Waals Dispersion Force Contribution to Works of Adhesion and Contact Angles on Basis of Macroscopic Theory," J. Chem. Soc., Faraday Trans. 2, **69**, pp. 1729–1738.

Computational Fracture Mechanics Analysis of Truck Tire Durability[1]

X. Allan Zhong
Technical Excellence Group,
Halliburton Energy Services,
2601 Belt Line Road,
Carrollton, TX 75006
e-mail: allan.zhong@halliburton.com

A three-dimensional fracture mechanics model is formulated to study fatigue crack growth and durability in tires. The application of this model in a radial medium truck tire reveals fracture characteristics of belt edge cracks and helps to explain mechanical and material changes in the tire subject to indoor accelerated durability tests. Along with a proprietary fatigue crack growth law, a fracture mechanics based durability analysis methodology is developed and successfully applied to rank durability of tires with different constructions or different rubber materials. [DOI: 10.1115/1.2069983]

1 Introduction

Tire durability, in one way or another, is frequently related to the fracture of tire components, such as belt edge separation, ply ending separation, or lug cracking in radial medium truck (RMT) tire [1]. Fatigue of tires, however, is a very complicated process. In addition to fracture, it involves microdamage accumulation and mechanical softening of rubber in the early stage of the fatigue process; it also involves temperature rise, and chemical aging. Chemical aging can be oxidation, or anaerobic aging which involves elastomer cross-link redistribution, among other aging processes. Cyclic deformation and associated temperature rise play dominant roles in tire fatigue process. Temperature plays an even more significant role when its elevation is large, such as in tires under severe service condition or under accelerated durability testing conditions. This is due to the fact that when temperature is high, significant chemical changes can take place in a relatively short time, leading to rapid degradation of rubber in tires.

Due to the complexity of the tire fatigue process, the tire durability analysis has been limited to some primitive approaches [2]. For belt edge separation induced fatigue failure in truck tires, such as RMT tires, interlaminar shear, which is a shear component in radial-meridian directions, has been used as durability indicator for a very long time. This approach is good when designs under consideration have only small difference in construction and materials at belt edge remain the same. Its limitation is obvious, it is limited to a single location—belt edge of tires, and it cannot be used to compare the durability of tires with different rubber compound. Other stress or strain measures of a similar nature, such as strain energy density, have been proposed as a durability indictor, to avoid the limitation of interlaminar shear. This type of durability indicators is *independent of materials*, and not directly related to failure mechanisms. Consequently when there are relatively large tire construction changes or material changes, this type of indicators may not yield the right durability ranking. Temperature can be a major factor for tire durability when it is high. Temperature at a critical region is proposed as an alternate durability indicator. It again suffers the same deficiency as the aforementioned simple mechanical indicators. To account for the combined effect of mechanical loading and temperature rise the *S-N* curve approach, which is a rudimentary method in metal fatigue analysis, was proposed for application in tires. This approach is based on the *S-N* curve of rubber composite, which has a construction similar to that at a critical region of interests in tires. This approach is quite appealing to many tire engineers since it fits their intuition. But the *S-N* curve of rubber composite is a structural response, not a material response. Relating the *S-N* curve of a composite sample to a corresponding tire is problematic theoretically and numerically.

Fracture mechanics, on the other hand, provides a fundamental approach to analyze the fracture and fatigue crack growth in tires. Numerically a major advantage of the three-dimensional (3D) fracture mechanics analysis approach is that it is applicable to all types of tires, and to any fracture mode or fracture location. It can be used to evaluate tire component fracture resistance, to predict the direction of potential crack growth, to study fracture mechanisms, and to assess tire durability ranking of different designs. Applications of fracture mechanics in rubber can be traced back to the work of Rivilin and Thomas [3] in early 1950s and the work of Gent et al. on fatigue crack growth [4,5] in 1960s. The significance of fracture mechanics to the evaluation of tire cracking and tire durability was recognized previously by Ebbott [1]. In open literature, most tire fracture mechanics analyses are performed on two-dimensional models, e.g., Ref. [1]. There are limited 3D fracture analyses of tires utilizing whole tire models, e.g., the work by Wei et al. [6] and Govindjee [7,8]. However, a straightforward finite element analysis (FEA) of tires with explicit cracks can be difficult due to convergence issues in numerical computations, especially for heavy tires such as RMT tires. There are few applications of 3D fracture mechanics in tire durability analysis reported in literature, one of them is by Govindjee [7].

To have a reliable fracture mechanics based durability analysis approach, one has to consider all dominant factors in tire durability. First of all, the rubber property evolution during tire life has to be accounted for in the analysis. The evolution of rubber properties involves two aspects, i.e., rubber constitutive properties and failure properties. However, it is unrealistic to describe rubber material evolution real time for life prediction of a RMT tire whose life is typically several hundred thousand miles, i.e., hundreds of millions of cycles, under normal service conditions. Major assumptions are made for an effective durability analysis.

Under cyclic loading condition, rubber reaches steady state stress-strain response in 5–10 cycles. Rubber stress-strain behavior will deviate from steady state at the end of its fatigue life. So the "appreciable" evolution of rubber constitutive behavior occurs only in a tiny portion of its fatigue life. Based on this observation, a major assumption is that under nominal conditions, rubber failure property evolution is the dominant factor, rubber constitutive

[1]This work was performed while the author was employed at the research division of the Goodyear Tire & Rubber Company. Scales in some figures in this paper are intentionally removed to protect Goodyear trade secrets.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 29, 2004; final manuscript received June 10, 2005. Review conducted by S. Govindjee. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

property evolution can be neglected. This assumption significantly simplifies the analysis of tire response in its life span. Instead of analysis of a continuous tire damage evolution process, only a "steady state" analysis of tire response under discrete loading steps is needed. Rubber constitutive properties in a loading step are determined via a steady state rubber response at the corresponding deformation level. Rubber failure property evolution is dictated by tire thermal mechanical responses. It determines tire durability.

One of the main features of the rubber cyclic deformation behavior is that the rubber modulus is dependent of strain amplitude in a cycle. Typical hypereleastic models for rubber cannot describe this behavior well. In typical tire analysis at Goodyear, a constitutive model, which is a modified neo-Hookean constitutive relation, is developed to describe this rubber behavior effectively by Ebbott et al. [9]. A thermal mechanical tire finite element analysis procedure based on the constitutive model has been developed and extensively validated. The numerical procedure used in the tire fracture mechanics analysis approach presented in this paper is built on the aforementioned procedure.

This paper is organized as following. The formulation of a (truck) tire fracture mechanics model is presented in Sec. 2. The results of the fracture mechanics analysis of a straight rolling truck tire under vertical load are presented in Sec. 3. In Sec. 4, the fracture mechanics based tire durability analysis approach is described. The emphasis is placed on the general description of modeling rubber fracture property evolution due to chemical aging. Some results of durability analysis for truck tires subject to indoor accelerated durability tests are presented to demonstrate the effectiveness of the fracture mechanics approach. The paper concludes with a brief discussion on some issues related to tire durability analysis under more complex conditions.

2 Formulation of Nonlinear Truck Tire Fracture Mechanics Models

In tire fracture mechanics, the general *energy release rate concept is used to describe the crack growth driving force* since tires are complex nonhomogeneous composite structures. Energy release rate G by definition is the rate of change in potential energy with respect to the newly created crack area. It is a measure of the energy available for an increment of crack extension. Suppose Π is the potential energy of a material system, and A is the fracture surface area, then

$$G = -\frac{d\Pi}{dA} \tag{1}$$

Numerically, the virtual crack closure technique (VCCT) is used to calculate the strain energy release rate at a crack tip in a tire. The method is based on the fact that the work required closing a crack an infinitesimal distance is equal to the energy released during an infinitesimal advance of the crack as discussed by Irwin [10]. For a review of the technique in tire fracture mechanics analysis, one may refer to Ref. [6]. The energy release rate calculated from the VCCT is solely determined by local deformation and load transfer, and therefore, is a local quantity. It is worthwhile to mention that local energy release rate has been applied to study failure mechanisms in other material systems successfully [11,12]. Using the VCCT, there is no need to resort to a new concept. The application of the VCCT in rubber composites has been validated at Goodyear.

Under a monotonic load, crack growth is determined by the critical energy release rate. Under cyclic loading conditions, crack growth is also called fatigue crack growth. Fatigue crack growth is typically determined by a fatigue crack growth law. One may refer to Ref. [13] for typical features of fatigue crack growth in rubber. Commonly used fatigue crack growth law has a power law form

$$\frac{dc}{dn} = A\Delta G^m \tag{2}$$

This equation is often referred as Paris' law, where ΔG is the range of energy release rate in a cycle, dc/dn is crack growth rate, A, m are material constants. It is noted that Paris' law is utilized in Ref. [7].

The proprietary fatigue crack growth law developed at Goodyear is a modification of Paris' law, which accounts for other factors that also influence fatigue crack growth, such as the R ratio (ratio between minimum and maximum energy release rate in a cycle), operating temperature, and chemical change. We will discuss this law in Sec. 4.

Basic Assumptions on Crack Placement. There are many locations in a tire where a crack might exist. However, cracks at limited locations play dominant roles in the tire failure process. To perform an efficient and effective tire fracture mechanics analysis, one has to identify critical locations first.

Critical Locations. The critical locations, such as belt edge and ply ending, can be determined either from experience, such as observations of tire failures in the field or in lab tests, or from finite element analysis of tires. Regions with highly localized deformation or stress concentrations or abrupt changes in material stiffness, such as the interfaces between cords and rubber, are likely locations for cracks. In tire FEA modeling, the critical locations can be selected based on maximum (cyclic) strain energy density, maximum (cyclic) principal stress, or maximum (cyclic) principal strain depending on material properties and local geometry.

Crack Orientation. The question of crack orientation can be easily resolved if there are observations of typical failures in the tire under consideration. Otherwise, one has to make an assumption. Since shear deformation is dominant in tire components, one can assume that a *macroscopic* crack orients in the direction of maximum shear at the critical location. This assumption typically leads to a crack parallel to tire component interfaces in a tire cross section. Depending on tire types and loading conditions, some tire components can be subject to tensile deformation, such as the lower apex in some RMT tires. In this case, the crack can be perpendicular to component interfaces, or perpendicular to the maximum principal tension at the location.

Crack Size and Shape. To have effective fracture mechanics analysis of tires, the fracture resistance of the critical regions should be evaluated at the early stage of crack growth. One typical way is to assume that the initial crack size in a tire cross section is of the order of the smallest characteristic length in a critical region. Near reinforcing belts, the smallest characteristic length scale is the effective steel cord radius or diameter. In tread groove roots, it is the smallest radius (inverse of the largest curvature) of the root. In regions where there is no apparent characteristic length scale, such as the apex (a tire component), one might consider taking one tenth of the smallest component dimension.

As for the overall geometry of the 3D crack, one can assume that the crack is of the shape of a closed ribbon if a steady-state scenario is considered (see illustration in Fig. 1(*a*)). The reason for this assumption is that all material points on a ring in the tire circumferential direction experience the same loading history and deformation, if there is a crack initiated at one point on the ring, one would expect that cracks are initiated along the whole ring. This is the simplest representation of a crack in a tire from the numerical standpoint.

Realistically, one usually does not see a crack of the shape displayed in Fig. 1(*a*). For example, for a crack at the belt edge of a tire, one is more likely to see a few disjoined cracks in the circumferential direction, as illustrated in Fig. 1(*b*). In a lab test or service, the disconnection between cracks is due to inhomogeneity of materials, nonuniformity of geometry, or random distributed

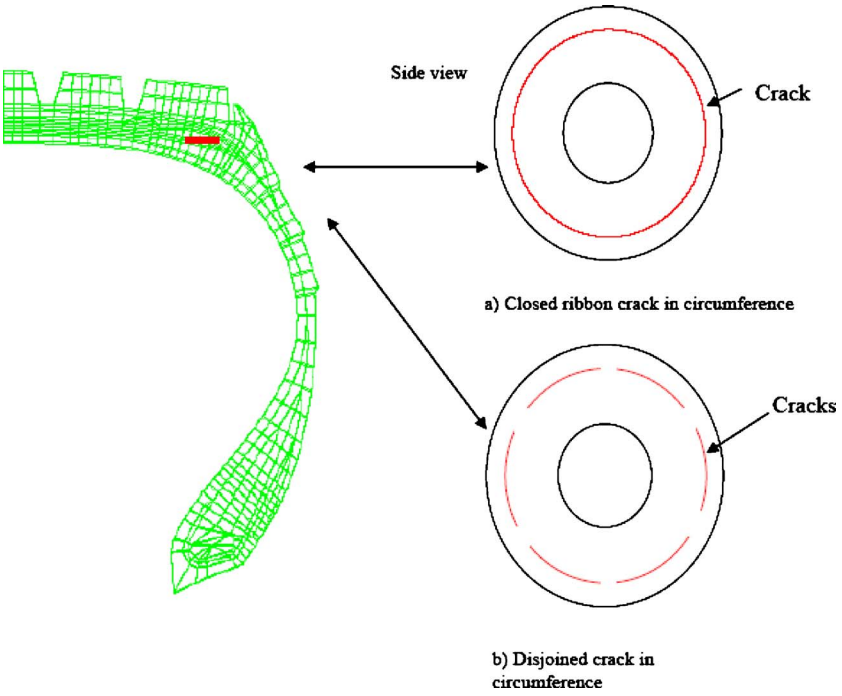


Fig. 1 A buried crack at belt edge (*a*) of the shape of a closed ribbon; (*b*) disjoined cracks in circumference

microcracks from manufacturing among other factors. The more realistic cracks illustrated in Fig. 1(*b*) are harder to model than the one illustrated in Fig. 1(*a*). In our fracture mechanics tire models, a closed ribbon-shape crack is assumed. This assumption is good unless tire noninfirmities are of concern.

In summary, to perform a tire fracture mechanics analysis, an explicit crack should be included in a tire model. The initial crack is assumed to be located in the critical region identified, and to orient in a predetermined direction. The crack is assumed to be of the shape of a closed ribbon with the size (in a tire cross section) of a certain characteristic length determined from geometrical features of the critical region. In the following, the fracture mechanics analysis of a RMT tire is conducted. In this RMT tire, the durability is primarily influenced by belt edge separation.

3 Fracture Mechanics Characteristics of a Belt Edge Crack in a RMT Tire

Using the formulation of a tire fracture mechanics model outlined in Sec. 2, a comprehensive study of a RMT tire with a belt edge crack is conducted. The tire has four belts (numbered from cavity to tire crown), see Fig. 2, and the crack is located at the belt 3 edge. A half tire cross-section model[2] is utilized since only straight rolling tire subject to vertical load are considered. It should be noted that a full tire cross-section model has to be utilized if a more complex load is applied on the tire, e.g., vertical and lateral load. A typical cross-sectional mesh of the tire fracture mechanics model is shown in Fig. 3. Local mesh refinement at the crack tip is used instead of uniform mesh refinement, and mesh density at the crack front should meet the VCCT requirement. Typically the element size at the crack tip is about one quarter of the local characteristic length scale, i.e., the wire radius for belt edge region. Furthermore, the crack surfaces are assumed to slide against each other frictionless. In the following discussions, the left crack front in Fig. 3 is referred to as the inboard crack tip, and the right crack front is referred to as the outboard crack tip.

Energy Release Rate at the Crack Fronts. It is noted that energy release rate can be decomposed into components in three orthogonal directions per the VCCT technique, thus one can determine the fracture mode from the calculation. The crack at belt 3 edge of the RMT tire is primarily of mode III, or tearing mode when the tire is under vertical footprint load. In Fig. 4, the total energy release rates at the inboard and outboard crack tips of a 0.1 in. (2.54 mm) crack at the belt 3 edge of the tire, that is under rated vertical load of 6715 lb (3045 kg) and tire cavity pressure of 110 psi (0.76 MPa), are displayed. It is noted that $c1=c2=0.05$ in. (1.27 mm) for this crack (see Fig. 3 for $c1, c2$ definition). In the figure the angle refers to the circumferential location of a position in the tire. Zero degree corresponds to the center of footprint of the tire, and the top of a tire is 180 deg from the footprint. Two major observations can be made from this figure: (1) the energy release rate is highest near the footprint center at both crack fronts; (2) the peak energy release rate at the inboard crack tip is much higher than that at the outboard crack tip for this

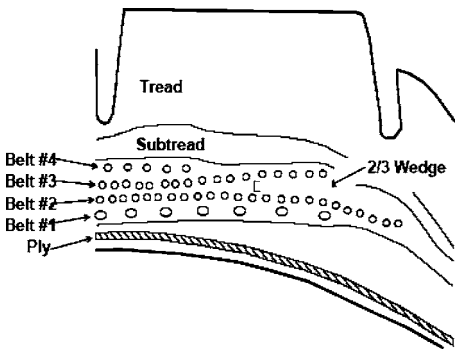


Fig. 2 Illustration of crown area construction of a RMT tire

[2]Skew symmetry exists at the centerline in typical tires under static vertical load or in a straight rolling tire under vertical load, and this skew symmetry should be accounted for in the half tire model.

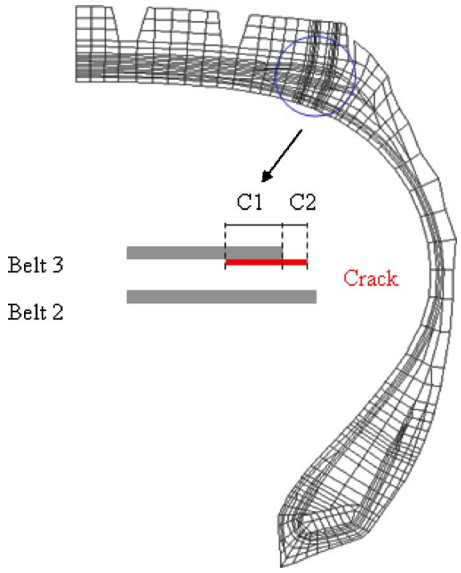


Fig. 3 Representative mesh in tire cross section for fracture mechanics analysis. The inlet is a description of the crack location with respect to the belt edge.

small crack. The second observation has significant implications in the direction the small crack will grow in a fatigue process. This will be discussed further in Sec. 4.

Effect of Crack Size on Interlaminar Shear Between Belt 2 and 3. Interlaminar shear at the belt edge has been an important belt edge durability indicator in the tire industry. The change of the interlaminar shear magnitude at the belt edge with tire mileage in an indoor accelerated durability test has been used as an argument in the debate about whether the tire fatigue process is a crack initiation dominant process or a crack propagation dominant process.

There are some experimental data available for interlaminar shear (ILS) measurement at the belt edge area at different mileage in an accelerated tire durability test. The comparison of the effect of the crack size on ILS as predicted by the fracture mechanics model to that measured experimentally in the tested tires should shed lights on the debate, at least for the accelerated durability test condition. In the following the ILS prediction from the fracture mechanics model is determined from an element right beneath the belt 3 edge in the wedge, (below the crack in the models), where the ILS is measured. Figure 5 shows the effect of the belt edge crack size on ILS. As expected, the larger the belt edge crack size, the lower the ILS and cyclic ILS, i.e., the difference of maximum ILS and minimum ILS along the circumference. This is because the wedge compound right beneath the belt 3 edge is unloaded due to the existence of a belt edge crack. It is noted that crack size

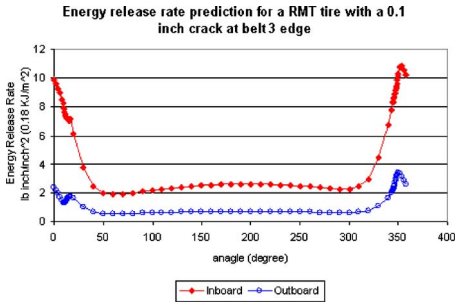


Fig. 4 Energy release rate distribution around the circumference at the belt edge crack tips in a RMT tire

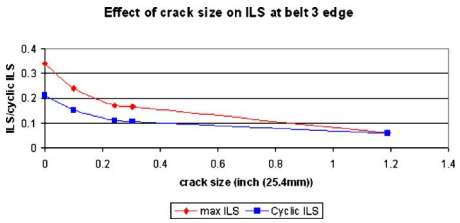


Fig. 5 Predicted effect of the belt edge crack size on the belt edge ILS and cyclic ILS

and its relative position with respect to the belt edge (see Fig. 3) is determined from a fatigue crack growth law, except the initial very small crack (which is the size of the effective radius of the steel wires in the belt, ~0.016 in. (0.4 mm). The initial crack is symmetric with respect to the belt edge, i.e., $c1=c2$.

In tire durability tests, it takes time for the belt edge crack to grow, so it is reasonable to assume that the larger crack size corresponds to the longer test mileage. Qualitatively, the predicted results shown in Fig. 5 compared to test measurements were very well for both max ILS and cyclic ILS, see Fig. 6. When comparing Figs. 5 and 6, one should note that the loading conditions are slightly different. In the numerical analysis rated single position load 6175 lb (3045 kg) vertical load and 110 psi(0.76 MPa) inflation is utilized, while in the measurement of fatigued tires, 5500 lb (2495 kg) vertical load and 90 psi (0.62 MPa) inflation is applied. It should be pointed out also that initial increase in the measured ILS value is due to rubber cyclic softening. Rubber cyclic softening is neglected in the current study. The good correlation between the fracture mechanics analysis results and the test measurement may indicate that there are belt edge cracks in the tires. As a matter of fact, loose cords are observed in the tested tires, as well as significant separation between the cord and rubber compound for a tire removed at ~ 7000 miles (11 265 km) from the test. Based on experimental observations, measurements, and fracture mechanics analysis, it is estimated that a *discernible* crack is formed in the tire under the test condition around 1000 miles (1609 km). Since the average life for the tire under the test condition is around 12 000 miles (19 311 km), it is determined that crack propagation accounts for at least 90% of the tire fatigue life in this test. The tire fatigue process is crack propagation dominated in this case.

The correlation between fracture mechanics analysis, the test results, and the conclusions reached are of fundamental importance since it settles, at least partially, a long standing debate over whether crack propagation or crack initiation is dominant in tires under accelerated durability testing conditions. It solidified the foundation of the fracture mechanics based tire durability analysis methodology.

Effect of Crack Size on Temperature Distribution at Belt Edge. Existence of a belt edge crack changes belt edge temperature distribution; the larger the cracks size, the bigger its impact.

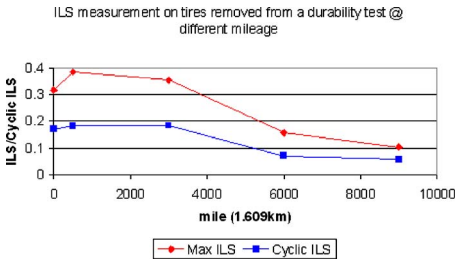


Fig. 6 Experimental measurement of ILS, cyclic ILS at the belt edge of RMT tires removed from an accelerated durability test at different mileage

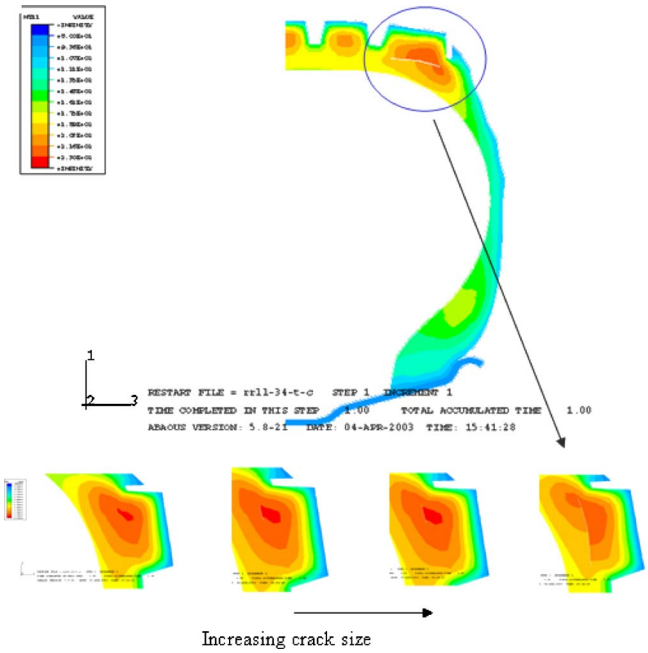


Fig. 7 Effect of the belt edge crack size on thermal conduction

In our models, crack surfaces are assumed to be adiabatic—none conducting because belt edge crack is closed only near footprint—it opens up away from the footprint under load. In the following, representative results are shown to demonstrate the main features of the temperature profiles. It is noted that the procedure for temperature prediction is described in Ref. [9]. In brief the heat source is viscoelastic heating of the rubber compounds in the tire due to cyclic loading. The rolling speed of the tire is 22 mile/h (35.41 km/h).

There is very small difference in temperature at the belt edge between the model without a crack and that with a small crack (less than 0.1 in. (2.54 mm) in size), as expected. When the belt edge crack size is large, its impact on temperature distribution becomes significant. As shown in Fig. 7, due to the existence of the crack (the white line in belt edge area), the temperature above belt 3 is higher than that below it. This observation is typical for all crack sizes. There are a few reasons for this observation. First, the wedge compound is unloaded due to the existence of belt edge crack. The larger the crack, the more it is unloaded and less heat is generated between belts. Second, the belt edge crack prevents heat generated in the tread from being conducted to the wedge area. A summary of temperature near the belt 3 edge with different crack sizes is given in Table 1.

While the belt edge temperature decreases with the increase of the crack size, it is observed that the overall temperature at the lower sidewall of the tire increases and the overall rolling resistance changes little with crack size.

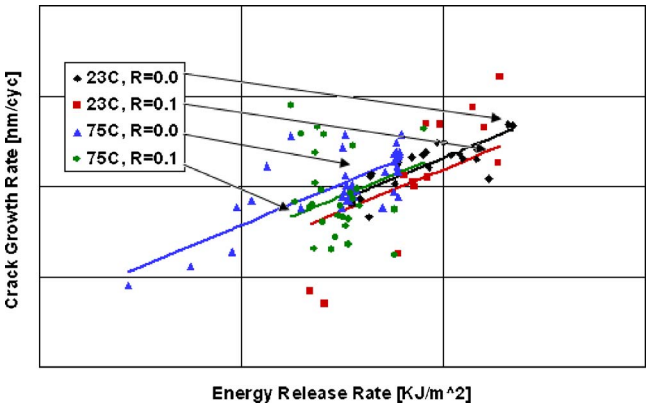


Fig. 8 Temperature effect and *R*-ratio effect on the crack growth rate in a compound (log-log scale)

4 Fracture Mechanics Based Truck Tire Durability Analysis

The main assumptions made in the fracture mechanics based durability analysis method include the following: (1) The evolution of rubber constitutive property can be neglected (see Sec. 1). Effect of chemical aging on rubber fracture property evolution is a dominant factor in rubber property evolution. Thus only a steady state analysis of tire response under discrete loading steps is needed. (2) The cyclic deformation at a material point in a tire during a tire revolution can be represented by the deformation on a material ring passing the material point of interests in the tire. (3) Crack size change due to fatigue is determined by Goodyear's proprietary fatigue crack growth law to be discussed. (4) Fatigue crack propagation is modeled through different mesh/model with different crack sizes. In this study, the rubber fatigue resistance is characterized by the fatigue crack growth rate for a given energy release rate cycle..

Literature [14,15] and Goodyear in-house test results indicated various degrees of sensitivity of the crack growth rate to temperature and *R* ratio (the ratio of minimum energy release versus maximum energy release at crack tip in a cycle) see Fig. 8. Despite the scatter of crack growth data, we have consistently observed the trends in various rubber compounds as indicated by the trend lines in Fig. 8, which is that a higher temperature leads to faster crack growth and a higher R ratio of the energy release rate leads to slower crack growth.

The crack growth rate is also affected by chemical aging. A sufficiently elevated temperature applied over time causes chemistry changes or chemical aging in rubber. Aging effects involve permanent changes in the chemistry of the material due to evolving processes such as crosslink changes or reversion. *At the belt edge of the truck tire under consideration, rubber aging is anaerobic*.

So the dominant factors that influence fatigue crack growth in tires include: energy release rate variation in a cycle of deformation, *R* ratio of the energy release rate, operating temperature, and

Table 1 Predicted temperature in the tire with different belt edge crack sizes

Crack size	Cavity	Wedge compound right beneath the belt 3 edge	Inboard crack tip	Outboard crack tip
0.10 in. (2.54 mm)	155.4 F (68.5 °C)	229.1 F (109.5 °C)	229.6 F (109.8 °C)	230.5 F (110.30°C)
0.24 in. (6.10 mm)	155.7 F (68.7 °C)	227.0 F (108.3 °C)	228.1 F (108.9 °C)	229.0 F ((109.4 °C)
0.3 in. (7.62 mm)	156.3 F (69.0 °C)	226.2 F (107.9 °C)	227.7 F (108.7 °C)	227.8 F (108.8 °C)
1.29 in. (32.77 mm)	156.0 F (68.9 °C)	214.7 F (101.5 °C)	194.0 F (90.0 °C)	220.5 F (104.7 °C)

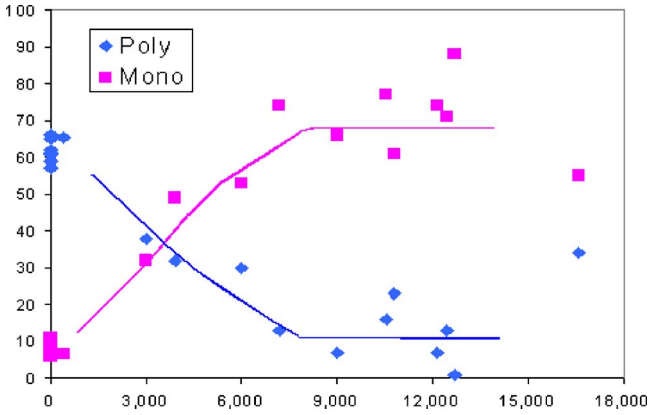


Fig. 9 Change of cross-link distribution vs test mileage (linear-linear scale)

chemical aging of elastomer, at the crack tip. A proprietary fatigue crack growth law, which accounts for the four primary factors, has been developed at Goodyear, which can be represented by the following functional format

$$\frac{dc}{dn} = A \cdot f(\Delta G, R, T, \text{Aging}) \tag{3}$$

The unique feature of this crack growth law is that it provides a mathematical way to describe the effects of chemical aging on fatigue crack growth rate. The key assumption for this description is that fracture (tear) property changes track changes in polysulfidic crosslink density. This assumption is supported by experimental observations of polysulfidic crosslink and monosulfidic crosslink distribution change in the rubber compound at the belt edge of the tire (Fig. 9) under accelerated durability test conditions and that of fracture toughness of the same compound (Fig. 10). It takes a lot of lab fatigue crack growth tests to determine the parameters in the Goodyear fatigue crack growth law, especially tests on properly aged samples.

Fatigue Crack Growth in Tires. The fatigue crack growth at the belt edge of a RMT tire in numerical models starts with a very small crack of the size of an effective wire radius of reinforcement steel wire in a tire cross section. Then the growth of the belt edge crack is determined by the energy release rate at the inboard and outboard crack tips and temperature there. The crack growth rate at either the crack front is calculated via the proprietary fatigue crack growth law, and the corresponding $c1, c2$ (see Fig. 3 for

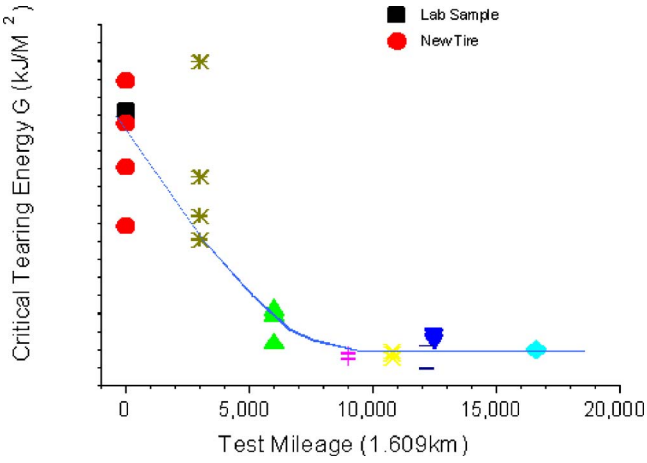


Fig. 10 Decay of critical energy release rate vs test mileage (linear-linear scale)

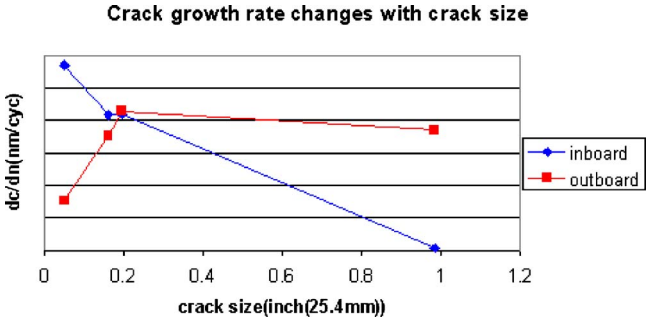


Fig. 11 Fatigue crack growth rate at the belt edge crack tips in a tire under rated load (linear-linear scale)

definition) is subsequently determined. If *the maximum energy release rate at the crack fronts remains smaller than the (evolved) critical energy release rate*, one can use the fatigue crack growth law (Eq. (3)) to calculate the fatigue crack growth rate at both crack fronts. As shown in Fig. 11, when the crack size is smaller than approximately a quarter inch (6.35 mm), the belt edge crack grows faster toward the centerline than toward the shoulder area.

Apparently inboard crack growth is stable for all crack sizes; i.e., crack growth becomes slower for larger crack sizes. Outboard crack growth becomes faster with the increase of crack size initially and then stabilizes. Unlike inboard crack growth, *the outboard crack growth is much more complex than the model represents, due to the fact that outboard crack front can* branch *(i.e., a crack front splits into multiple crack fronts) when it propagates into the shoulder area where no reinforcements exist, and effective crack growth can be substantially lower than predicted here*.

It should be emphasized that Fig. 11 is obtained based on the assumption that the energy release rate remains smaller than the critical energy release rate. But in reality, the critical energy release rate decreases with tire service life (see Fig. 10), the belt edge crack growth can be critical. The crack growth law (Eq. (3)) does not apply to critical crack growth.

Ranking Truck Tire Durability Under Indoor Accelerated Test Conditions. The fracture mechanics based tire durability approach, along with the proprietary fatigue crack growth law, has been successfully applied to many different truck tires under different accelerated tire testing conditions at Goodyear. Here one of the applications is presented to demonstrate the success of the fracture mechanics based tire durability analysis approach.

In this test case, the control tire is the aforementioned RMT tire, four experimental tires are variations of the control tire. Two of the experimental tires have the same tire construction but with a change in rubber compounds at certain locations in the tire. One involves replacing the wedge compound in the control tire with a softer wedge compound; the other involves replacing the tread compound of the control tire with a cooler tread compound (i.e., with lower loss modulus). The other two experimental tires differ from the control tire only in the tire's tread arc radius, one has a flat tread (infinite tread arc radius), the other has a tread arc radius half that of the control tire.

In this fracture mechanics based durability analysis, tire life is designated as the mileage in the simulated test, at which the assumed initial crack grows to a critical size, e.g., 1 in. (25.4 mm). Due to the large "time marching"[3] in calculation of the fatigue crack growth, and many other assumptions made, instead of comparing the predicted tire life to a measured one, rankings of tire durability are compared. The ranking, as shown in Fig. 12, is

[3]Time marching refers to the number of cycles needed from one (model) crack size to the next one. To save time, the number of cycles used is at a million cycle level. For simple screening analysis, only results from the model with the initial crack size are used to project time for the crack to reach a critical size.

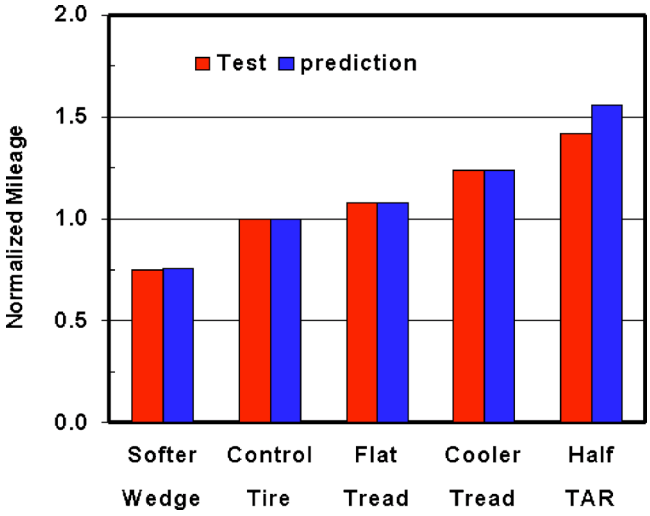


Fig. 12 Comparison of predicted tire durability (mileage) to measurement from tests

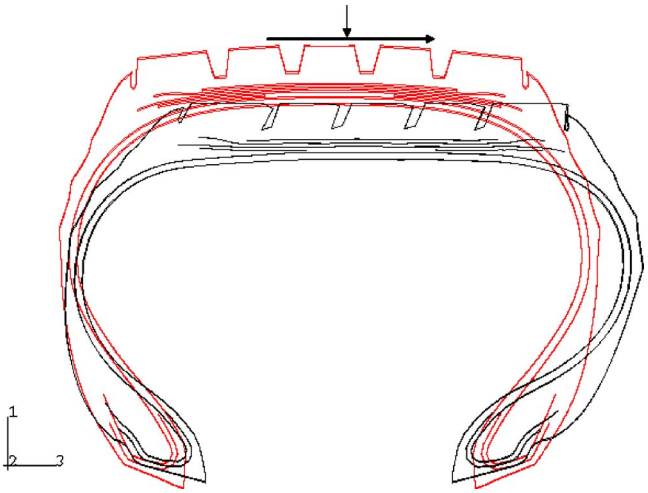


Fig. 13 A tire under combined vertical and lateral load

determined as following. The failure mileage of each tire is normalized with the failure mileage of the control tire for either prediction or test. The fracture mechanics based durability approach predicts the correct ranking and reasonable relative percentage change for this test case.

5 Further Discussions on Analysis of Truck Tire Durability

The tire fatigue process is so complicated that even when experimental data are available, different interpretations can be made to the same data, or a reasonable interpretation cannot be established. In many cases, fracture mechanics provided the missing link in understanding phenomena observed in the tire fatigue process. For example, the decay of interlaminar shear at the belt edge in a tire at increasing test mileage (see Fig. 6), cannot be explained by those who believed that the tire fatigue process is a crack initiation dominant process. With fracture mechanics analysis, the phenomenon was easily explained.

The effect of the belt edge crack size on temperature distribution (Sec. 3) may help us to better understand the observations made in the accelerated durability test regarding wedge compound material degradation. It is observed that the critical tearing energy of the wedge compound for the RMT tire degrades very fast initially and approaches a plateau with the testing mileage, see Fig. 10. The plateau was initially thought to be due to the characteristics of the rubber compound aging chemistry. Current fracture mechanics analysis indicates that the plateau might be due to the change of thermal-mechanical load on the wedge compound—lower mechanical load and lower temperature at wedge compound right beneath the belt 3 edge due to the increasing belt edge crack size.

It should be pointed out that the correlation between the crosslink distribution change and the fracture toughness evolution is dependent of aging chemistry and rubber compounding. When compounding changes, the correlation between rubber crosslinks and mechanical properties will be different from the one presented here. When aging chemistry changes, for example, from anaerobic aging to oxidation, the material structure-property relation will be totally different.

In the tire durability analysis, it is natural to consider one tire revolution corresponds to one fatigue cycle. But in reality, the energy release rate variation at a crack tip can have multiple cycles in one tire rotation from fatigue analysis point of view; see multiple local maximums in Fig. 4. In general, the cycles of deformation at a crack tip has to be counted via special techniques, e.g., rainflow cycle counting method. Appropriate cycle counting is very important in tire durability analysis.

In this paper, only a straight rolling tire is considered when the load on the tire can be approximated as the vertical load only. When a tire changes rolling direction, the load on the tire is much more complex. An approximation of the latter situation is that the tire is subject to a vertical and lateral load. The impact of the lateral load on the belt edge deformation is large. Under combined vertical and lateral load, interlaminar shear at the 2/3 belt edge at the right-hand side of the footprint center (see Fig. 13) was found to increase with increasing lateral load. Interlaminar shear at the 2/3 belt edge at the left-hand side, however, decreased with increasing lateral load. This is because, under lateral loads, the belt package goes through an in-plane bending. The right-hand side experiences additional loading due to tension, while the left-hand side experiences unloading due to compression.

The impact of the lateral load on the energy release rate at a small belt edge crack is significant. It is shown numerically that, under combined rated vertical load and lateral load (which is 35% of the rated vertical load), the maximum energy release rate increased by about 90%, the range of the energy release rate increased 100% when compared to the energy release rate at the rated vertical load only. This significant increase of the energy release rate due to lateral loads explains why lateral loads have more adverse effects on crown area durability than vertical load. However, despite the significant lateral load, the fracture mode of belt edge crack remains to be primarily a tearing mode at both crack fronts.

It is noted that there are other important factors that influence tire durability under service conditions, such as interfacial failure, tire nonuniformity, rubber healing, etc. Studies on these factors are beyond the scope of current work. One can find a more comprehensive discussion by Govindjee [7] on factors that influence tire durability under service conditions.

This study demonstrated that fracture mechanics, along with proper fatigue crack growth law, can rank RMT tire durability under indoor accelerated durability testing conditions. More importantly, fracture mechanics analysis of an as-designed tire provides insights into fundamental tire durability mechanisms due to tire construction and tire material. This fundamental understanding of tire durability mechanisms will be beneficial for identification of directions in design for improving tire durability performance.

Acknowledgment

The author conducted work presented in this paper during his employment at the research division of the Goodyear Tire & Rubber Company. Professor Wolfgang Knauss of the California Institute of Technology was an academic consultant to the research project on RMT tire durability at Goodyear during the time. The author is grateful to the management of Goodyear's research division for its support of fracture mechanics research during the author's tenure there and its approval of a request to present tire fracture mechanics related work. The request was made before the author left the company in 2003. The author benefited from discussions, collaborations with many colleagues at Goodyear, including Tom Ebbott, Tom Fleischman, Ming Du, Ed Terrill, Soojaa Kim, and Haitao Xu.

References

[1] Ebbott, T. G., 1993, "An Application of Finite Element Based Fracture Mechanics Analysis to Cord-Rubber Structure," *12th Annual Meeting and Conference of the Tire Society, 23–24 March 1993*; also appeared in Tire Sci. Technol., **24**(3), pp. 220–235 (1996).

[2] Zhong, X. A., and Ebbott, T. G., 2001, "Aspects of Durability of Elastomeric Structures," *ASME/ASCE/SES the Mechanics and Materials Summer Conference 2001 (MMC2001)*, San Diego, 27–29 June 2001.

[3] Rivilin, R. S., and Thomas, A. G., 1953, "Rupture of Rubber I. Characteristic Energy for Tearing," J. Polym. Sci., **10**(3), pp. 291–318.

[4] Gent, A. N., Lindley, P. B., and Thomas, A. G., 1964, "Cut Growth and Fatigue of Rubber, Part 1: The Relation Between Cut Growth and Fatigue," J. Appl. Polym. Sci., **8**, pp. 455–465.

[5] Lake, G. J., and Lindley, P. B., 1964, "Cut Growth and Fatigue of Rubber, Part 2: Experiments on Noncrystalizing Rubber," J. Appl. Polym. Sci., **8**, pp. 707–712.

[6] Wei, Y.-T., Tian, Z.-H, and Du, X. W., 1999, "A Finite Element Model for the Rolling Loss Prediction and Fracture Analysis of Radial Tires," Tire Sci. Technol., **27**(4), pp. 250–265.

[7] Govindjee, S., 2001, *Firestone Tire Failure Analysis*, see http://www.ce.berkeley.edu/~sanjay/REPORT_WEB_Secure.PDF

[8] Govindjee, S., 2003, "Finite Deformation Fracture of Tires," *Solid Mechanics and Its Applications: IUTAM Symposium on Computational Mechanics of Solid Materials at Large Strains*, C. Miehe, ed., Kluwer, Dordrecht, pp. 417–424.

[9] Ebbott, T., Hohman, R., Jeusette, J.-P, and Kerchman, V., 1999, "Tire Temperature and Rolling Resistance Prediction With Finite Element Analysis," Tire Sci. Technol., **27**(1), pp. 1–8.

[10] Irwin, G. R., 1957, "Analysis of Stress and Strain Near the End of a Crack Transversing a Plate," ASME J. Appl. Mech., **24**, pp. 361–364.

[11] Gao, H., Zhang, T.-Y., and Tong, P., 1997, "Local and Global Energy Release Rate for an Electrically Yielded Crack in Piezoelectric Ceramic," J. Mech. Phys. Solids, **45**(4), pp. 491–510.

[12] Marie, S., and Chapuliot, S., 1998, "Tearing Simulation Based on a Local Energetic Criterion," Fatigue Fract. Eng. Mater. Struct., **21**(2), pp. 215–227.

[13] Gent, A. N., 1992, *Engineering With Rubber*, Hanser, Fachbuch.

[14] Gent, A. N., and McGrath, J. E., 1965, "Effect of Temperature on Ozone Cracking of Rubber," J. Polym. Sci. A **3**(4), pp.1473–1483.

[15] Mars, W. V., and Fatemi, A., 2003, "A Phenomenological Model for the Effect of R Ratio on Fatigue of Strain Crystallizing Rubber," Rubber Chem. Technol., **76**(5), pp. 1241–1258.

On the Crushing Stress of Open Cell Foams

Lixin Gong

Stelios Kyriakides[1]
e-mail: skk@mail.utexas.edu

Research Center for Mechanics of Solids,
Structures & Materials,
The University of Texas at Austin,
WRW 110, Austin, TX 78712

The compressive response of many foams is characterized by an initial linearly elastic regime which is terminated by instability. For open cell foams instability leads to localized buckling and collapse of zones of cells. Local collapse in these zones is terminated by contact between cell ligaments. In the process collapse spreads to neighboring cells hitherto intact. The spreading of collapse occurs at a well-defined load plateau and continues until most of the cells are thus affected when the material response regains stiffness once more. This type of three-regime compressive response was reproduced numerically by idealizing such foams to be assemblages of space-filling Kelvin cells. The onset of instability involves a long wavelength mode. It has been established by considering a fully periodic column of cells tall enough to accommodate this mode. The crushing response has been evaluated by considering finite size microsections which allow localized deformation to develop. This paper shows that the crushing stress can also be established from the local response of the fully periodic column of cells through an energy argument leading to a Maxwell-type construction. [DOI: 10.1115/1.2047608]

1 Introduction

Solid foams comprise a class of lightweight cellular materials with exceptional energy absorption characteristics. As a result, they are widely used in shock mitigation, in packaging, and in cushioning. Most cellular materials exhibit a compressive stress-displacement ($\sigma-\delta$) response with the characteristic three-regime shape shown in Fig. 1 [1]. In regime I the response is stiff and essentially linearly elastic (modulus E^*). This terminates into a limit stress (σ_I) which is followed by an extended stress plateau which comprises regime II (mean stress level $\bar{\sigma}_P$ and extent $\Delta\bar{\varepsilon}_P$). In regime III the response recovers stiffness once more. The limit stress indicates the onset of unstable behavior. During the stress plateau, localized deformation bands of buckled cells form and spread throughout the material. The relatively low initial stress peak and the extended plateau are defining features of the excellent energy absorption of such materials. This three regime behavior is indeed very similar to the one observed in in-plane crushing of honeycombs [2–4] where the plateau was clearly associated with the propagation of a crushing front through the material.

Gong and Kyriakides [5,6] used experiment and analysis to study a class of polyester urethane foams which exhibit the behavior shown in Fig. 1. They considered foams with nominal cell sizes ranging from 3 to 100 pores per inch (ppi). Their relative densities (ρ^*/ρ) varied between approximately 0.022 and 0.028. The experimental part of the study included (a) characterization of the cell and ligament morphology; (b) measurement of the mechanical properties of the polymer using ligaments extracted from the foam; and (c) crushing of blocks of foams between rigid parallel plates, at various constant displacement rates. The major findings were as follows:

- The microstructure consisted of interconnected polyhedra each with an average of 13.7 faces, each face having an average of nearly five sides.
- The size of the cells did not vary significantly.
- The cells were elongated in the rise direction (ratio of diameters in rise-to-lateral directions—λ—ranged from 1.23 to 1.43).
- The ligaments had a three-cusp hypocycloid cross section (*Plateau borders*) and the cross-sectional area varied along the length.
- To first order, the microstructure was found to scale with cell size.
- The base material is an elastomer and, as a result, it is viscoelastic. For simplicity the rate dependence was neglected and the material was assumed to be linearly elastic with modulus E and Poisson's ratio ν.
- Because of the anisotropy in the microstructure, the compressive response shown in Fig. 1 is characteristic of the rise direction. The response in the transverse direction has a much smaller elastic modulus but is essentially a monotone. Overall buckling of the test specimens was not observed in this direction.

A sequence of models for predicting all mechanical foam properties of interest was developed. The foam was idealized to be periodic using the space-filling Kelvin cell, shown in Fig. 2, assigned the major geometric characteristics of the actual foams. The cells were elongated in the rise direction ($\lambda=\tan\alpha$); the ligaments were assumed to be straight, to have *Plateau border* cross sections and variable cross-sectional area distribution as shown in Fig. 3 (see [7,8]). The following area distribution function developed from measurements was adopted:

$$A(\xi)=A_o f(\xi)=A_o(a\xi^4+b\xi^2+1), \quad -0.5\leqslant\xi=x/\ell\leqslant 0.5 \tag{1}$$

where ℓ is the length of the ligament (unless otherwise stated $a=86$ and $b=1$ will be used). The ligaments were modeled as shear deformable beams. The sectional properties of the beams are

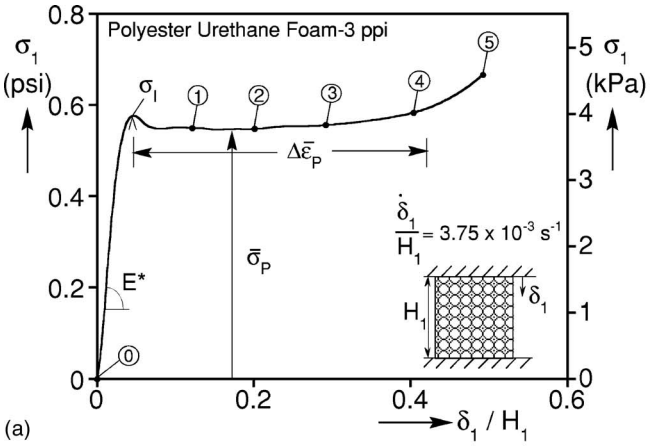
$$A=\left(\sqrt{3}-\frac{\pi}{2}\right)r^2, \quad I_y=I_z=\frac{1}{24}(20\sqrt{3}-11\pi)r^4, \quad J=0.0021r^4 \tag{2}$$

(GJ is an equivalent torsional rigidity which accounts for warping [9]).

A Timoshenko-type correction for the additional deformation resulting from shear stresses was adopted. The strain energy due to the shear force (V) is modified as follows:

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received December 16, 2004; final manuscript received April 25, 2005. Review conducted by R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.



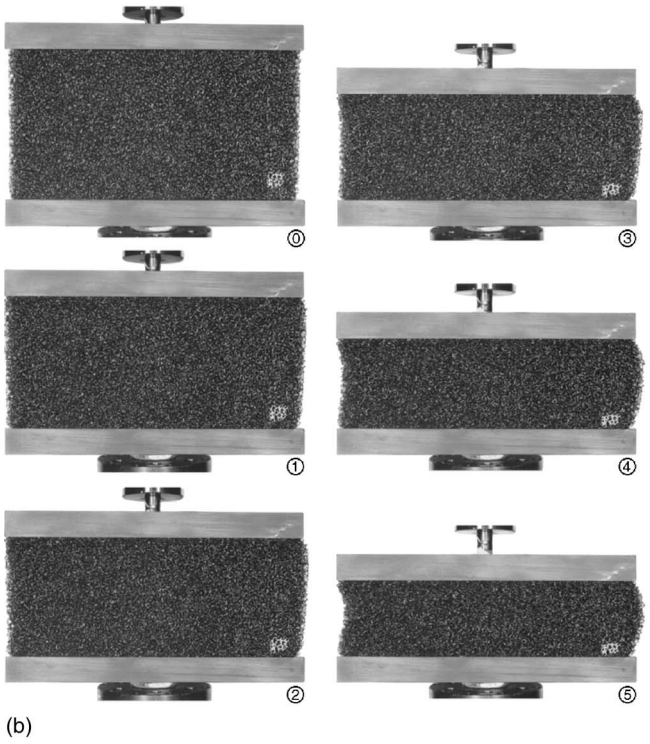


Fig. 1 (*a*) Compressive stress-displacement response of polyester-urethane foam. (*b*) Sequence of deformed configurations of crushed foam.

$$U_s = \int_{-1/2}^{1/2} \beta \frac{V^2(\xi)}{2GA} \ell d\xi, \quad \text{where } \beta = \frac{A}{I_y^2} \int_z \frac{Q^2(z)}{b(z)} dz. \qquad (3)$$

The second integral is over the cross section of the ligament shown in Fig. 3 with Q being the first moment of area about the y axis [$b(z) \equiv$ width]. For this cross section $\beta = 1.24$.

The initial elastic constants of the anisotropic foam were evaluated in closed form [5] using microsections appropriate for each type of loading. The rise and transverse direction moduli predicted for the five foams tested were found to be in good agreement with the measured values. The rest of the properties were calculated numerically using domains based on the characteristic cell shown in Fig. 4.

Uniaxial compression of a domain consisting of one or several characteristic cells is accomplished by prescribing incrementally the relative displacement (δ_1) between its upper and lower periodic boundaries. Figure 5(*a*) shows the calculated average stress (σ_1)-displacement response for a generic anisotropic foam with

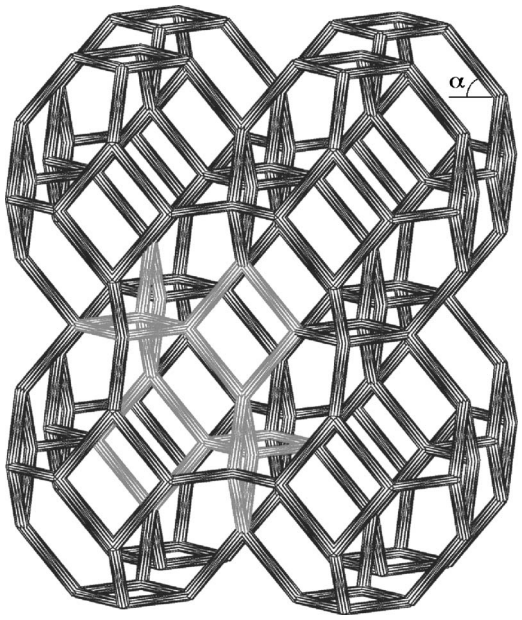


Fig. 2 Cluster of anisotropic Kelvin cells

$\lambda = 1.3$ (=tanα, see Fig. 2) and relative density of 0.025. The foam is compressed in the rise direction. We start by considering one fully periodic unit cell. Initially (OAB), it deforms uniformly and symmetrically about a vertical axis through the center of the cells (inset shows side view of deformed characteristic cell). The response is stiff but progressively becomes nonlinear because of the increasingly larger deflections of the ligaments. At some point along OAB an eigenvalue is identified indicating that an alternate equilibrium configuration is possible corresponding to a mode at the cell level. Similar calculations and eigenvalue checks are conducted for columns of N characteristic cells with periodicity conditions at the top, bottom, and sides. For the present microstructure $N=2$ yields the same first eigenvalue and mode, but for $N=3$ the eigenvalue is smaller. Increasing the number of cells to 4, 6, 8, and 12 reduces further the critical stress and increases the mode wavelength. For $N>12$ the critical stress remains unchanged although the mode wavelength increases. Similar calculations were performed for wider columns in the x_2, x_3 directions. It was found that widening the column does not affect the critical stress or the corresponding mode. In summary, the critical mode

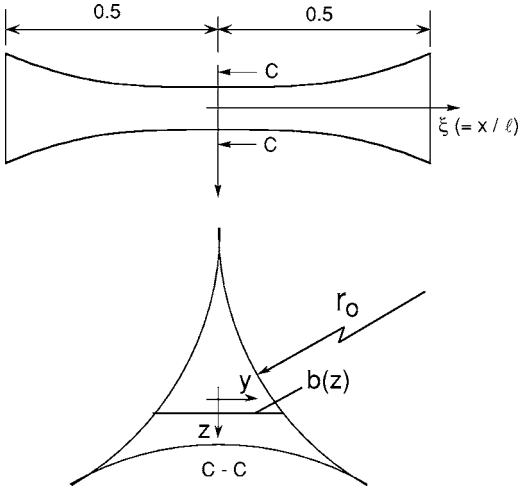


Fig. 3 Geometry of foam ligaments

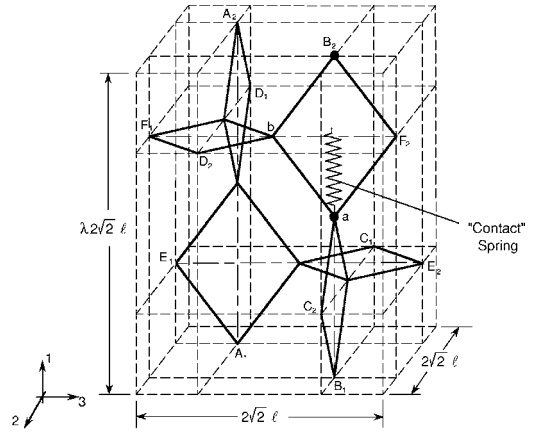


Fig. 4 The Kelvin foam characteristic cell

involves the whole domain (often called a long wavelength mode). A more systematic method for identifying the critical state based on the Bloch wave theory of Triantafyllidis and co-workers [10–12] can be found in Gong et al. [13] (see also [14]).

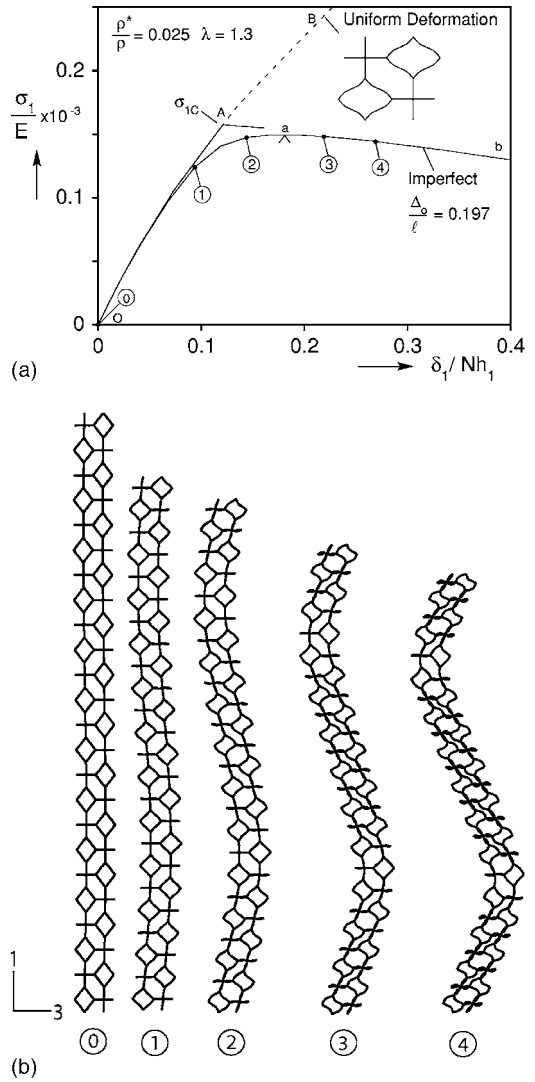


Fig. 5 (*a*) Calculated prebuckling and postbuckling responses in rise direction. (*b*) Deformed configurations of periodic column of cells.

The postbuckling response of the periodic column of cells was established by adding a small initial geometric imperfection to the domain corresponding to the critical mode ($N=12$ is sele cted here). The microsection is then compressed in the x_1 direction by prescribing incrementally δ_1. Results for the case (oab) with imperfection amplitude of $\Delta_o/\ell=0.197$ are included in Fig. 5(*a*) [$\Delta_o=|\mathbf{u}_o(\mathbf{x})|_\infty$ and $\mathbf{u}_o(\mathbf{x})$ is the imperfection]. The initial and several deformed configurations of a column of cells in the 1-3 plane corresponding to points marked with solid bullets on the response are shown in Fig. 5(*b*). Initially the response follows that of the perfect case. As σ_{1C} is approached, the domain is distorted (see configurations ① and ②) resulting in reduction of its stiffness. With further distortion the response increasingly deviates from that of the perfect case and reaches a limit stress at an average strain of about 18%. Subsequently, the stress decreases with deformation. The deformed configurations in Fig. 5(*b*) show the progressive growth of the buckled configuration.

By contrast, the critical stress of a characteristic domain compressed in the transverse direction is associated with a mode local to the cell. Furthermore, the postbuckling response was found to be monotonically increasing, as was the case in the experiments [6]. The present work is concerned with the stress plateau that follows the limit stress as well as its extent. This behavior was observed strictly for compression in the rise direction of the foams and, as a result, attention will be limited to this type of loading.

2 Modeling of Foam Crushing

The periodicity conditions for a single characteristic Kelvin unit cell (Fig. 4) can be expressed as follows: Let the three pairs of opposite bounding faces of the cell be $(\partial R_{i1},\partial R_{i2})$, $i=1,2,3$. The displacements and rotations of points on these faces are respectively denoted by (u_{i1},u_{i2}) and $(\theta_{i1},\theta_{i2})$, $i=1,2,3$. The following relationships of degrees of freedom are prescribed for points on each pair of faces $(\partial R_{i1},\partial R_{i2})$, $i=1,2,3$:

$$u_{i1}-u_{i2}=u_{i1}^{ref}-u_{i2}^{ref},\quad i=1,2,3$$
$$\theta_{i1}-\theta_{i2}=0,\quad i=1,2,3 \tag{4}$$

where u_{ij}^{ref} $(j=1,2)$ are the displacements of corresponding points on opposite sides chosen as reference points [e.g., (A_1,A_2), (C_1,C_2), etc.].

The characteristic cell is discretized with finite elements within the nonlinear code ABAQUS using the B32, three-noded, quadratic, space-beam elements. Each ligament is represented by eight elements of uniform cross-sectional area. The area of each element is based on the symmetric function $f(\xi)$ in (1) as follows:

$$f(\xi)=1,\quad 0\leqslant|\xi|\leqslant 0.2,$$
$$f(\xi)=1.482,\quad 0.2<|\xi|\leqslant 0.3,$$
$$f(\xi)=2.574\quad 0.3<|\xi|\leqslant 0.4, \tag{5}$$
$$f(\xi)=4.993\quad 0.4<|\xi|\leqslant 0.5.$$

By using the *beam general section* feature, the sectional characteristics of each beam $\{A,I_y,J\}$ are prescribed to correspond to the values in (2). The shear correction factor $\beta=1.24$ is included. The resultant model has 24 ligaments, 192 elements, 378 (951 code) nodes, and a total of 2268 variables. The foam material is assumed to be linearly elastic with Young's modulus E and Poisson's ratio of 0.49.

2.1 Ligament Contact. Experimental observations confirm that the crushing of the foam in the rise direction is associated with local collapse of zones. This of course is reminiscent of very similar behavior observed in our honeycomb crushing studies [2–4]. Experimental observations also confirm that collapse of these zones is terminated by contact between the buckled liga-

ments. This is again in concert with the arrest of collapse in honeycombs. The complexity of the foam microstructure makes modeling of ligament contact formidable. Indeed, despite the simplifications introduced to the modeling effort by the adoption of the periodic Kelvin cell, modeling of ligament contact remains a difficult task which often causes numerical convergence difficulties. We bypassed this difficulty by adopting the following approximation for ligament contact. The corner nodes of all vertical rhombi (squares for $\lambda=1$) are connected with springs. Thus, for example, points a and B_2 in rhombus abB_2F_2 in Fig. 4 are connected. The springs (ABAQUS SPRING2) become activated when the vertical distance between the two nodes exceeds a chosen threshold. Once activated, the spring force depends only on the vertical distance of the connected nodes; in other words, relative displacement of the nodes in the x_2-x_3 plane is uninhibited. Thus, the force in the spring is given by

$$F=\begin{cases} 0, & (u_{1B_2}-u_{1a})<\psi, \\ k(u_{1B_2}-u_{1a}), & (u_{1B_2}-u_{1a})\geqslant\psi, \end{cases} \tag{6}$$

where ψ is the gap that must be closed first before the spring is activated and k is the spring constant. The selection of the variables ψ and k and their effects on the results will be discussed below.

2.2 Crushing of a Finite Size Foam Domain. We have seen that the postbuckling response of the strictly periodic column of cells analyzed in Fig. 5 exhibits a limit stress. This indicates that localized deformation patterns will probably be energetically preferred if the periodicity conditions are relaxed to some degree. To this end we consider a domain of the same foam with a finite width. For numerically expediency we choose a domain with 12 characteristic cells in the rise direction (x_1) and 28 in the transverse direction (x_3) [see undeformed configuration, ⊙, in Fig. 6(*b*)]. The domain is one characteristic cell thick (x_2 direction) with periodicity conditions applied to the front and back surfaces, as well as at the top and bottom edges. The two lateral edges are however free. An initial imperfection (amplitude $\Delta_o=0.197\ell$) corresponding to the critical mode of the fully periodic microstructure of this height is introduced to the model. In addition, the pseudo-contact scheme described above with the parameters $\psi=0.670\lambda\ell$ and $k=0.202Er_o^2/\ell$ was implemented.

Figure 6(*a*) shows the calculated average stress-displacement response and Fig. 6(*b*) shows the undeformed and a sequence of deformed configurations corresponding to the numbered points on the response in Fig. 6(*a*). Initially the response follows the trivial one closely. At higher stress levels the imperfection is excited (see configuration ①) and nontrivial deformation grows causing a softening in the response. This results in a stress maximum close to configuration ①. As might be expected, freeing the lateral ends makes the limit stress about 6% lower than the corresponding value for the fully periodic domain with the same imperfection value. Beyond the stress maximum localized deformation starts to become visible in the concave parts of the free ends in configurations ① and ②. At these sites deformation grows to levels that cause the contact springs to be activated. Further local deformation is then inhibited and collapse spreads across to neighboring cells while the stress drops somewhat. By configuration ③, highly deformed zones and less deformed ones are distinctly visible in the domain while simultaneously overall buckling of the domain takes place. As this continues contact springs are activated in increasingly more cells resulting in the upturn in the average stress seen to occur around an average strain of about 45%. By configuration ⑤ most cells have developed contact. For the pseudo-contact adopted the strain at the upturn depends on ψ. This dependence was discussed in [6] and will be discussed again in the next section. The response is not sensitive to the spring constant k provided its value exceeds $0.202Er_o^2/\ell$. In summary, making the domain finite by freeing the sides coupled with the pseudo-contact adopted has reproduced all three response regimes observed experimentally.

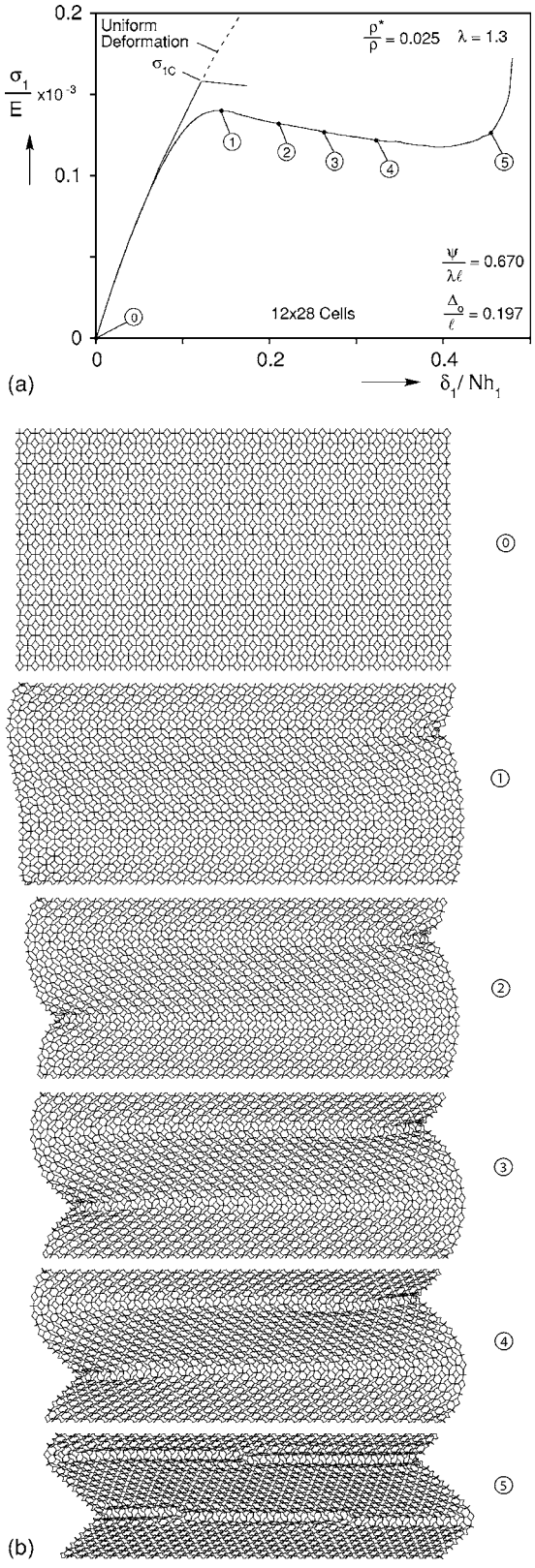


Fig. 6 (*a*) Stress-displacement responses of a finite size foam microsection. (*b*) Sequence of deformed configurations corresponding to response in (*a*).

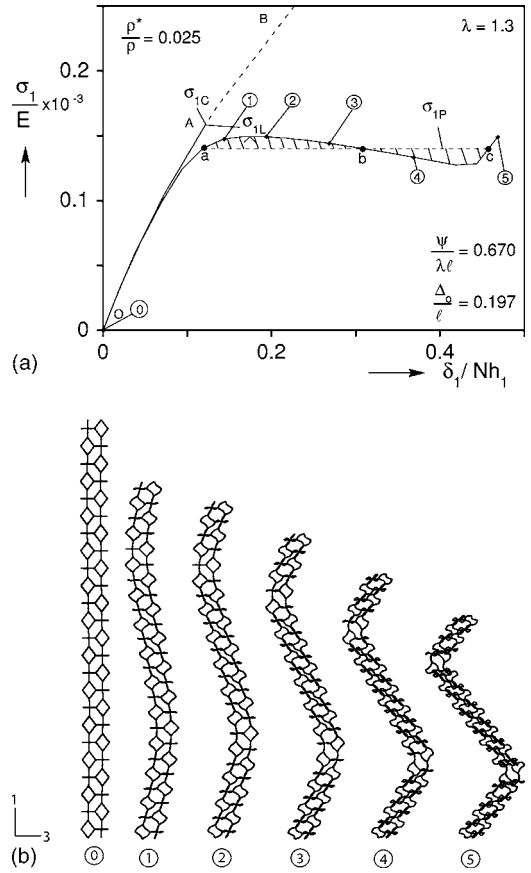


Fig. 7 (*a*) Response of a fully periodic column of cells and the Maxwell construction. (*b*) Sequence of deformed configurations corresponding to response in (*a*).

2.3 Energy Based Prediction of Foam Crushing Stress. It was noted that the response yielded by the fully periodic domain of a single column of characteristic cells does not allow the development of localized deformation nor the stress plateau associated with its spreading. Despite this, the characteristic domain response is worth revisiting. To this end we incorporate into this model the pseudo-contact model discussed above and pursue the solution further into the postbuckling regime. In this case we consider a fully periodic domain consisting of a column of 12 characteristic cells. The calculated response is shown in Fig. 7(*a*). The initial part of the response is the same as the one in Fig. 5(*a*). The domain buckles in the long wavelength mode shown in Fig. 7(*b*). A limit load instability is developed at an average strain of about 18% and subsequently the stress drops with increasing displacement. The response stabilizes again at an average strain of 42.4% when ligaments come into contact in some cells. A second stable branch develops during which most of the cells have collapsed as seen in configuration ⑤.

This *up-down-up* nature of the response of this characteristic domain is one typical of problems which exhibit propagating instabilities (see [15–17]). The first ascending branch up to the limit stress represents stable, uncollapsed configurations. The second ascending branch following the local minimum represents collapsed configurations which in this case is accompanied by the long wavelength buckling of the domain. The two are joined by a third branch which is unstable. For a range of stress levels, cells initially at equilibrium in the first ascending branch, given a large enough disturbance, can snap to the second stable branch. Indeed in this stress regime collapsed and uncollapsed zones of cells can coexist.

This response is unrealizable in practice as it represents a uniformly collapsing foam. It will be referred to as the *local response* of the material. However, it can be used to evaluate the plateau or crushing stress as follows. Consider a loaded foam domain in which the highly deformed phase has been initiated and coexists with the relatively undeformed phase. Apply a far field displacement so that the highly deformed phase consumes, in a quasistatic fashion, an additional unit volume of the domain. In other words the unit volume of material moves from equilibrium state a to state c at a constant stress σ_{1P}. If A is the cross-sectional area of the domain, the external work done in this process is

$$\sigma_{1P}A(\delta_c - \delta_a).$$

The material is elastic and, as a result, the process is path independent. The collapsed configuration c can be reached by moving along abc at a constant stress due to the propagation of the collapse front as discussed above, or by following the up-down-up response. Thus, the change in strain energy induced can be evaluated from the local response. The change in strain energy between configuration a and c is equal to the external work done yielding

$$\sigma_{1P}A(\delta_c - \delta_a) = \int_{\delta_a}^{\delta_c} \sigma_1(\delta)Ad\delta.$$

The plateau stress is then

$$\sigma_{1P} = \frac{1}{(\delta_c - \delta_a)} \int_{\delta_a}^{\delta_c} \sigma_1(\delta)d\delta. \tag{7}$$

This equation is satisfied when abc is at a level that makes the shaded area under the response and above the line equal to the shaded area below the line and above the response. σ_{1P} is also called *Maxwell* stress.

For the case shown in Fig. 7(*a*) $\sigma_{1P}/E = 0.140 \times 10^{-3}$. The corresponding finite domain case in Fig. 6(*a*) yielded a value of about 0.128×10^{-3}. As is the case for the limit stress, the free ends tend to lower slightly the plateau stress. It is also interesting to compare the deformed configurations from the two cases at the end of the process. Figure 8(*a*) shows the deformed configuration at an average strain of about 45% for the finite domain calculation. To accommodate such a comparison we assembled together 28 fully periodic column deformed configurations at the same average strain to generate the domain shown in Fig. 8(*b*). Although not exactly the same, the similarity between the two deformed patterns is quite clear. This is especially true regarding the distributions of collapsed cells. Small differences are once more due to the influence of the free edges. This influence should be diminishing as the width of the finite domain is increased. The strength of the Maxwell construction is the simplicity with which the crush-

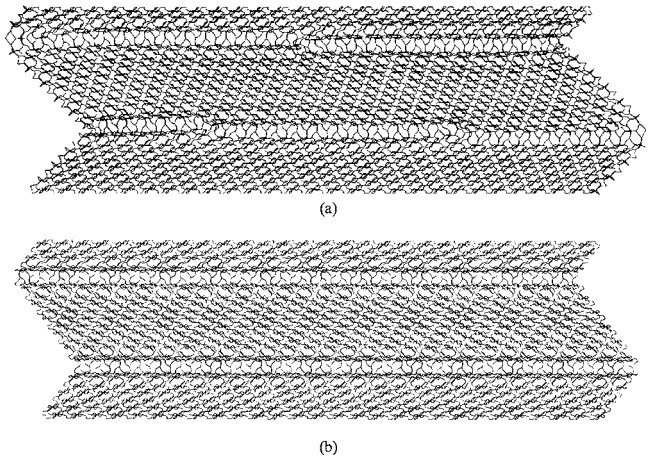


Fig. 8 Comparison of crushed configurations from the finite width domain (*a*) and the fully periodic domain (*b*)

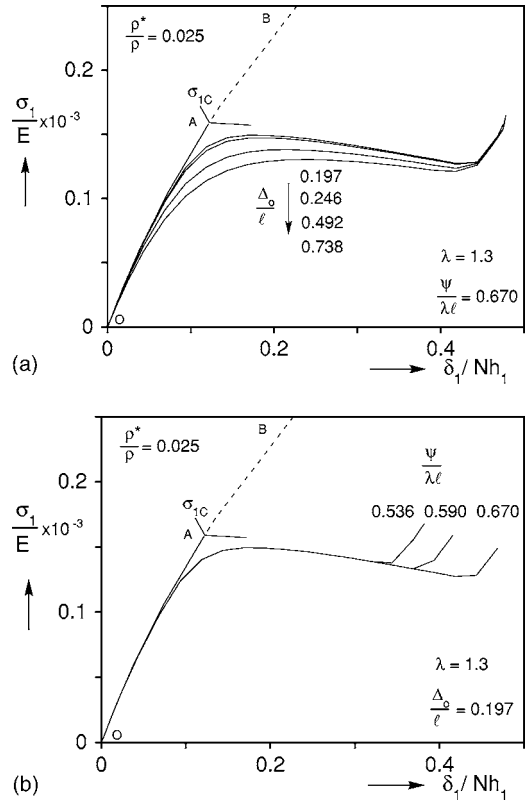


Fig. 9 (*a*) Fully periodic domain responses for various imperfection amplitudes. (*b*) Fully periodic domain responses for various contact parameter ψ values.

ing strength can be established. By comparison, the finite domain calculation is computationally very intensive indeed.

3 Parametric Study of Crushing Stress

The postbuckling response and, by extension, the crushing stress are influenced first by the amplitude of the imperfection (Δ_o) adopted and second by the value of the gap variable ψ. The influence of the imperfection amplitude is illustrated in Fig. 9(*a*) which shows responses for four imperfection amplitudes for the same generic foam analyzed in the previous section. (In all cases in this section the domain consists of a fully periodic column of 12 cells.) Increasing the imperfection reduces the limit stress and as a result lowers the crushing stress. Numerical values of these stresses are given in Table 1. The difference in σ_{1P} between the smallest and largest imperfections used is about 9%. For this idealized Kelvin cell microstructure the accuracy of σ_{1P} will increase as the imperfection amplitude is decreased. However, as pointed out in [6], the imperfection amplitude is one way of making the calculated critical stresses approach those measured in actual foams which have irregular microstructures. Thus, the optimal choice of Δ_o must be left to the user.

The effect of the gap variable ψ on such results is illustrated in Fig. 9(*b*) where similar responses are shown for three different values of ψ. This variable does not affect the limit stress but affects the point of recovery of the response. The quantitative effect this has on σ_{1P} is given in Table 2. The difference in σ_{1P} between the smallest and largest values of ψ used is about 3.5%.

Gong et al. [5] developed a volume correction scheme in order to account for the duplication of material at nodes joining four beam ligaments. In all cases Eq. (1) was adopted for $A(\xi)$ and the extra material was removed by special surface functions. This was done for a range of r_o/ℓ and for several anisotropy values. The results were fitted with the following powerlaw relationship for the relative density

$$\frac{\rho^*}{\rho} = k\left(\frac{r_o}{\ell}\right)^n \tag{8}$$

where k and n depend on the anisotropy of the foam in the manner given in Table 3. Because of this nonlinear relationship r_o/ℓ will be used in parametric studies instead of ρ^*/ρ as is more customary.

To this end, foams of different r_o/ℓ were analyzed keeping the anisotropy, the imperfection, and the gap variables fixed. The results are listed in Table 4. (Note that the density of all these foams had been calculated in the course of developing (8). For this reason, the relative densities listed in Table 4 are slightly different from the values yielded by the fit in (8).] The bifurcation stress, the limit stress, and the crushing stress are listed in Table 4. As expected, σ_C is the largest of the three and σ_P the smallest. The three stresses are also plotted in log-log scales against r_o/ℓ in Fig. 10. All three follow powerlaw relationships to r_o/ℓ.

Table 1 Comparison of calculated critical stresses for foams with four initial imperfection amplitudes ($\lambda=1.3$, $\rho^*/\rho=2.5\%$, $\psi/\lambda\ell=0.670$)

$\frac{\Delta_o}{\ell}$	$\frac{\sigma_{1C}}{E}\times 10^3$	$\frac{\sigma_{1L}}{E}\times 10^3$	$\frac{\sigma_{1P}}{E}\times 10^3$
0.197	0.156	0.149	0.140
0.246	0.156	0.147	0.139
0.492	0.156	0.138	0.133
0.738	0.156	0.130	0.127

Table 2 Comparison of calculated critical stresses for foams with three gap values ($\lambda=1.3$, $\rho^*/\rho=2.5\%$, $\Delta_o/\ell=0.197$)

$\frac{\psi}{\lambda\ell}$	$\frac{\sigma_{1C}}{E}\times 10^3$	$\frac{\sigma_{1L}}{E}\times 10^3$	$\frac{\sigma_{1P}}{E}\times 10^3$
0.536	0.156	0.149	0.145
0.590	0.156	0.149	0.143
0.670	0.156	0.149	0.140

Table 3 Fit parameters for corrected relative density powerlaw relationship to r_o/ℓ for variable XSC

λ	n	k
1	1.7392	0.1803
1.1	1.7426	0.1791
1.2	1.7433	0.1637
1.3	1.7449	0.1580
1.4	1.7474	0.1350

Table 4 Comparison of calculated critical stresses for foams of several relative densities ($\lambda=1.3$, $\Delta_o/\ell=0.246$, $\psi/\lambda\ell=0.670$)

$\frac{r_o}{\ell}$	$\frac{\rho^*}{\rho}(\%)$	$\frac{\sigma_{1C}}{E}\times 10^3$	$\frac{\sigma_{1L}}{E}\times 10^3$	$\frac{\sigma_{1P}}{E}\times 10^3$
0.2185	1.11	0.025	0.0236	0.022
0.3485	2.50	0.156	0.147	0.139
0.4720	4.14	0.506	0.472	0.451
0.5352	5.08	0.810	0.767	0.738
0.5917	5.96	1.170	1.130	1.086

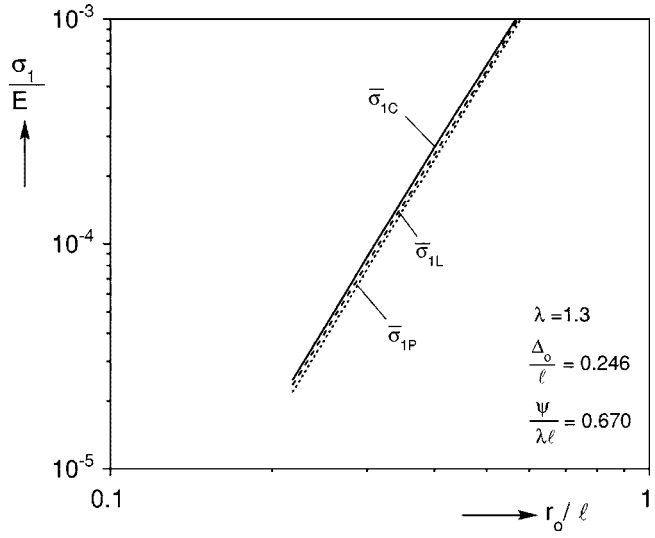


Fig. 10 Critical stress, limit stress, and crushing stress as a function of the foam geometric parameter r_o/ℓ

The next variable to be varied is the foam anisotropy λ, keeping the relative density, the imperfection amplitude, and the gap variable $(\psi/\lambda\ell)$ constant. Foams with five different values of λ between 1.0 and 1.4 were analyzed. Three of the calculated responses are shown in Fig. 11 while the main results from the five cases are summarized in Table 5. As reported in [5], increasing the anisotropy increases the foam modulus E_1^* in the x_1 direction. Thus, for example, when λ=1.4 the modulus is 69% higher than the isotropic case modulus. This is because the ligaments with a vertical projection have a smaller inclination to the vertical axis. Simultaneously, the length of these ligaments has increased (see Fig. 4) and, as a result, the critical buckling stress of the foam decreases but by a smaller amount. By contrast, the critical strain decreases quite significantly. In all cases the postbuckling response exhibits a limit load instability. The limit load also decreases as λ is increased essentially for the same reason. The crushing stress decreases with λ more or less in the same mild manner as the other two stresses.

4 Summary and Conclusions

The compressive response and crushing of a class of anisotropic foams has been examined. Compressing the foam along the rise direction leads to localized buckling and collapse of the foam cells. Local collapse is terminated by contact between cell ligaments. In the process neighboring cells hitherto intact start to collapse and in this manner the crushing spreads throughout the material. The spreading of collapse occurs at a well-defined load plateau and continues until most of the cells are thus affected and the material response regains stiffness once more.

Different aspects of this three regime response have been established by idealizing the foam to consist of periodic Kelvin cells. Anisotropy is introduced by elongating the cells in the rise direction. The onset of instability was previously established by considering a fully periodic column of cells tall enough to accommodate long wavelength modes found to govern the behavior. The crushing response has been evaluated by considering finite size microsections which allow localized deformation to develop. In particular a domain with a width of 28 cell columns having free sides was adopted in the examples shown in this paper. The present extension of this work showed that the crushing stress can also be established from the local response of the fully periodic column of cells. Inclusion of cell ligament contact results in an up-down-up response. This response results in crushed configurations similar to those produced via the computationally more elaborate finite width model. In view of this similarity in the final configurations, an energy balance-type argument results in the Maxwell stress which is then associated with the crushing stress.

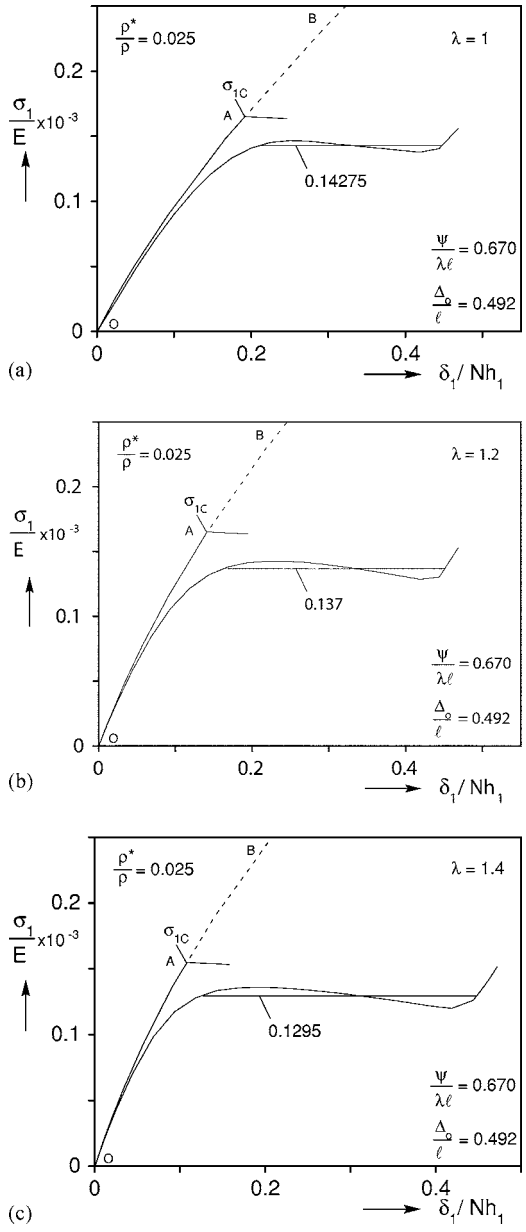


Fig. 11 Fully periodic domain stress-displacement responses for foams of three different anisotropy values

Table 5 Comparison of calculated critical stresses for foams of different anisotropy (ρ^*/ρ=2.5%, Δ_o/ℓ=0.492, $\psi/\lambda\ell$=0.670)

λ	$\frac{r_o}{\ell}$	$\frac{E_1^*}{E}\times10^3$	$\frac{\sigma_{1C}}{E}\times10^3$	$\frac{\sigma_{1L}}{E}\times10^3$	$\frac{\sigma_{1P}}{E}\times10^3$
1	0.3222	1.115	0.166	0.147	0.143
1.1	0.3324	1.315	0.165	0.146	0.141
1.2	0.3411	1.512	0.162	0.143	0.137
1.3	0.3485	1.701	0.156	0.138	0.133
1.4	0.3549	1.881	0.152	0.136	0.130

Acknowledgment

The work was conducted with financial support from the National Science Foundation through Grant No. CMS-0245485. This support is acknowledged with thanks. The authors also wish to thank Foamex International Inc. for providing foam samples and technical support.

References

[1] Gibson, L. J., and Ashby, M. F., 1997, *Cellular Solids: Structure and Properties*, 2nd Ed. Cambridge University Press, Cambridge, UK.
[2] Papka, S. D., and Kyriakides, S., 1994, "In-Plane Compressive Response and Crushing of Honeycomb," J. Mech. Phys. Solids, **42**, pp. 1499–1532.
[3] Papka, S. D., and Kyriakides, S., 1998, "In-Plane Crushing of a Polycarbonate Honeycomb," Int. J. Solids Struct., **35**, pp. 239–267.
[4] Papka, S. D., and Kyriakides, S., 1998, "Experiments and Full-Scale Numerical Simulations of In-Plane Crushing of a Honeycomb," Acta Mater., **46**, pp. 2765–2776.
[5] Gong, L., Kyriakides, S., and Jang, W.-Y., 2005, "Compressive Response of Open-Cell Foams. Part I: Morphology and Elastic Properties," Int. J. Solids Struct., **42**(5-6), pp. 1355–1379.
[6] Gong, L., and Kyriakides, S., 2005, "Compressive Response of Open-Cell Foams. Part II: Initiation and Spreading of Crushing," Int. J. Solids Struct., **42**(5-6), pp. 1381–1399.
[7] Weaire, D., and Hutzler, S., 1999, *The Physics of Foams*, Oxford University Press, Oxford, UK.
[8] Kraynik, A. M., 2003, "Foam Structure: From Soap Froth to Solid Foams," MRS Bull., **24**(4), pp. 275–278.
[9] Warren, W. E., Nielsen, M. K., and Kraynik, A. M., 1997, "Torsional Rigidity of a Plateau Border," Mech. Res. Commun., **24**, pp. 667–672.
[10] Schraad, M. W., and Triantafyllidis, N., 1997, "Scale Effects in Media With Periodic and Nearly Periodic Microstructures. Part I: Macroscopic Properties," ASME J. Appl. Mech., **64**, pp. 751–762.
[11] Schraad, M. W., and Triantafyllidis, N., 1997, "Scale Effects in Media With Periodic and Nearly Periodic Microstructures. Part II: Failure Mechanisms," ASME J. Appl. Mech., **64**, pp. 763–771.
[12] Triantafyllidis, N., and Schraad, M. W., 1998, "Onset of Failure in Aluminum Honeycombs Under General In-Plane Loading," J. Mech. Phys. Solids, **46**, pp. 1089–1124.
[13] Gong, L., Kyriakides, S., and Triantafyllidis, N., 2004, "On the Stability of Kelvin Cell Foams Under Compressive Loads," J. Mech. Phys. Solids, **53**, pp. 771–794.
[14] Laroussi, M., Sab, K., and Alaoui, A., 2002, "Foam Mechanics: Nonlinear Response of an Elastic 3D-Periodic Microstructure," Int. J. Solids Struct., **39**, pp. 3599–3623.
[15] Ericksen, J. L., 1975, "Equilibrium of Bars," J. Elast., **5**, pp. 191–201.
[16] Kyriakides, S., 1993, "Propagating Instabilities in Structures," *Advances in Applied Mechanics*, Vol. 30, J. W. Hutchinson and T. Y. Wu, eds., Academic Press, Boston, MA, pp. 67–189.
[17] Kyriakides, S., 2001, "Propagating Instabilities in Materials," in *Materials Science for the 21th Century*, Society of Materials Science, Japan, Vol. 1, pp. 316–325.

An Anisotropic Hyperelastic Constitutive Model With Fiber-Matrix Shear Interaction for the Human Annulus Fibrosus

X. Q. Peng

Z. Y. Guo

B. Moran
e-mail: b-moran@northwestern.edu

Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208

Based on fiber reinforced continuum mechanics theory, an anisotropic hyperelastic constitutive model for the human annulus fibrosus is developed. A strain energy function representing the anisotropic elastic material behavior of the annulus fibrosus is additively decomposed into three parts nominally representing the energy contributions from the matrix, fiber and fiber-matrix shear interaction, respectively. Taking advantage of the laminated structure of the annulus fibrosus with one family of aligned fibers in each lamella, interlamellar fiber-fiber interaction is eliminated, which greatly simplifies the constitutive model. A simple geometric description for the shearing between the fiber and the matrix is developed and this quantity is used in the representation of the fiber-matrix shear interaction energy. Intralamellar fiber-fiber interaction is also encompassed by this interaction term. Experimental data from the literature are used to obtain the material parameters in the constitutive model and to provide model validation. Determination of the material parameters is greatly facilitated by the partition of the strain energy function into matrix, fiber and fiber-matrix shear interaction terms. A straightforward procedure for computation of the material parameters from simple experimental tests is proposed. [DOI: 10.1115/1.2069987]

1 Introduction

The annulus fibrosus of the human intervertebral disk can be described as a highly organized fiber reinforced material. The collagen fibers of the annulus form lamellae and encircle the gelatinous nucleus pulposus of the intervertebral disk, as shown in Fig. 1. The fibers are attached to the superior and inferior vertebral endplates and run a roughly uniform course in each lamina. The fibers in adjoining lamellae cross each other at an angle of approximately 120 deg [1–3]. The annulus fibrosus is an essential constituent of the intervertebral disk. Its major mechanical functions are to distribute loads and provide motion and flexibility to the spine. Extreme stresses from excessive physical activities may cause fiber disorganization and rupture and, hence, result in a degenerated intervertebral disk, which is considered to be the major etiologic factor of low back pain and spinal instability [4]. To elucidate the mechanisms in the normal and pathologic functions of the intervertebral disk, the load deformation properties of the annulus fibrosus must be understood. This has stimulated extensive research efforts on the material behavior of the annulus fibrosus.

Experimental investigations of the human annulus fibrosus have been widely carried out in the past decades [5–17]. Galante [5] investigated the tensile properties of the human lumbar annulus fibrosus under various testing environments. Wu and Yao [6] demonstrated that the annulus fibrosus was nearly incompressible. Adams et al. [7,8] explored size effects on the tensile properties of the annulus fibrosus. They found that fiber-matrix interaction made a significant contribution to the annulus tensile stiffness and strength. Best et al. [9] reported the radial compressive modulus for the annulus fibrosus and studied the relations between biochemical composition, structure, and function of intervertebral disks. Skaggs et al. [10] investigated the tensile properties of the lumbar annulus fibrosus with single lamella samples. They concluded that regional variations in the tensile properties of the annulus fibrosus resulted predominantly from structural (site-dependent) rather than from biochemical composition variations. Acaroglu et al. [11] explored the effects of degeneration and aging on the tensile behavior of the human lumbar annulus fibrosus. Ebara et al. [12] reported the tensile properties of the nondegenerated annulus fibrosus at different anatomical regions. Fujita et al. [13] investigated the radial tensile properties of the annulus fibrosus. Dynamic and static shear moduli of the annulus fibrosus were obtained by Iatridis et al. [14]. Fujita et al. [15] investigated the anisotripic shear behavior of the annulus fibrosus by relaxed simple shear deformation. The contribution of annular collagen fibers to the shear modulus was studied by applying a prestretch to the fibers perpendicular to the direction of the applied shear stress. They further showed that the shear behavior of the annulus fibrosus significantly increased with tensile strain in the fibers. Bass et al. [16] performed uniaxial and biaxial tensile tests on the annulus fibrosus. Holzapfel et al. [17] performed an in vitro study of single lamellar annulus lamellae to investigate the tensile properties of the annulus fibrosus and the regional variation of fiber orientation. These experimental studies demonstrate that the intrinsic mechanical properties of the annulus fibrosus are nonlinear, anisotropic, and are dependent on anatomical regions and degeneration.

The above experimental studies shed some light on the intrinsic material behavior of the annulus fibrosus. They are also vital for the development of theoretical constitutive models by providing original data and means of validation for constitutive models. Several constitutive models have been proposed for the annulus fibrosus based on Spencer's theoretical approach for fiber reinforced composites [18,19]. Wu and Yao [6] presented a hyperelastic constitutive model for the annulus fibrosus by defining a strain energy function with two new invariants which combined the conven-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received December 13, 2004; final manuscript received May 16, 2005. Review conducted by S. Kyriakides. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

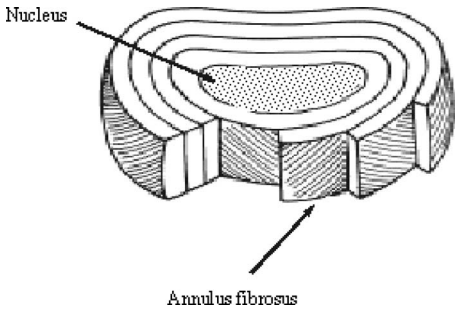


Fig. 1 Schematic of intervertebral disk showing the laminated structure of the annulus fibrosus

tional invariants in the fiber reinforced theory for two families of fibers. Interlamellar fiber-fiber interaction was included in their constitutive model which complicated the material characterization. By using simultaneous regression analysis on uniaxial experimental data, Klisch and Lotz [20] proposed a nonlinear, orthotropic constitutive model for the annulus fibrosus. They assigned all principal stretches in the uniaxial tensile tests. This will most likely violate the traction free boundary conditions. By decomposing the strain energy contribution into two parts namely ground substance and embedded fibers, Holzapfel et al. [21,22] proposed an anisotropic hyperelastic fiber reinforced constitutive model for soft tissue. They then performed a finite element analysis on a multisegment portion of the human lumber spine using their constitutive model for the annulus fibrosus [23]. Elliott and Setton [24,25] developed a linear fiber-induced anisotropic model for the annulus fibrosus by using a quadratic strain energy function involving interlamellar fiber-fiber interaction. An infinitesimal strain tensor was used in the model development. They first carried out a series of uniaxial tensile tests on the human annulus fibrosus to obtain the engineering constants. The material parameters in their linear model were then determined from these engineering constants. This constitutive model is valid in the small strain regime but does not extend readily to the large strain regime encountered in the physiological response of the annulus fibrosus. Taking into account the effect of nutrient and electrolyte transport and swelling deformation, Sun and Leong [26] developed a mixture model for the annulus fibrosus. The solid phase (ground matrix and collagen fibers) was modeled as a hyperelastic composite reinforced with one family of fibers by defining a local coordinate system for each annulus lamella. The elimination of interlamellar fiber-fiber interaction simplified the constitutive model. Wagner and Lotz [27] proposed a strain energy function with separate terms representing the matrix, fiber, and interaction energies to characterize the nonlinear elastic behavior of the annulus fibrosus.

In this paper, we develop a constitutive model for the annulus fibrosus which is relatively simple yet which captures the material behavior over the physiological range of response. First, similar to the approach of Sun and Leong [26], a continuum formulation based on Spencer's approach [19] is used with a single family of aligned fibers in each lamina. This eliminates the interlamellar fiber-fiber interaction and simplifies the strain energy function. Second, by explicitly dividing the strain energy function into three parts representing energy stored in the matrix, fibers and in fiber-matrix shear interaction, the determination of the material parameters involved in the constitutive model is greatly facilitated. A simple geometric description for shearing between the fiber and matrix is provided and this quantity is used in the representation of the fiber-matrix shear interaction energy.

An outline of the remainder of this paper is given as follows: In Sec. 2, a general anisotropic hyperelastic constitutive model for composites with one family of fibers is developed based on Spencer's continuum mechanics framework for fiber reinforced composites. A strain energy function with explicit definition of matrix, fiber, and fiber-matrix shear interaction energy is introduced to represent the nonlinear anisotropic material behavior of soft tissues with one family of fibers. A simple geometric description for the shearing between the fiber and the matrix is derived and this quantity is used to represent the fiber-matrix shear interaction. An alternative representation of the fiber shear angle used in Ref. [27] is shown to correspond to the present one for the special case of simple shear but to differ in the case of uniaxial deformation. A straightforward procedure for computing the material parameters from simple experimental tests is proposed. The method avoids the simultaneous determination of material parameters through least-square regression, which can sometimes give unrealistic values for the parameters. The specific application to the annulus fibrosus is given in Sec. 3 through a simple homogenization process for two families of fibers. Experimental data on the human anterior outer (AO) annulus fibrosus are used in Sec. 4 to obtain the material parameters in the constitutive model and to provide model validation. In Sec. 5, a detailed discussion of the fiber-matrix shear interaction term is presented. The model is then used to simulate biaxial tensile tests on the annulus fibrosus. A brief summary and conclusion are provided in Sec. 6.

2 Fiber Reinforced Constitutive Model With Fiber-Matrix Interaction

Soft tissue such as the annulus fibrosus of the human intervertebral disk and collateral ligaments can be considered as laminated composite structures with one family of aligned fibers embedded in a ground matrix material in each lamina. In the development of the proposed constitutive model, the fibers are assumed to be extensible and uniformly distributed in the ground substance. The fibers are assumed to be perfectly bonded to the matrix and no separation between the fiber and matrix is considered. The matrix is assumed to be initially isotropic and hyperelastic with or without compressibility. Based on Spencer's continuum mechanics theory for fiber-reinforced composites [19], we first develop a nonlinear anisotropic constitutive model with fiber-matrix shear interaction for soft tissue. Extension of this model to the lamellar structure of the annulus fibrosus is given in Sec. 3.

2.1 General Format for the Constitutive Model. It is assumed that the strain energy function W is a scalar function of the right Cauchy-Green deformation tensor $\mathbf{C}=\mathbf{F}^T\mathbf{F}$ and the original fiber directional vector $\mathbf{a}_0$ [19], i.e., $W=W(\mathbf{C},\mathbf{a}_0)$. Here $\mathbf{F}$ is the deformation gradient tensor. The elastic response of the annulus fibrosus (for a single layer) is assumed to originate from the resistance of the matrix, fibers, and their interaction. Accordingly, the strain energy function W can be divided into three parts, namely

$$W = W(\mathbf{C},\mathbf{a}_0) = W^M + W^F + W^{FM} \tag{1}$$

where nominally W^M is the strain energy contribution from the ground substance, W^F is the contribution from the fiber stretch and W^{FM} is the strain energy caused by the fiber-matrix shear interaction. This approach was previously applied by Quapp and Weiss [28] and Gardiner and Weiss [29] in the material characterization of human collateral ligaments. However, they did not consider fiber-matrix shear interaction in their studies due to the predominantly tensile loading of ligaments, whereas, the annulus fibrosus undergoes multiaxial and shear loadings and the shear interaction contribution may be significant.

The strain energy function may be written in terms of principal invariants I_i as

$$W(\mathbf{C},\mathbf{a}_0) = W(I_1,I_2,I_3,I_4,I_5) \tag{2}$$

where the invariants are given by

$$I_1 = \operatorname{tr}\mathbf{C}, \quad I_2 = \tfrac{1}{2}[(\operatorname{tr}\mathbf{C})^2 - \operatorname{tr}\mathbf{C}^2], \quad I_3 = \det\mathbf{C},$$

$$I_4 = \mathbf{a}_0\cdot\mathbf{C}\cdot\mathbf{a}_0 = \lambda_F^2, \quad I_5 = \mathbf{a}_0\cdot\mathbf{C}^2\cdot\mathbf{a}_0 \tag{3}$$

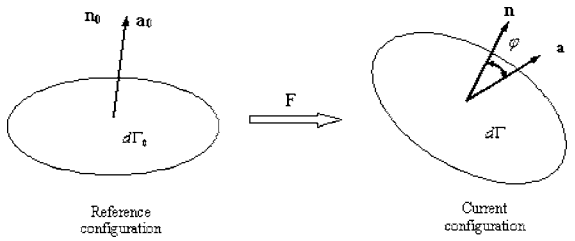


Fig. 2 Exact geometric description of the fiber-matrix interaction

and where λ_F is the fiber stretch.

The ground substance in soft tissue is usually modeled as a hyperelastic material. Here, for simplicity, we use the well-developed Neo-Hookean hyperelastic model with compressibility. Other forms such as exponential [30] could also be used, the particular form depending on the experimental data to be fitted. The energy W^M stored in the matrix is therefore given by the compressible Neo-Hookean form

$$W^M = C_{10}(\bar{I}_1 - 3) + \frac{1}{D_1}(J-1)^2 \tag{4}$$

where $J=I_3^{1/2}$ is the total volume change and $\bar{I}_1$ is the first deviatoric invariant

$$\bar{I}_1 = J^{-2/3} I_1 \tag{5}$$

For an incompressible matrix, only one material parameter, C_{10}, is needed.

The fiber part of the strain energy function originates from the fiber elongation. We define a simple nonlinear fiber strain energy function W^F based on the fiber stretch as

$$W^F = \begin{cases} C_2(I_4-1)^2 + C_3(I_4-1)^4 & I_4 > 1 \\ 0 & I_4 \leqslant 1 \end{cases} \tag{6}$$

Here we assume that the fibers do not contribute to the total strain energy when they are under contraction ($\lambda_F < 1$), since the fibers are generally crimped and buckle readily and thus have negligible compression stiffness. The material parameters C_{10}, C_2, and C_3 have units of MPa and D_1 has units of MPa^{-1}, and are all positive quantities. Now the task is to define the fiber-matrix shear interaction part of the strain energy function.

2.2 Fiber-Matrix Interaction. Because the fibers and matrix are assumed to be perfectly bonded together, the strain energy contribution from the fiber-matrix interaction is assumed to originate from the shearing between the fiber and the matrix or shearing between fibers within the same lamina resulted from collagen cross linking [27]. Next we provide a simple geometric description for the shearing between the fiber and matrix and use this quantity with representation of the strain energy contribution from fiber-matrix shear interaction.

For a fiber with an original fiber orientation vector $\mathbf{a}_0$, we select a differential area element $d\Gamma_0$ which has a normal $\mathbf{a}_0$, i.e., the reference normal $\mathbf{n}_0=\mathbf{a}_0$, as shown in Fig. 2. After a deformation, the normal $\mathbf{n}$ of the deformed differential area element, $d\Gamma$, can be calculated by using Nanson's relation [31]

$$\mathbf{n}d\Gamma = J\mathbf{n}_0 \cdot \mathbf{F}^{-1} d\Gamma_0 \tag{7}$$

Hence, we have

$$\mathbf{n} = \frac{1}{|\mathbf{n}_0 \cdot \mathbf{F}^{-1}|}\mathbf{n}_0 \cdot \mathbf{F}^{-1} = \frac{1}{|\mathbf{a}_0 \cdot \mathbf{F}^{-1}|}\mathbf{a}_0 \cdot \mathbf{F}^{-1} \tag{8}$$

where $||$ denotes the norm of a vector and

$$|\mathbf{a}_0 \cdot \mathbf{F}^{-1}| = \sqrt{\mathbf{a}_0 \cdot \mathbf{F}^{-1} \cdot \mathbf{F}^{-T} \cdot \mathbf{a}_0} = \sqrt{\mathbf{a}_0 \cdot \mathbf{C}^{-1} \cdot \mathbf{a}_0} \tag{9}$$

The deformed fiber direction vector $\mathbf{a}$ can be calculated from the deformation gradient tensor $\mathbf{F}$ as

$$\mathbf{a} = \frac{1}{\lambda_F}\mathbf{F} \cdot \mathbf{a}_0 \tag{10}$$

The shearing between the matrix and the fiber can be represented by the angle φ in Fig. 2

$$\cos\varphi = \mathbf{n} \cdot \mathbf{a} = \frac{1}{|\mathbf{a}_0 \cdot \mathbf{F}^{-1}|}\frac{1}{\lambda_F}\mathbf{a}_0 \cdot \mathbf{F}^{-1} \cdot \mathbf{F} \cdot \mathbf{a}_0 = \sqrt{\frac{1}{(\mathbf{a}_0 \cdot \mathbf{C}^{-1} \cdot \mathbf{a}_0)}\frac{1}{I_4}} \tag{11}$$

From the Cayley-Hamilton theorem [32], a second-order tensor $\mathbf{C}$ satisfies its own characteristic equation

$$\mathbf{C}^3 - I_1\mathbf{C}^2 + I_2\mathbf{C} - I_3\mathbf{I} = 0 \tag{12}$$

where $\mathbf{I}$ is the 2nd order identity tensor. Multiplying Eq. (12) by $\mathbf{C}^{-1}$ and then solving for $\mathbf{C}^{-1}$ gives

$$\mathbf{C}^{-1} = \frac{1}{I_3}(\mathbf{C}^2 - I_1\mathbf{C} + I_2\mathbf{I}) \tag{13}$$

Substituting this expression for $\mathbf{C}^{-1}$ into Eq. (11) yields

$$\cos\varphi = \mathbf{n} \cdot \mathbf{a} = \sqrt{\frac{I_3}{(I_5 - I_1 I_4 + I_2)I_4}} \tag{14}$$

From Eq. (14), we have

$$\tan^2\varphi = \frac{1}{\cos^2\varphi} - 1 = \frac{I_4}{I_3}(I_5 - I_1 I_4 + I_2) - 1 \tag{15}$$

By defining a new invariant $\chi=\tan^2\varphi$, we propose the following representation for the fiber-matrix shear interaction energy

$$W^{FM} = W^{FM}(I_4, \phi) = f(I_4)\chi^2 = f(I_4)\left[\frac{I_4}{I_3}(I_5 - I_1 I_4 + I_2) - 1\right]^2 \tag{16}$$

Equation (16) indicates that the fiber-matrix shear interaction may depend on fiber stretch status. This was verified by experimental studies [15] and our analysis based on composite mechanics [33]. If the fiber is under contraction or has very small stretch, the fibers may easily bend and rotate, and there may be little fiber-matrix shear interaction. Consequently, the fiber-matrix shear interaction factor $f(I_4)$ should assume a small value for this case. With more stretching, the fibers become stiffer, and thus provide more resistance to fiber rotation (or they activate cross linking in tension). Hence, $f(I_4)$ should increase with the fiber stretch. However, there should have an upper limit for the fiber-matrix shear interaction factor $f(I_4)$, considering the idealized case of rigid fibers. Based on this characteristic behavior of $f(I_4)$, we represent the fiber-matrix interaction factor by a sigmoid function as

$$f(I_4) = \frac{\gamma}{1 + \exp[-\beta(\lambda_F - \lambda_F^*)]} \tag{17}$$

where γ corresponds to the upper limit for $f(I_4)$ and $f(I_4)=0.5\gamma$ at $\lambda_F=\lambda_F^*$. The quantity λ_F^* may be related to the transition point between the toe region and the linear region in uniaxial tensile stress-strain curves. In Eq. (17), the material parameter γ has units of MPa and β is dimensionless.

We further investigate the fiber-matrix shear interaction by considering special cases of simple shear and uniaxial tensile deformation. Similar studies can be found in Merodio and Ogden [34]. For a simple shear deformation shown in Fig. 3, the deformation gradient tensor $\mathbf{F}$ and the right Cauchy-Green deformation tensor $\mathbf{C}$ are

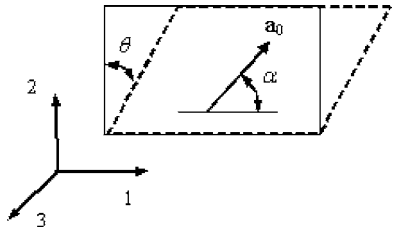


Fig. 3 Simple shear deformation of fiber reinforced composites

$$\mathbf{F}=\begin{bmatrix}1 & \kappa & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\quad \mathbf{C}=\mathbf{F}^T\mathbf{F}=\begin{bmatrix}1 & \kappa & 0\\ \kappa & 1+\kappa^2 & 0\\ 0 & 0 & 1\end{bmatrix} \tag{18}$$

where $\kappa=\tan\theta$. The invariants I_1, I_2, and I_3 are given by

$$I_1=3+\kappa^2,\quad I_2=3+\kappa^2,\quad I_3=1 \tag{19}$$

The invariants I_4 and I_5 for fibers with an original direction vector of $\mathbf{a}_0=[\cos\alpha\ \ \sin\alpha\ \ 0]$ shown in Fig. 3 are

$$I_4=\lambda_F^2=1+\kappa^2\sin^2\alpha+\kappa\sin 2\alpha \tag{20}$$

$$I_5=\kappa^2+\cos^2\alpha+(1+\kappa^2)^2\sin^2\alpha+(2\kappa+\kappa^3)\sin 2\alpha \tag{21}$$

Substituting Eqs. (19)–(21) into Eq. (15) yields

$$\chi=\tan^2\varphi=\frac{I_4}{I_3}(I_5-I_1I_4+I_2-1)$$
$$=\kappa^2\cos^2 2\alpha+\kappa^4\sin^2\alpha\cos^2\alpha+\kappa^3\sin 2\alpha\cos 2\alpha=I_5-I_4^2 \tag{22}$$

Hence, for the simple shear deformation, the shear angle between the fiber and matrix can be expressed in terms of the invariants I_4 and I_5 by Eq. (22) and correspondingly the fiber-matrix shear interaction energy function Eq. (16) could be rewritten in the simple form

$$W^{FM}=f(I_4)(I_5-I_4^2)^2 \tag{23}$$

For a uniaxial tensile deformation in which the principal axes are shown in Fig. 4, the deformation gradient tensor $\mathbf{F}$ and the right Cauchy-Green deformation tensors $\mathbf{C}$ are

$$\mathbf{F}=\begin{bmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{bmatrix}\quad \mathbf{C}=\mathbf{F}^T\mathbf{F}=\begin{bmatrix}\lambda_1^2 & 0 & 0\\ 0 & \lambda_2^2 & 0\\ 0 & 0 & \lambda_3^2\end{bmatrix} \tag{24}$$

where λ_i is the principal stretch along the i-th principal direction. The strain invariants, I_1, I_2, and I_3 can be expressed in terms of the principal stretches as

$$I_1=\lambda_1^2+\lambda_2^2+\lambda_3^2,\quad I_2=\lambda_1^2\lambda_2^2+\lambda_2^2\lambda_3^2+\lambda_1^2\lambda_3^2,\quad I_3=\lambda_1^2\lambda_2^2\lambda_3^2 \tag{25}$$

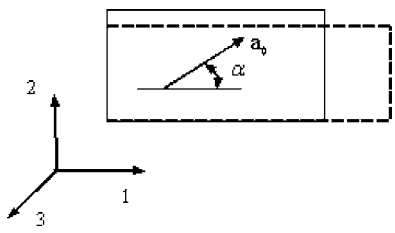


Fig. 4 Uniaxial tensile deformation of fiber reinforced composites

The invariants I_4 and I_5 for fibers in the 1–2 plane with an original direction vector of $\mathbf{a}_0=[\cos\alpha\ \ \sin\alpha\ \ 0]$ shown in Fig. 4 are

$$I_4=\lambda_1^2\cos^2\alpha+\lambda_2^2\sin^2\alpha \tag{26}$$

$$I_5=\lambda_1^4\cos^2\alpha+\lambda_2^4\sin^2\alpha \tag{27}$$

Substituting Eqs. (25)–(27) into Eq. (15) yields

$$\chi=\tan^2\varphi=\frac{I_4}{I_3}(I_5-I_1I_4+I_2)-1$$
$$=\frac{(\lambda_1^2-\lambda_2^2)^2}{\lambda_1^2\lambda_2^2}\sin^2\alpha\cos^2\alpha=\frac{1}{\lambda_1^2\lambda_2^2}(I_5-I_4^2) \tag{28}$$

As can be seen from Eq. (28), for uniaxial tensile deformations with fibers in the 1–2 plane, the relationship can be expressed as $\tan^2\varphi=(I_5-I_4^2)$ if and only if $\lambda_1\lambda_2=1$. However, for an incompressible material ($I_3=\lambda_1^2\lambda_2^2\lambda_3^2=1$), which is roughly the case for the annulus fibrosus, and for which $\lambda_3=1$ (this is closely satisfied in uniaxial tensile tests on the annulus fibrosus), then $\lambda_1\lambda_2=1$. Wagner and Lotz [27] used the expression $(I_5-I_4^2)$ for the shear strain along the fiber direction to represent the intralamellar fiber-fiber interaction, which they assumed is the dominant interaction term.

Summing the matrix, fiber, and fiber-matrix shear interaction contributions represented by Eqs. (4), (6), and (16), the total strain energy function for our fiber-reinforced model is given by

$$W=W^M+W^F+W^{FM}=C_{10}(\bar{I}_1-3)+\frac{1}{D_1}(J-1)^2+C_2(I_4-1)^2$$
$$+C_3(I_4-1)^2+f(I_4)\chi^2 \tag{29}$$

As can be seen from Eq. (29), we have, at this point, a total of seven material parameters (only six are needed if the matrix is taken to be incompressible) and the physical meaning of each term is very clear. This makes the constitutive model very simple and easy to implement. These material parameters can be obtained from experimental data through the following procedure:

(i.) Match experimental results of uniaxial tensile tests transverse to the fiber direction to obtain the matrix material parameters C_{10} and D_1;
(ii.) with C_{10} and D_1 obtained, match experimental results of uniaxial tensile tests along the fiber direction to obtain the fiber material parameters C_2 and C_3; and
(iii.) with C_{10}, D_1, C_2, and C_3 obtained, match experimental results of simple shear tests or uniaxial tensile tests skew to the fiber direction to obtain the parameters γ, β and λ_f^* in the fiber-matrix shear interaction factor $f(I_4)$.

This procedure provides a guideline for experimental setup for the material characterization of human soft tissue. Furthermore, it facilitates the determination of the material parameters by avoiding the determination of material parameters through least-squares regression on all the parameters *simultaneously*, which is sometimes difficult to accomplish.

The second Piola-Kirchhoff stress tensor is obtained directly from the hyperelastic potential energy equation (29) as $\mathbf{S}=2\,\partial W/\partial\mathbf{C}$ and the Cauchy stress tensor, $\boldsymbol{\sigma}$, is given by $\boldsymbol{\sigma}=J^{-1}\mathbf{F}\mathbf{S}\mathbf{F}^T$ from which we obtain the expression

$$\boldsymbol{\sigma}=\frac{2}{J}\mathbf{F}\frac{\partial W}{\partial\mathbf{C}}\mathbf{F}^T$$
$$=\frac{2}{J}[(I_2W_2+I_3W_3)\mathbf{I}+W_1\mathbf{B}-I_3W_2\mathbf{B}^{-1}+I_4W_4\mathbf{a}\otimes\mathbf{a}$$
$$+I_4W_5(\mathbf{a}\otimes\mathbf{B}\mathbf{a}+\mathbf{a}\mathbf{B}\otimes\mathbf{a}] \tag{30}$$

where $\mathbf{I}$ is the 2nd order unit tensor, $\mathbf{B}$ is the left Cauchy-Green

tensor, $\mathbf{B}=\mathbf{F}\mathbf{F}^T$, and W_i denotes $\partial W/\partial I_i$. When the fibers are under contraction ($\lambda_F<1$), the terms involving C_2 and C_3 in Eqs. (29) and (30) are eliminated due to the fact that the fibers have negligible compressive stiffness.

3 Specific Application to the Human Annulus Fibrosus

The human annulus fibrosus is a complex composite tissue with a series of concentric encircling layered lamellae which surround the gelatinous nucleus pulposus of the intervertebral disk [3]. The collagen fibers run a roughly uniform course in each layer and cross each other at an angle of approximately 120 deg in adjoining lamina, as illustrated in Fig. 1. These anatomical features make it feasible to treat the annulus fibrosus as a laminated composite with one family of fibers in each individual ply.

The annulus fibrosus usually consists of 15–25 distinct layers [35]. As a result, it is computationally costly to model each layer in the annulus fibrosus by using the formulation developed in Sec. 2 for soft tissues with one family of reinforced fibers. By assuming neighboring lamellae are perfectly bonded together, we can homogenize the annulus fibrosus as a composite with two families of reinforced fibers. With the physically realistic assumption of no interlamellar fiber-fiber interaction, we can still apply the previously developed fiber reinforced model to the annulus fibrosus. Consequently, the strain energy function for the annulus fibrosus with the homogenization of the two families of fibers a and b can be written as

$$\begin{aligned} W &= W^M + \frac{1}{2}(W_a^F + W_b^F) + \frac{1}{2}(W_a^{FM} + W_b^{FM}) \\ &= C_{10}(\bar{I}-3) + \frac{1}{D}(J-1)^2 + \frac{1}{2}\{C_2(I_{4a}-1)^2 \\ &\quad + C_3(I_{4a}-1)^4 + C_2(I_{4b}-1)^2 + C_3(I_{4b}-1)^4\} + \frac{1}{2}\{f(I_{4a})\chi_a^2 \\ &\quad + f(I_{4b})\chi_b^2\} \end{aligned} \tag{31}$$

Here we assume that the two families of fibers are approximately mechanically equal except for their directions, which was experimentally verified by Skaggs et al. [10]. Due to the homogenization of two neighboring plies of annulus fibrosus into one ply, the strain energy contributions from the fiber and the fiber-matrix shear interaction from each family of fibers need to be halved in Eq. (31). From Eq. (30), the Cauchy stress tensor can be written as

$$\begin{aligned} \sigma = \frac{2}{J}\Bigg\{ &(I_2W_2 + I_3W_3)\mathbf{I} + W_1\mathbf{B} - I_3W_2\mathbf{B}^{-1} + \frac{1}{2}(I_{4a}W_{4a}\mathbf{a}\otimes\mathbf{a} \\ &+ I_{4b}W_{4b}\mathbf{b}\otimes\mathbf{b}) + \frac{1}{2}[I_{4a}W_{5a}(\mathbf{a}\otimes\mathbf{B}\mathbf{a} + \mathbf{a}\mathbf{b}\otimes\mathbf{a}) + I_{4b}W_{5b}(\mathbf{b} \\ &\otimes\mathbf{B}\mathbf{b} + \mathbf{b}\mathbf{B}\otimes\mathbf{b})]\Bigg\} \end{aligned} \tag{32}$$

where $\mathbf{b}$ is the deformed fiber direction vector for the second family of fibers.

To be consistent with experimental uniaxial tensile test setups, the x_1 and x_2 axes are chosen to be in the fiber plane and bisect the angle between the two families of fibers. The x_3 axis is normal to the x_1–x_2 plane, as shown in Fig. 5. The angle between the two families of fibers is initially at 2α deg. The two original fiber direction vectors are then given by

$$\mathbf{a}_0 = \begin{bmatrix}\cos\alpha & \sin\alpha & 0\end{bmatrix}, \quad \mathbf{b}_0 = \begin{bmatrix}\cos\alpha & -\sin\alpha & 0\end{bmatrix} \tag{33}$$

For uniaxial tensile deformation, the deformation gradient tensor $\mathbf{F}$ and the right Cauchy-Green tensor $\mathbf{C}$ were given by Eq. (24). The left Cauchy-Green tensor $\mathbf{B}$ is identical to the right Cauchy-Green tensor $\mathbf{C}$

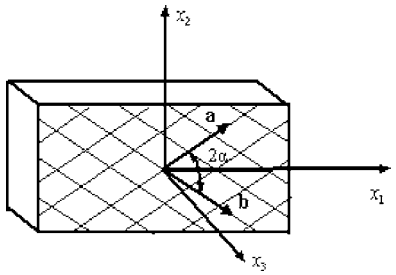


Fig. 5 Annulus fibrosus with two families of fibers

$$\mathbf{B} = \mathbf{C} = \begin{bmatrix}\lambda_1^2 & 0 & 0\\ 0 & \lambda_2^2 & 0\\ 0 & 0 & \lambda_3^2\end{bmatrix} \tag{34}$$

The fiber vectors in the deformed configuration are

$$\mathbf{a}_0 = \frac{1}{\lambda_F}\begin{bmatrix}\lambda_1\cos\alpha & \lambda_2\sin\alpha & 0\end{bmatrix},$$

$$\mathbf{b}_0 = \frac{1}{\lambda_F}\begin{bmatrix}\lambda_1\cos\alpha & -\lambda_2\sin\alpha & 0\end{bmatrix} \tag{35}$$

The invariants I_1, I_2, and I_3 in the uniaxial tensile deformation were given by Eq. (25). Fiber a and fiber b have the same invariants I_4 and I_5, therefore $W_{4a}=W_{4b}=W_4(I_4)$, $W_{5a}=W_{5b}=W_5(I_5)$. I_4 and I_5 were given by Eqs. (26) and (27), respectively.

Substituting Eqs. (25)–(27), (34), and (35) into Eq. (33) gives expressions for the Cauchy stresses in uniaxial tensile deformation as

$$\sigma_1 = \frac{2}{J}\left[I_2W_2 + I_3W_3 + W_1\lambda_1^2 - \frac{I_3W_2}{\lambda_1^2} + (W_4 + 2W_5\lambda_1^2)\lambda_1^2\cos^2\alpha\right] \tag{36a}$$

$$\sigma_2 = \frac{2}{J}\left[I_2W_2 + I_3W_3 + W_1\lambda_2^2 - \frac{I_3W_2}{\lambda_2^2} + (W_4 + 2W_5\lambda_2^2)\lambda_2^2\sin^2\alpha\right] \tag{36b}$$

$$\sigma_3 = \frac{2}{J}\left[I_2W_2 + I_3W_3 + W_1\lambda_3^2 - \frac{I_3W_2}{\lambda_3^2}\right] \tag{36c}$$

where W_i denotes $\partial W/\partial I_i$ and can be calculated from Eq. (29).

In uniaxial tensile tests, the traction free boundary conditions on the two directions other than the loading direction (say x_1) need to be satisfied. Imposing the constraints $\sigma_2=0$, $\sigma_3=0$ permits Eqs. (36*a*)–(36*c*) to be solved for $\sigma_1,\lambda_2,\lambda_3$ for a given principal stretch λ_1 (or for λ_1, λ_2, and λ_3, given σ_1). In some papers in the literature, the principal stretches λ_2 and λ_3 are assigned by approximate relations in terms of λ_1 obtained from experimental measurements. This may lead to unrealistic material parameters and is avoided here.

Next we determine the material parameters needed for the annulus fibrosus by using experimental tensile test data available in the literature. Engineering stresses and engineering strains were generally used in experimental studies. To compare with the experimental data, we transform the Cauchy stresses in Eq. (36*a*)–(36*c*) into engineering stresses in the following numerical simulations.

4 Numerical Results and Model Validation

4.1 Matrix Parameters. Experimental studies have shown that the mechanical tensile behavior of the human annulus fibrosus is site and degeneration dependent [10–13]. This dependence is beyond the scope of this paper. For the purpose of illustrating

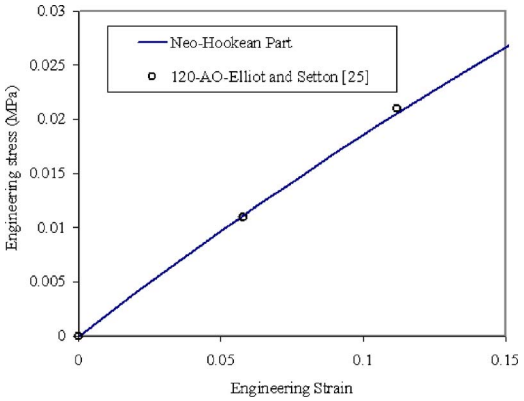


Fig. 6 Initial tensile stress-strain behavior of multilayer AO annulus fibrosus with $2\alpha=120$ deg (used to obtain fiber material parameters C_{10} and D_1)

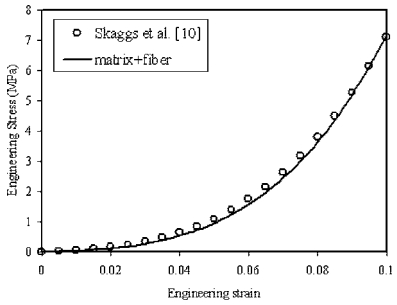


Fig. 7 Tensile stress-strain behavior along fiber direction of single-layer AO annulus fibrosus (used to obtain fiber material parameters C_2 and C_3)

the present constitutive model, we will choose experimental data for the nondegenetated lumbar AO annulus fibrosus, which has been extensively investigated (Refs. [6,10,12,25]). The seven material parameters C_{10}, D, C_2, C_3, and γ, β and λ_f^* in the fiber-matrix shear interaction factor $f(I_4)$ in the constitutive model are obtained *step-by-step* (see Sec. 2) by nonlinear least-square regression analyses on experimental uniaxial tensile test data available in the literature. The traction free boundary conditions on the two directions other than the loading direction are imposed as constraints in the nonlinear least-square regression analyses for determining the material parameters in the constitutive model.

Following the general procedure for obtaining the material parameters in the constitutive model provided in Sec. 2, we first determine the two material parameters C_{10} and D in the Neo-Hookean hyperelastic part for the ground matrix substance. Unfortunately, we cannot find any experimental data of uniaxial tensile tests on a single lamella of the annulus fibrosus with the loading direction transverse to the fiber direction. A feasible alternative is to match the initial elastic response in uniaxial tensile tests on multi-layer samples with an original angle of $2\alpha=120$ deg between the two families of fibers (see Fig. 5). In the initial stage of this deformation mode, the fibers should be under contraction (no fiber strain energy contribution) and the shearing angle between each fiber and the matrix should be very small (fiber-matrix shear interaction is negligible). In addition, to model a slight compressibility for the annulus fibrosus, we will predefine a Poisson's ratio ν for the matrix. The material parameter D_1 is then defined in terms of C_{10} and ν as [36]

$$D_1 = \frac{1-2\nu}{2(1+\nu)C_{10}} \tag{37}$$

We investigated the effect of different values of Poisson's ratio ν in the range of $0.4 \leqslant \nu \leqslant 0.49$ and found that the value of ν in this range had negligible influence on the stress-strain response for the loading configurations considered. For convenience, we use the value $\nu=0.49$ through the paper.

Figure 6 shows the experimental uniaxial tensile stress versus strain by circles on multilayer samples with an original inter-fiber angle of $2\alpha=120$ deg [25]. By matching this stress-strain curve under moderate deformations and with $\nu=0.49$ for a slight compressibility, we obtain the matrix material parameter C_{10} as

$$C_{10} = 0.034 \text{ (MPa)}, \tag{38}$$

The matrix material parameter D_1 is then obtained from Eq. (37) as

$$D_1 = 0.197 \text{ (MPa}^{-1}\text{)} \tag{39}$$

The predicted matrix response is shown in Fig. 6 with a solid line. A good agreement with the experimental data under moderate deformation is expected because we obtain the matrix material parameters by fitting the experimental data.

4.2 Fiber Parameters. Skaggs et al. [10] experimentally investigated the tensile behavior along the fiber direction of a single lamella of the annulus fibrosus. Significant tensile property variations of the annulus fibrosus with respect to sites and directions were found in their experiments. The experimental data exhibited considerable scatter. They found that the average tensile stress-strain behavior for this orientation (loading along the fiber direction) could be approximated by a cubic function as

$$\sigma = A\varepsilon + B\varepsilon^3 \tag{40}$$

where σ is the engineering stress and ε is the engineering strain. The parameters A and B can be found in Ref. [10] as $A=5.7$, $B=6532$. The average experimental tensile behavior along the fiber direction of a single lamella annulus fibrosus, given by Eq. (40) is plotted in Fig. 7 with circles.

With the matrix parameters C_{10} and D_1 known (from Eqs. (38) and (39)), we can determine the fiber material parameters C_2 and C_3 by matching the averaged experimental tensile behavior along the fiber direction of a single lamella shown in Fig. 7. The fiber material parameters so obtained are

$$C_2 = 0.45 \text{ (MPa)}, \quad C_3 = 82.6 \text{ (MPa)} \tag{41}$$

The predicted tensile behavior along the fiber direction of a single lamella is shown in Fig. 7 with a solid line.

4.3 Shear Interaction Factor $f(I_4)$. With the matrix and fiber material parameters (C_{10}, D_1, C_2, and C_3) obtained, our final objective in the material characterization is to determine the material parameters γ, β, and λ_f^* in the fiber-matrix shear interaction factor $f(I_4)$. This is achieved by simultaneously matching the tensile behaviors of multilayer AO annulus fibrosus samples with original inter-fiber angles of $2\alpha=60$ deg and $2\alpha=120$ deg, respectively. Figure 8 shows experimental data reported by different research groups for the case of $2\alpha=60$ deg. In Fig. 8, the stars represent the experimental results from Wu and Yao [6]; the blank circles denote the experimental data from Ebara et al. [12]; and the filled circles represent the experimental data from Elliott and Setton [25]. As can be seen from the figure, there are large discrepancies between the experimental results for the AO annulus fibrosus samples reported by different research groups. Possible factors behind these discrepancies are the variations in testing environments and sample preparation as well as anatomical variations in intervertebral discs from different people. The experimental tensile data for multilayer AO annulus fibrosus samples with an original interfiber angle of $2\alpha=120$ deg reported by Elliott and Setton [25] are shown in Fig. 9 by filled circles. Because Elliott and Setton [25] reported experimental data for both cases of $2\alpha=60$ deg and $2\alpha=120$ deg, we will choose their experimental data (the

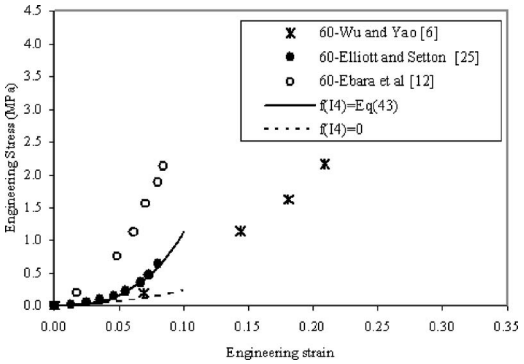


Fig. 8 Tensile stress-strain behavior of multilayer AO annulus fibrosus with $2\alpha=60$ deg (used to obtain fiber-matrix shear interaction factor $f(I_4)$)

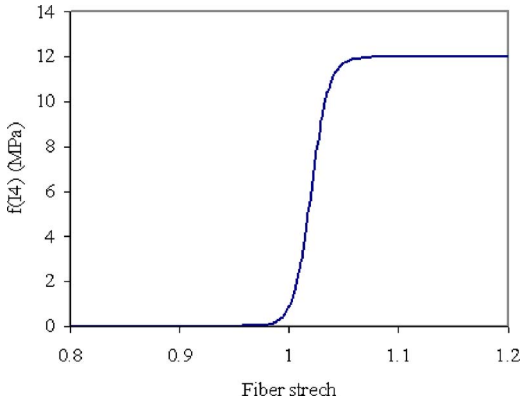


Fig. 10 Fiber-matrix interaction factor $f(I_4)$ expressed by a sigmoid function as Eq. (43)

filled circles in Figs. 8 and 9) to obtain the material parameters γ, β and λ_f^* in the fiber-matrix shear interaction factor $f(I_4)$. With the matrix and fiber material parameters known, the parameters γ, β, and λ_f^* in the fiber-matrix shear interaction factor $f(I_4)$ are determined by simultaneously performing nonlinear least-square regression analysis on the experimental data for multilayer AO annulus fibrosus samples with interfiber angles of $2\alpha=60$ deg and $2\alpha=120$ deg in Elliott and Setton [25]. The traction free boundary conditions are imposed as constraints in the optimization process. The parameters obtained are

$$\gamma = 12.0 \text{ (MPa)}, \quad \beta = 125, \quad \lambda_F^* = 1.02 \tag{42}$$

Hence, the fiber-matrix shear interaction factor $f(I_4)$ is

$$f(I_4) = \frac{12.0}{1 + \exp[-125(\lambda_F - 1.02)]} \tag{43}$$

Figure 10 shows the relation between the fiber-matrix shear interaction factor $f(I_4)$ and fiber stretch according to Eq. (43). As can be seen from Fig. 10, the fiber-matrix interaction factor $f(I_4)$ assumes a very small value when the fiber is under contraction ($\lambda_F < 1$). It increases with the fiber stretch and then approaches an upper limit.

The predicted stress-strain curves for multilayer AO annulus fibrosus samples with interfiber angles of $2\alpha=60$ deg and $2\alpha=120$ deg are shown in Figs. 8 and 9 by solid lines, respectively. Good agreements with the experimental data are expected because we obtain the interaction parameters by nonlinear least-square regression analysis on the experimental data.

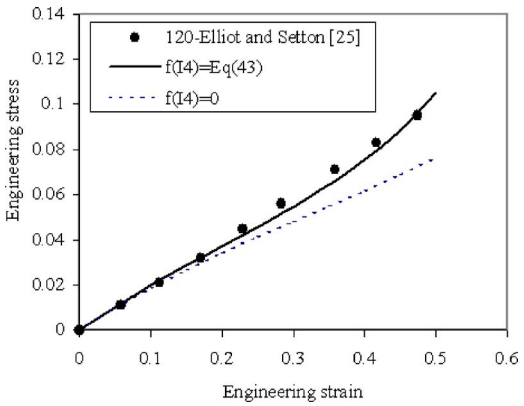


Fig. 9 Tensile stress-strain behavior of multilayer AO annulus fibrosus with $2\alpha=120$ deg (used to obtain fiber-matrix shear interaction factor $f(I_4)$)

4.4 Model Validation. Next we attempt to validate the constitutive model. The general approach for model validation in the literature is to compare the stress-strain curves from the constitutive model with the corresponding experimental data. This kind of comparison alone, however, is not enough for the task of model validation, as will be demonstrated in the following.

By setting the fiber-matrix shear interaction factor $f(I_4)=0$ in the constitutive model and keeping all other matrix and fiber parameters the same, we remove the effect of the fiber-matrix shear interaction. With $f(I_4)=0$, the predicted tensile stress-strain curve for the multilayer AO annulus fibrosus samples with interfiber angles of $2\alpha=60$ deg and $2\alpha=120$ deg are shown in Figs. 8 and 9 by dotted lines, respectively. Taking into account the scatter in the experimental data, we might conclude that a constitutive model without fiber-matrix shear interaction can provide an acceptable stress prediction for the annulus fibrosus. However, the tensile behavior of an annulus fibrosus sample is not solely determined by the stress-strain curve. We should consider other factors such as geometric shape changes of the sample during the deformation. As a step in this direction, we further investigate the constitutive model by comparing the angle changes between the two families of fibers as well as the principal stretches in the x_2 and x_3 directions with experimental data.

Figure 11 shows the angle changes between the two families of fibers as a function of the principal stretches λ_1 in the loading direction (x_1 direction in Fig. 5) for the case of $2\alpha=60$ deg. The results predicted by the constitutive model with fiber-matrix shear interaction are denoted by a solid line in Fig. 11. Those predicted by the constitutive model without fiber-matrix shear interaction are represented by a dotted line in Fig. 11. The only measured experimental data point was from Wu and Yao [6]. As can be seen

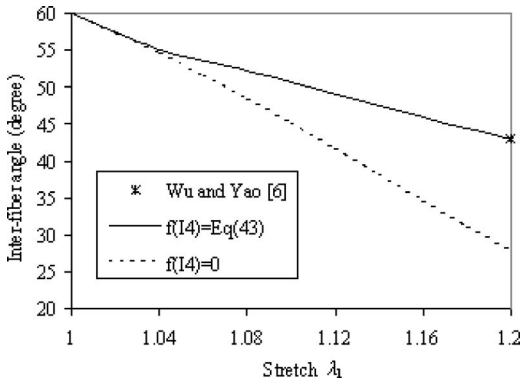


Fig. 11 Angle changes between the two fibers for multilayer AO annulus fibrosus with $2\alpha=60$ deg

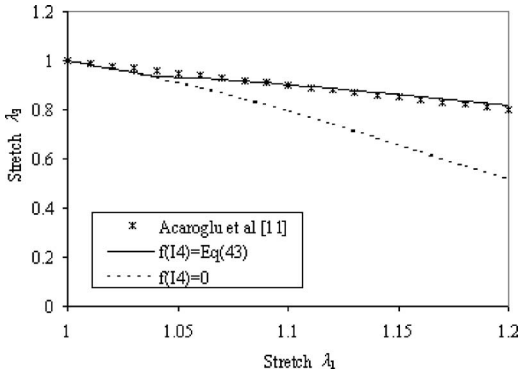


Fig. 12 Stretches in x_2 direction versus stretches in x_1 direction for multilayer AO annulus fibrosus with $2\alpha=60$ deg

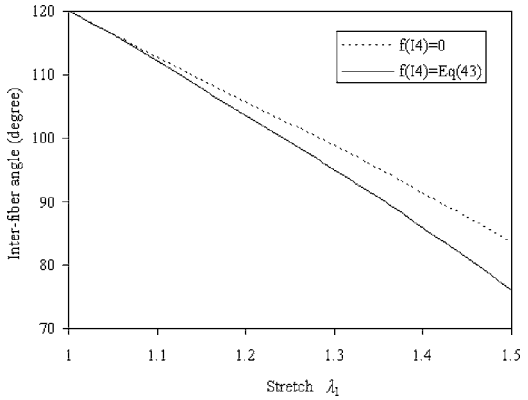


Fig. 14 Angle changes between the two fibers for multilayer AO annulus fibrosus with $2\alpha=120$ deg

from Fig. 11, the constitutive model with fiber-matrix shear interaction provides a good match with the experimental data point. On the contrary, the constitutive model without fiber-matrix shear interaction overpredicts the angle changes between the two families of fibers. This is expected. Without fiber-matrix shear interaction, the fibers can more easily rotate in the ground matrix substance. This excess rotation leads to lower fiber stretch and a less stiff response, as seen in Fig. 8.

Figures 12 and 13 show the numerically predicted principal stretches λ_2 and λ_3 under the traction free boundary conditions ($\sigma_2=\sigma_3=0$) versus the principal stretches λ_1 in the loading direction, for the case of $2\alpha=60$ deg. Again, the principal stretches λ_2 and λ_3 predicted by the constitutive model with the fiber-matrix shear interaction are denoted by solid lines in Figs. 12 and 13. Those predicted by the constitutive model without fiber-matrix interaction are represented by dotted lines in Figs. 12 and 13. Acaroglu et al. [11] reported that the experimental transverse stretches λ_2 can be approximated by the expression $\lambda_2=2-\lambda_1$. This relation is plotted in Fig. 12 by stars. As can be seen from Fig. 12, the curve predicted by the constitutive model with a fiber-matrix shear interaction factor $f(I_4)=$Eq. (43) matches the experimental curve fairly well. On the contrary, the constitutive model without fiber-matrix shear interaction results in excessively large transverse contractions in the x_2 direction. On the other hand, with the nearly incompressible characterization for the annulus fibrosus, which was experimentally demonstrated by Wu and Yao [6], and the relation $\lambda_2=2-\lambda_1$, we should have little change in the x_3 directions, i.e., $\lambda_3\approx 1$. As shown in Fig. 13, the stretch λ_3 predicted by the constitutive model with fiber-matrix shear interaction shows such a tendency. However, that predicted by the constitutive model without fiber-matrix shear interaction deviates from $\lambda_3\approx 1$ significantly. The comparisons in Figs. 11–13 demonstrate that the fiber-matrix shear interaction plays a significant role in the tensile behavior of the annulus fibrous and should not be neglected.

5 Further Investigations of the Constitutive Model

5.1 Discussion on Fiber-Matrix Shear Interaction. By using experimental uniaxial tensile test data from the literature, we obtained values for all of the material parameters in the constitutive model. The material parameters were systematically obtained from tests where their contributions could be expected to be significant. Thus we obtained the matrix parameters C_{10} and D_1 (assuming a Poisson's ratio $\nu=0.49$) from the initial slopes of the stress-strain curve of the tensile tests on multilayer samples with $2\alpha=120°$. We then obtained the fiber parameters C_2 and C_3 from tensile tests along the fiber direction on a single lamella. Finally, the shear interaction term $f(I_4)$ was obtained from the tensile tests on multilayer samples with interfiber angles of $2\alpha=60$ deg and $2\alpha=120$ deg.

Figure 14 shows the numerically predicted angle changes between the two fibers with or without shear interaction for multilayer samples with $2\alpha=120$ deg. In Fig. 14, the solid line represents the results obtained from a fiber-matrix shear interactions factor $f(I_4)=$Eq. (43) and the dotted line denotes those from $f(I_4)=0$. We can see from Fig. 14 that quite different angle changes are predicted by using different constitutive models with or without fiber-matrix shear interaction. Detailed experiments recording interfiber angle changes would be very beneficial for further exploration of the significance of the fiber-matrix shear interaction.

Figures 15 and 16 shows the numerically predicted principal

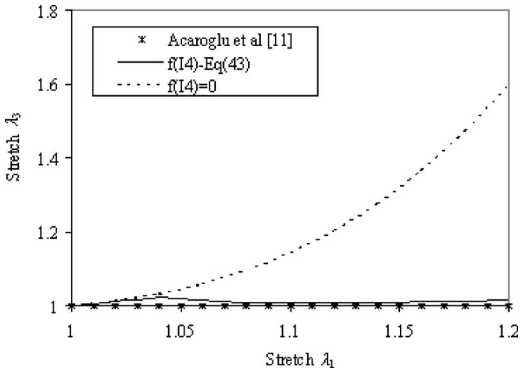


Fig. 13 Stretches in x_3 direction versus stretches in x_1 direction for multilayer AO annulus fibrosus with $2\alpha=60$ deg

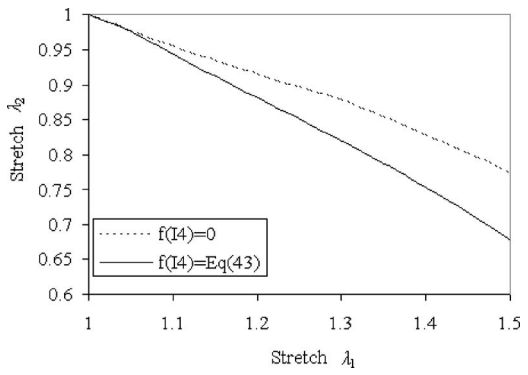


Fig. 15 Stretches in x_2 direction versus stretches in x_1 direction for multilayer AO annulus fibrosus with $2\alpha=120$ deg

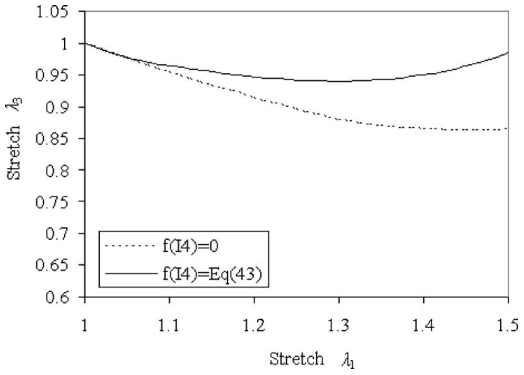


Fig. 16 Stretch in x_3 direction versus stretch in x_1 direction for multilayer AO annulus fibrosus with $2\alpha=120$ deg

stretches λ_2 and λ_3 (by applying different fiber-matrix interaction factors) under the traction free boundary conditions ($\sigma_2=\sigma_3=0$) versus the principal stretches λ_1 in the loading direction, respectively. The principal stretches λ_2 and λ_3 predicted by the constitutive model with a fiber-matrix shear interactions factor $f(I_4)$=Eq. (43) are denoted by solid lines in Figs. 15 and 16, respectively. Those predicted by $f(I_4)=0$ are represented by dotted lines in Figs. 15 and 16. Again, different predictions are obtained using the two different fiber-matrix interaction parameters. Additional experiments, especially for the case of $2\alpha=120$ deg (multilayer samples) would be very useful in further validation and refining this aspect of the model.

5.2 Biaxial Tensile Deformation. We further investigate the proposed constitutive model by applying it to biaxial tensile deformation. The biaxial tensile deformation is implemented by keeping the circumferential direction (1-direction in Fig. 5) prestrained at a certain strain (E_{11}) while pulling in the axial direction (2-direction in Fig. 5). Traction free boundary condition in the radial direction ($\sigma_3=0$) is imposed. The fiber-matrix interaction factor $f(I_4)$ is taken to be given by Eq. (43). Figure 17 shows the axial and circumferential second Piola-Kirchhoff stresses versus the axial Lagrangian strain under different circumferential prestrains. The results in Fig. 17 follow the same trends as the experimental results of Bass et al. [16].

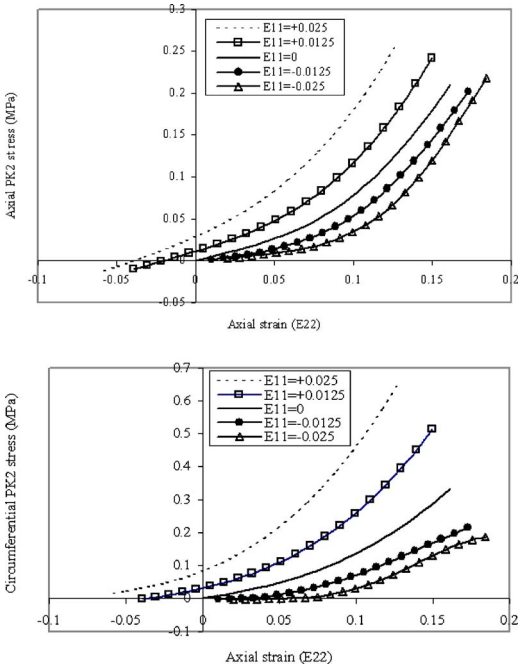


Fig. 17 Stress versus Lagrangian strain in biaxial deformation

6 Summary

The annulus fibrosus is an important integral constituent of the intervertebral disk and plays a vital role in the function and pathology of the disk. The correct characterization of the mechanical behavior of the intervertebral disk requires knowledge about the load deformation characteristics of the annulus fibrosus. For this purpose, a simple fiber reinforced constitutive model to characterize the nonlinear anisotropic of the human annulus fibrosus has been presented. By representing the strain energy function as a sum of matrix, fiber, and fiber-matrix shear interaction energies, the proposed constitutive model has only seven material parameters (six for incompressible ground matrix substances). The seven material parameters can be obtained consecutively by matching experimental uniaxial data. A simple geometric description of the shearing between the fiber and matrix is provided and this quantity is used to represent the fiber-matrix shear interaction energy. Moreover, a detailed investigation of the fiber-matrix shear interaction shows that this quantity is of great importance for the effectiveness of a constitutive model for the annulus fibrosus. It is found that comparison of stress-strain curves alone is inadequate for model validation. Comparisons with angle changes between the two families of fibers and with principal stretches are essential for model validation.

Acknowledgment

The authors would like to thank Medtronic-Sofamor Danek for research support. Helpful discussions with Professor Stephen Ondra and Professor Aruna Ganju, Department of Neurological Surgery, Feinberg School of Medicine, Northwestern University, are greatly appreciated.

B. M. is grateful to Wolfgang Knauss for a post-doctoral fellowship in 1987 and for his many insights into constitutive modeling of nonlinear materials.

References

[1] Hickey, D. S., and Hukins, D. W. L., 1980, "X-Ray Diffraction Studies of the Arrangement of Collagenous Fibers in Human Fetal Intervertebral Disc," J. Anat., **131**, pp. 81–90.
[2] Cassidy, J. J., Hiltner, A., and Baer, E., 1989, "Hierarchical Structure of the Intervertebral Disk," Connect. Tissue Res., **23**, pp. 75–88.
[3] White, A. A., and Panjabi, M. M., 1990, *Clinical Biomechanics of the Spine*, Lippincott, Williams & Wilkins, New York.
[4] Kelsey, J. L., and White, A. A., 1980, "Epidemiology and Impact of Low-Back Pain," Spine, **5**, pp. 133–142.
[5] Galante, J. O., 1967, "Tensile Properties of the Human Lumbar Annulus Fibrosus," Acta Orthop. Scand. Suppl., **100**, pp. 4–91.
[6] Wu, H. C., and Yao, R. F., 1976, "Mechanical Behavior of the Human Annulus Fibrosus," J. Biomech., **9**, pp. 1–7.
[7] Adams, M. A., and Green, T. P., 1993, "Tensile Properties of the Annulus Fibrosus, I The Contribution of Fiber-Matrix Interactions to Tensile Stiffness and Strength," Eur. Spine J., **2**, pp. 203–208.
[8] Green, T. P., Adams, M. A., and Dolan, P., 1993, "Tensile Properties of the Annulus Fibrosus: II Ultimate Tensile Strength and Fatigue Life," Eur. Spine J., **2**, pp. 209–214.
[9] Best, B. A., Guilak, F., Setton, L. A., Zhu, W., Sae-Nejad, F., Ratcliffe, A., Weidenbaum, M., and Mow, V. C., 1994, "Compressive Mechanical Properties of the Human Annulus Fibrosus and Their Relationship to Biochemical Composition," Spine, **19**, pp. 212–221.
[10] Skaggs, D. L., Weidenbaum, M., Iatridis, J. C., Ratcliffe, A., and Mow, V. C., 1994, "Regional Variation in Tensile Properties and Biochemical Composition of the Human Lumbar Annulus Fibrosus," Spine, **19**, pp. 1310–1319.
[11] Acarroglu, E. R., Latridis, J. C., Setton, L. A., Foster, R. J., Mow, V. C., and Weidenbaum, M., 1995, "Degeneration and Aging Affect the Tensile Behavior of Human Lumbar Annulus Fibrosus," Spine, **20**, pp. 2690–2701.
[12] Ebara, S., Latridis, J. C., Setton, L. A., Foster, R. J., Mow, V. C., and Weidenbaum, M., 1996, "Tensile Properties of Nondegenerate Human Lumbar Annulus Fibrosus," Spine, **21**, pp. 452–461.
[13] Fujita, Y., Duncan, N. A., and Lotz, J. C., 1997, "Radial Tensile Properties of the Lumbar Annulus Fibrosus are Site and Degeneration Dependent," J. Orthop. Res., **15**, pp. 814–819.
[14] Itradis, J. C., Kumar, S., Foster, R. J., Weidenbaum, M., and Mow, V. C., 1999, "Shear Mechanical Properties of Human Lumbar Annulus Fibrosus," J. Or-

thop. Res., **17**, pp. 732–737.
[15] Fujita, Y., Wagner, D. R., Biviji, A. A., Duncan, N. A., and Lotz, J. C., 2000, "Anisotropic Shear Behavior of the Annulus Fibrosus: Effect of Harvest Site and Tissue Prestrain," Med. Eng. Phys., **22**, pp. 349–357.
[16] Bass, E. C., Ashford, F. A., Segal, M. R., and Lotz, J. C., 2004, "Biaxial Testing of Human Annulus Fibrosus and its Implications for a Constitutive Formulation," Ann. Biomed. Eng., **32**, pp. 1231–1242.
[17] Holzapfel, G. A., Schulze-Bauer, C. A. J., Feigl, G., and Regitinig, P., 2005, "Single Lamellar Mechanics of the Human Lumbar Annulus Fibrosus," Biomechanics and Modeling in Mechanobiology **3**, pp. 125–140.
[18] Spencer, A. J. M., 1972, *Deformations of Fiber-reinforced Materials*, Oxford Science, NY.
[19] Spencer, A. J. M., 1984, *Continuum Theory of the Mechanics of Fiber-Reinforced Composites*, Springer, New York.
[20] Klisch, S. M., and Lotz, J. C., 1999, "Application of a Fiber-Reinforced Continuum Theory to Multiple Deformations of the Annulus Fibrosus," J. Biomech., **32**, pp. 1027–1036.
[21] Holzapfel, G. A., Gasser, T. C., and Ogden, R. W., 2000, "A New Constitutive Framework for Arterial Wall Mechanics and a Comparative Study of Material Models," J. Elast., **61**, pp. 1–48.
[22] Eberlein, R., Holzapfel, G. A., and Schulze-Bauer, C. A. J., 2001, "An Anisotropic Model for Annulus Tissue and Enhanced Finite Element Analyses of Intact Lumbar Disc Bodies," Comput. Methods Biomech. Biomed. Eng., **4**, pp. 209–230.
[23] Eberlein, R., Holzapfel, G. A., and Fröhlich, M., 2004, "Multi-Segment FEA of the Human Lumbar Spine Including the Heterogeneity of the Annulus Fibrosus," Comput. Mech., **34**, pp. 147–163.
[24] Elliott, D. M., and Setton, L. A., 2000, "A Linear Material Model for Fiber-Induced Anisotropy of the Annulus Fibrosus," J. Biomech. Eng., **122**, pp. 173–179.
[25] Elliott, D. M., and Setton, L. A., 2001, "Anisotropic and Inhomogeneous Tensile Behavior of the Human Annulus Fibrosus: Experiments Measurement and Material Model Predictions," J. Biomech. Eng., **123**, pp. 256–263.
[26] Sun, D. N., and Leong, K. W., 2004, "A Nonlinear Hyperelastic Mixture Theory for Anisotropy, Transport, and Swelling of Annulus Fibrosus," Ann. Biomed. Eng., **32**, pp. 92–102.
[27] Wagner, D. R., and Lotz, J. C., 2004, "Theoretical Model and Experimental Results for the Nonlinear Elastic Behavior of Human Annulus Fibrosus," J. Orthop. Res., **22**, pp. 901–909.
[28] Quapp, K. M., and Weiss, J. A., 1998, "Material Characterization of Human Medial Collateral Ligament," ASME J. Biomech. Eng., **120**, pp. 757–763.
[29] Gardiner, J. C., and Weiss, J. A., 2001, "Simple Shear Testing of Parallel-fibered Planar Soft Tissue," ASME J. Biomech. Eng., **123**, pp. 170–175.
[30] Fung, Y. C., 1981, *Biomechanics—Mechanical Properties of Living Tissues*, Springer-Verlag, New York.
[31] Malvern, L. E., 1969, *Introduction to the Mechanics of Continuous Medium*, Prentice-Hall, Englewood Cliffs, NJ.
[32] Ogden, R. W., 1984, *Non-Linear Elastic Deformation*, Dover, Mineola, NY.
[33] Guo, Z. Y., Peng, X. Q., and Moran, B., 2005, "A Composite-Based Hyperelastic Constitutive Model for Soft Tissue and its Application to the Human Annulus Fibrosus," (unpublished).
[34] Merodio, J., and Odgen, R. W., 2005, "Mechanical Response of Fiber-reinforced Incompressible Non-Linearly Elastic Solid," Int. J. Non-Linear Mech., **40**, pp. 213–227.
[35] Marchand, F., and Ahmed, A. M., 1990, "Investigation of the Laminate Structure of Lumbar Disc Annulus Fibrosus," Spine, **15**, pp. 402–410.
[36] ABAQUS, 2001, *ABAQUS/Standard User's Manual, Version 6.2*, Hibbitt, Karlson & Sorensen, Inc., Pawtucket, R.I.

Quasi-Static and Dynamic Buckling of Thin Cylindrical Shape-Memory Shells

Sia Nemat-Nasser
e-mail: sia@ucsd.edu

Jeom Yong Choi

Jon B. Isaacs

David W. Lischer

Center of Excellence for Advanced Materials,
University of California,
San Diego,
9500 Gilman Drive,
La Jolla, CA 92093-0416

To investigate the buckling behavior of thin and relatively thick cylindrical shape-memory shells, uniaxial compression tests are performed at a 295 K *initial temperature, using the CEAM/UCSD's modified split Hopkinson bar systems and an Instron hydraulic testing machine. The quasi-static buckling response of the shells is directly observed and recorded using a digital camera with a close-up lens and two back mirrors. To document the dynamic buckling modes, a high-speed Imacon 200 framing camera is used. The shape-memory shells with an austenite-finish temperature of* A_f=281 K*, buckle gradually and gracefully in quasi-static loading, and fully recover upon unloading, showing a superelastic property, whereas when suitably annealed, the shells do not recover spontaneously upon unloading, but they do so once heated, showing a shape-memory effect. The thin shells had a common thickness of* 0.125 mm *a common outer radius of* 2.25 mm *(i.e., a common radius, R, to thickness, t, ratio, R/t, of 18). A shell with the ratio of length, L, to diameter, D (L/D) of 1.5 buckled under a quasi-static load by forming a nonsymmetric chessboard pattern, while with a L/D of 1.95 the buckling started with the formation of symmetrical rings which then changed into a nonsymmetric chessboard pattern. A similar buckling mode is also observed under a dynamic loading condition for a shell with L/D of 2. However, thicker shells, with* 0.5 mm *thickness and radius* 4 mm *(R/t=8), buckled under a dynamic loading condition by the formation of a symmetrical ring pattern. For comparison, we have also tested shells of similar geometry but made of steel and aluminum. In the case of the steel shells with constrained end conditions, the buckling, which consists of nonsymmetric (no rings) folds (chessboard patterns), is sudden and catastrophic, and involves no recovery upon unloading. The gradual buckling of the shape-memory shells is associated with the stress-induced martensite formation and seems to have a profound effect on the unstable deformations of thin structures made from shape-memory alloys.* [DOI: 10.1115/1.2165241]

1 Introduction

Shape-memory alloys can sustain large strains, up to 6%–8%, and recover without macroscopically measurable residual deformations. This is called *superelasticity*. While the deformation is recovered, the process is dissipative, and the area within the closed loop of the material's stress-strain relation represents the energy lost to heating per unit volume of the sample. In addition, these alloys can display the so-called *shape-memory effect*. In this case, unloading leaves the material in its martensite phase, with a residual deformation that may be recovered upon heating. These and related unique properties of shape-memory alloys have been studied in connection with a variety of potential applications [1–5]; see Miyazaki and Otsuka [6], Otsuka and Wayman [7], and Otsuka and Ren [8,9] for more information on shape-memory alloys.

Recently, shape-memory alloys have been considered for potential application in developing multifunctional structures with superior load-bearing/energy absorption, and morphing capabilities [10–12], as well as actively controlling the buckling of the resulting flexible structures [13]. The buckling behavior of tubular structures made of various conventional materials has been studied under both quasi-static and dynamic loading conditions [14–19]. It is reported that the energy absorption capability of the NiTi plates under quasi-static compression loading is superior to that of aluminum plates [20]. There has been a limited amount of experimental work in characterizing the dynamic compression behavior of shape-memory alloys [21–28], particularly because of difficulties with controlling the strain rate under dynamic conditions. Recently, this issue has been carefully examined by Nemat-Nasser and co-workers [27,28], who produce stress-strain relations over a broad range of strain rates, and report the existence of a critical strain rate, below which material deforms by the formation of stress-induced martensites, whereas above this critical strain rate, the material deforms by dislocation-induced slip of the parent austenite, even though it may be loaded in the superelastic temperature range. However, we are not aware of any study that addresses the direct observation of the buckling behavior of relatively thin ($R/t=18,8$) shape-memory alloy shells under quasi-static and dynamic loading conditions. For this class of materials, the buckling modes are critically affected by the concomitant austenite to martensite and reverse phase transformations which can prolong the process, dissipate energy, and, because of the superelastic and shape-memory effects, can allow for a full recovery after unloading. Here, we report direct experimental observations of the buckling modes of thin shells made of shape-memory materials, under both quasi-static and dynamic loads, with full recovery due to both superelastic and shape-memory effects. For comparison, we have also tested shells of similar geometries made of steel and aluminum. In the case of the steel shells, with similar end conditions, the buckling consists of nonsymmetric chessboard folds, is sudden and catastrophic, and involves no recovery upon unloading.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received February 3, 2005; final manuscript received November 22, 2005. Review conducted by K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of applied Mechanics, Department of Mechanical and Environmental Engineering, University of California – Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

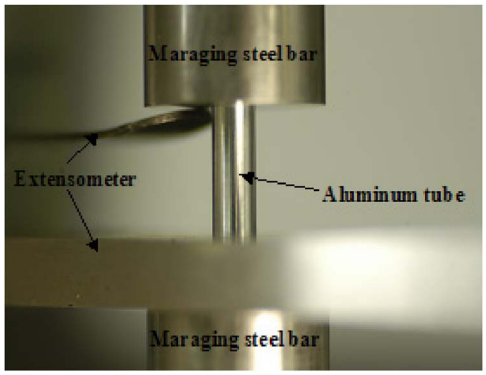


Fig. 1 Photograph of the buckling-test setup, showing the upper and lower platens made of maraging steel, an aluminum sample, and an extensometer

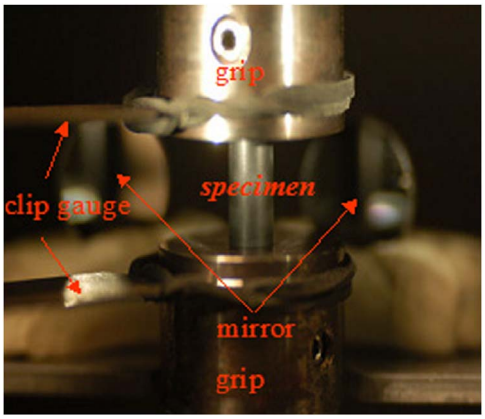


Fig. 3 Photograph of the fixed-end bucking-test setup, showing the sample and its reflections in the back mirrors

2 Experimental Procedures

2.1 Materials. The material used in the thin-tube experiments is a shape-memory alloy of composition 55.85 wt% Ni-Ti. The thin tubes have 4.5 mm outer diameter, 0.125 mm wall thickness (radius, R, to thickness, t, of 18), and various lengths. In the as-received state, the material's austenite-finish temperature, A_f is 281 K. To obtain an alloy with an A_f temperature above 295 K, some of the samples are annealed at 773 K for 150 min. Tubes of 0.5 mm wall thickness and 8 mm outer diameter ($R/t=8$) are machined from a cold-drawn 55.0 wt% Ni-Ti shape-memory alloy bar that has been annealed at 973 K for 60 min. After annealing, it was verified by quasi-static loading that the material is in fact superelastic at room temperature. Samples with the length-to-diameter ratio from 0.8 to 2.6 were produced and tested. In addition and for comparison with the shape-memory shells, maraging steel and 7075 aluminum samples were also produced and tested.

2.2 Quasi-Static Experimental Setup. Quasi-static compression tests are performed using an Instron hydraulic testing machine with a specially designed arbor. Figure 1 is a photograph of the buckling-test setup, showing the upper and lower platens made of maraging steel, an aluminum sample, and the extensometer to measure the sample's axial displacement. In this arrangement, the ends of the sample rest on the platens with no lateral constraint. Three views of the specimen are recorded using two back mirrors, as shown in Fig. 2. The photograph in Fig. 3 shows the buckling-test setup for a tube both of whose ends are constrained in maraging steel grips, each consisting of a pin and a set-screw, as shown in Fig. 4. The pin is designed for transferring the load and at the same time aligning the specimen.

The tests are conducted at 295 K (room temperature) under a displacement-controlled condition, with crosshead speed of the order of 10^{-3} mm/s. The axial specimen displacement is measured by an extensometer/clip gauge that has been calibrated before testing. The load and displacement signals are collected through an SCXI 1001 data acquisition system, using Labview software. Simultaneously, photographs are taken by a digital camera with a close-up lens, and the photos are synchronized with the corresponding axial load.

2.3 Dynamic Experimental Setup. Dynamic compression tests are performed at an initial temperature of 295 K, using the CEAM/UCSD's modified split Hopkinson bar systems [29,30]. Generally, the split Hopkinson bar consists of a striker bar, an incident bar, a transmission bar, and strain gauges [29,31]. In the present study, we use two different Hopkinson bar systems: a 1/2 in. Hopkinson bar, and a 3/16 in. mini-Hopkinson bar. The 1/2 in. Hopkinson bar is equipped with a momentum trap facility for recovery experiments [29]. The mini-Hopkinson bar can produce strain rates exceeding 10,000/s, using a suitably small specimen. To capture the dynamic buckling behavior of tube speci-

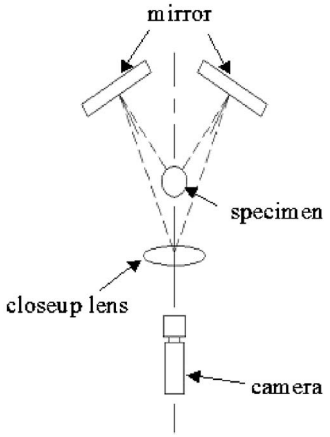


Fig. 2 Schematic diagram of photographic setup, showing two back mirrors, a focusing lens, and the recording camera

Fig. 4 Photograph of the grip parts and the specimen

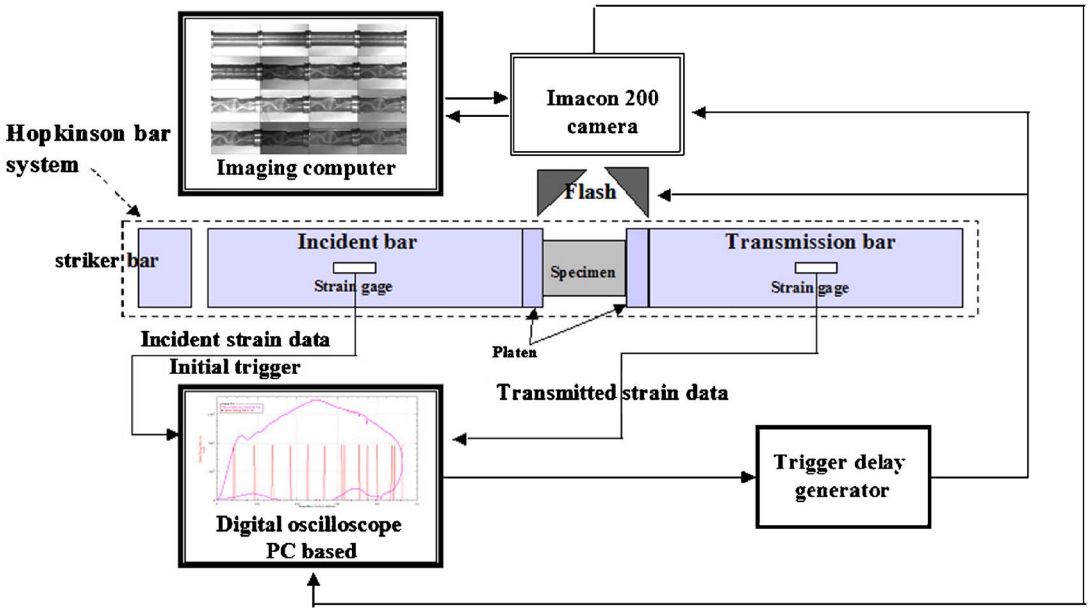


Fig. 5 Schematic diagram of the dynamic-test setup, showing the Hopkinson bar, a PC-based digital oscilloscope, a trigger delay generator, an Imacon 200 framing camera, and an image processing computer displaying a set of actual images

mens, we have used the high-speed Imacon 200 framing camera.

Figure 5 is a schematic diagram of the dynamic buckling-test setup. It shows a Hopkinson bar, a PC-based digital oscilloscope, a trigger delay generator, an Imacon 200 framing camera, and an image processing computer. The tube specimen with end platens is sandwiched between the incident and transmission bars. A gas gun propels a striker bar that impacts the incident bar, generating in it a compressive pulse that travels along the bar toward the specimen. Some of the pulse's momentum is transmitted through the specimen into the transmission bar, and the remaining momentum is reflected back into the incident bar as a tensile release pulse. The strain gauge mounted on the incident bar records the incident compressive and the reflected tensile pulses, and the strain gauge mounted on the transmission bar records the transmitted compressive pulse. These recorded incident, transmitted, and reflected pulses are analyzed through a Viewpoint program to obtain the load-displacement response of the specimen.

The strain gauge mounted on the incident bar also operates a trigger delay generator to turn on the camera which captures the buckling response of the specimen. The Imacon 200 framing camera, manufactured by Hadland Inc., can be programmed to record a sequence of separate images. The time delay of each individual channel can be controlled to acquire 16 images at prescribed time intervals. The accompanying software allows for digital manipulation of the pictures and for image analysis to obtain maximum information from the pictorial data.

To constrain the ends of the tube specimen, we used a platen whose impedance is matched with that of the bars. Figure 6 shows an engineering drawing of the platen. Figure 7 shows a photograph of the platens and the tube specimen. In the mini-Hopkinson bar, some tests are conducted with constrained end conditions, using a pin, placed in the tube, to limit the tube's total axial shortening.

3 Experimental Results and Discussion

3.1 Buckling Under Quasi-Static Loading Condition. Figure 8 displays three views of a NiTi shape-memory shell of 0.125 mm wall thickness ($R/t=18$) at various indicated stages of the experiment. These three views are obtained by direct photographic recording of the deformation of the sample (center) and the two back mirrors (left and right), as discussed in connection with Fig. 3. The NiTi thin shell buckles gradually and with a triple-fold symmetry, as the axial load is increased (Fig. 8(*b*)), and the folds recover, once the shell is fully unloaded (Fig. 8(*c*)), showing a superelastic response.

To confirm that the recovery shown in Fig. 8 is in fact due to the superelastic properties of the material and not just elastic recovery common to most metals, we produced a geometrically similar shell made of maraging steel, and subjected it to the same loading conditions. Figure 9 shows the three views of the buckled shell. The buckling occurred when the load reached a value near the yield stress of the alloy, did not involve any axisymmetric (rings) patterns, and was catastrophic without any recovery.

Figure 10 displays the variation of the load with the displacement under a displacement-controlled loading with a cross-head speed of 6.99×10^{-3} mm/s (nominal strain rate $\approx10^{-3}$/s). The L/D is 1.5. Both ends of the tube are fixed, using special grips shown in Fig. 4. Numbers in parentheses in this figure refer to the deformed states shown in Fig. 11. The load initially increases linearly with an increasing axial displacement, and then, once a maximum load is reached, it begins to decrease slowly with further increase in the axial displacement. By limiting the maximum

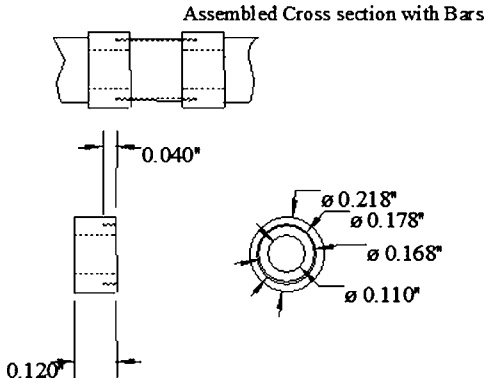
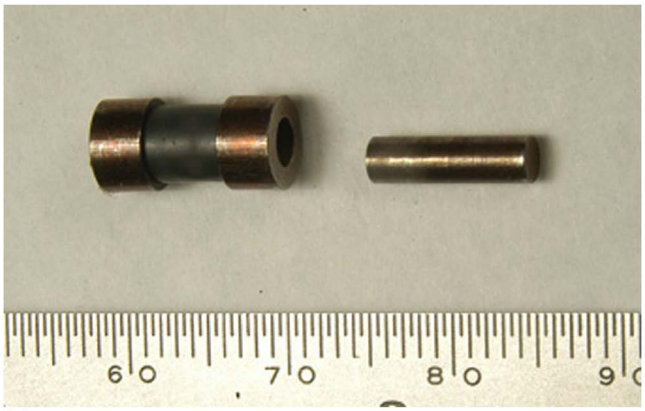
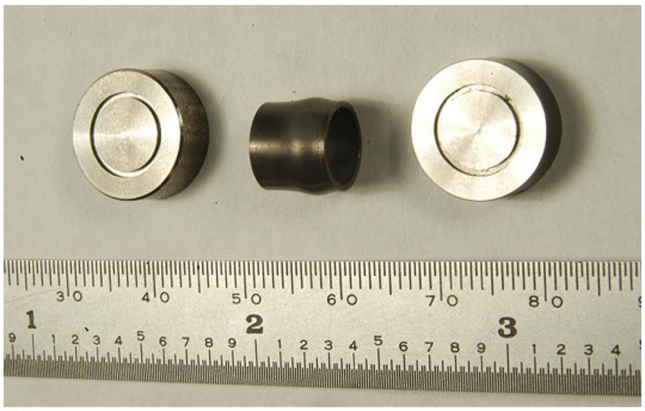


Fig. 6 Schematic diagram of constraining platens used in the mini-Hopkinson bar tests

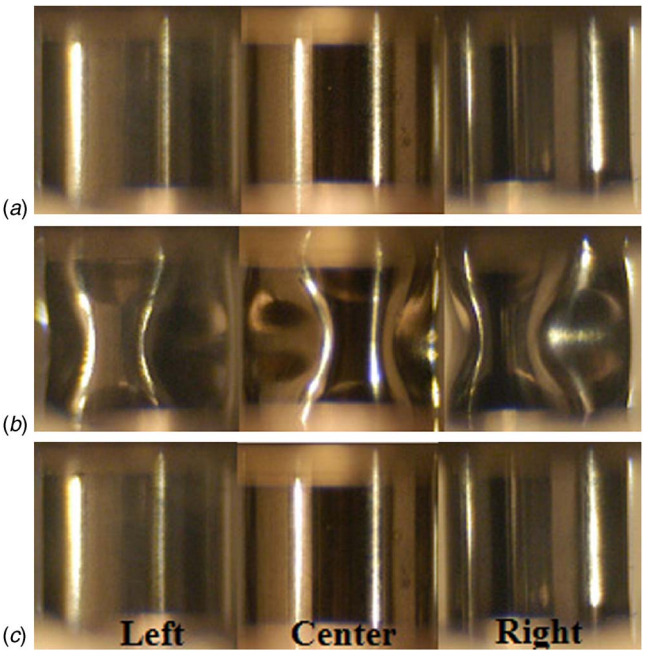


(*a*)



(*b*)

Fig. 7 Photograph of platens used in the (*a*) mini-Hopkinson bar with displacement limiting pin, and (*b*) the 1/2 in. Hopkinson bar



(*a*) (*b*) (*c*)

Fig. 8 Photographs of NiTi tube (R/t=18) buckling in uniaxial compression, showing the tube at its (*a*) initial stage, (*b*) the maximum displacement, and (*c*) the unloaded stage

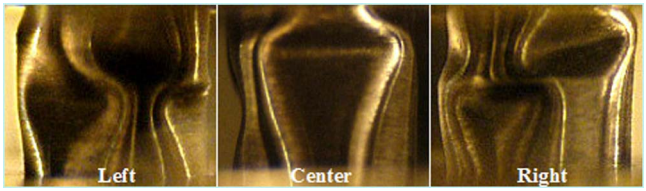


Fig. 9 Photograph of the buckled maraging steel tube (R/t=18)

axial displacement, the material remains superelastic, displaying total recovery upon unloading. During unloading, the load first decreases with a decrease in the axial displacement, and then levels off to a more or less constant value even though the displacement continues to decrease. Finally, the load rapidly decreases as the shell regains its initial configuration. The load-displacement curve forms a hysteresis loop, typical of the superelastic loop of

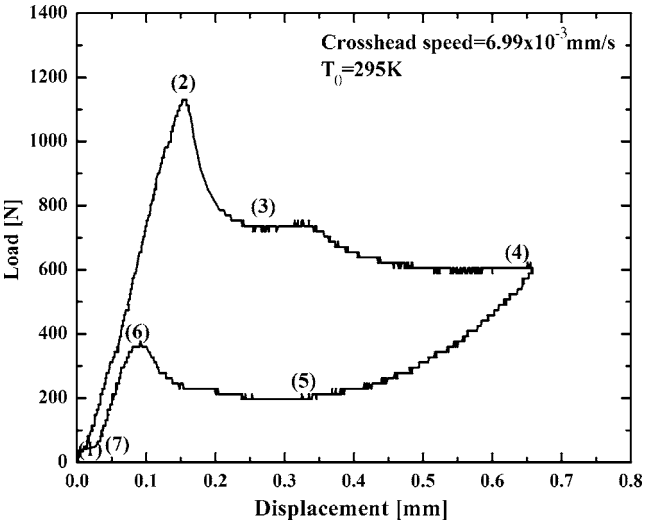


Fig. 10 The load-displacement relation for the NiTi tube (R/t=18) shown in Fig. 11, obtained under a displacement-controlled loading with a cross-head speed of 6.99×10^{-3} mm/s (nominal strain rate $\approx10^{-3}$/s)

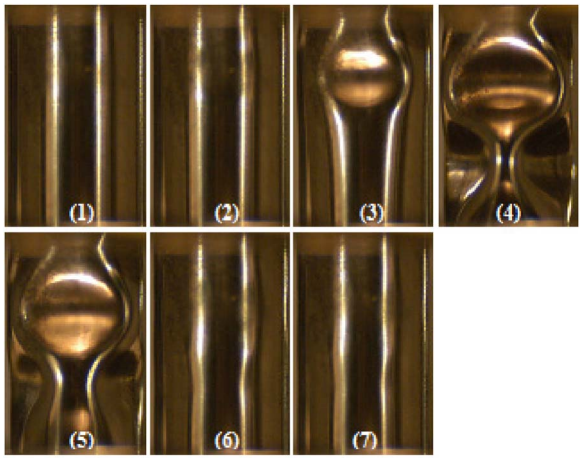


Fig. 11 Photographs of NiTi tube (R/t=18) buckling in uniaxial compression under a displacement-controlled loading with a cross-head speed of 6.99×10^{-3} mm/s (nominal strain rate $\approx10^{-3}$/s); numbers correspond to the load-displacement states of Fig. 10

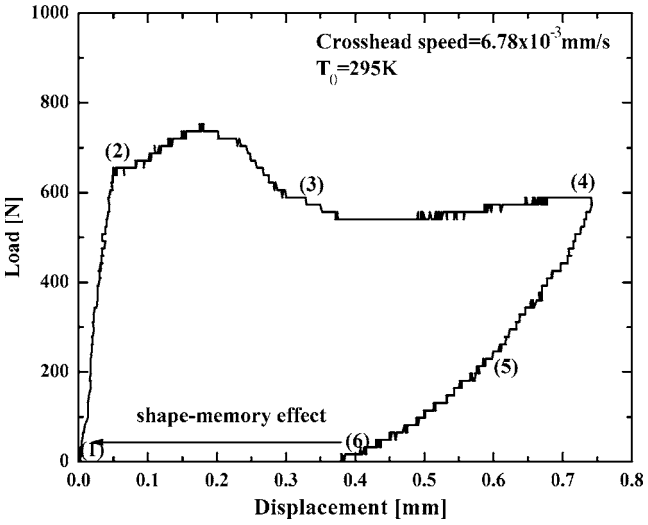


Fig. 12 Variation of load with displacement for a NiTi tube, obtained under a displacement-controlled loading with a cross-head speed of 6.78×10^{-3} mm/s (nominal strain rate $\approx10^{-3}$/s); the tube ($R/t=18$) is annealed at 773 K for 150 min, resulting in an A_f temperature higher than 295 K

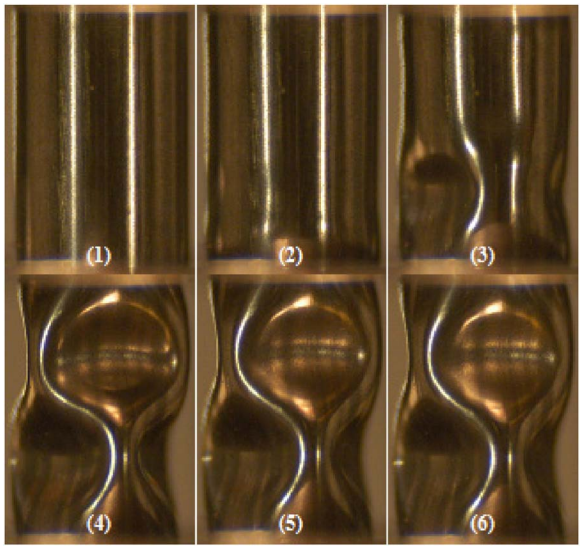


Fig. 13 Photographs of NiTi tube ($R/t=18$) buckling in uniaxial compression under a displacement-controlled loading with a cross-head speed of 6.78×10^{-3} mm/s (nominal strain rate $\approx10^{-3}$/s); numbers indicate the load state in Fig. 12, and the tube is annealed at 773 K for 150 min, resulting in an A_f temperature higher than 295 K

shape-memory alloys [6,7].

Figure 11 shows the progressive development of buckling of the shape-memory alloy shell. These pictures are taken directly from the sample at the load levels marked by the corresponding numbers in Fig. 10. The buckling begins close to the peak load level (2), after which the load continuously decreases from (2) to (4). At the maximum displacement of state (4), the sample has buckled into a chessboard pattern. As the displacement is decreased from (4) to (6) in Fig. 10, the buckled folds slowly disappear, as shown in Fig. 11, with a complete recovery at stage (6).

The buckling process in Fig. 11 relates directly to the material's phase transformation. During the loading within the superelastic range, the austenitic NiTi transforms into martensites when the buckling begins. As the deformation is continued, transformation from austenite to stress-induced martensite continues, resulting in a gradual and buckling response of the shell. Upon unloading, a reverse phase transformation, from martensite to austenite, takes place, leading to a full recovery. We conclude that this buckling recovery is due to the *superelastic property* of the material and not an elastic recovery common to some metals, e.g., steels. This is further supported by the fact that, under exactly the same test conditions, and with exactly the same sample geometry, the spontaneous recovery of the buckled sample (here attributed to superelasticity of the material) can be changed by simply changing the material's A_f temperature in such a manner that its superelasticity changes into shape-memory effect. This is discussed in the following.

Consider now the buckling response of a thin NiTi shell that is annealed at 773 K for 150 min, resulting in an A_f temperature higher than 295 K (room temperature). Figure 12 displays the load-displacement curve of such a shell with an L/D of 1.5, $R/t=18$, tested at room temperature (nominal strain rate $\approx10^{-3}$/s). The unloading now leaves a residual displacement, and, unlike in the superelastic case of Fig. 10, the hysteresis loop is no longer closed. Upon heating however, the residual deformation disappears and the shell recovers its initial configuration. Figure 13 shows the configuration of the shell at the loading stages that are indicated by the numbers in parentheses in Fig. 12. The buckling is again gradual and graceful, occurring at an almost constant load over a rather large displacement range, as seen in Fig. 12 from stage (2) to stages (3) and (4). The sample in stage (6) in Fig. 13 has not recovered, however, but, as shown in Fig. 14, it fully recovers upon heating, displaying a typical *shape-memory effect*. Since all the experimental conditions are kept the same for the tests shown in Figs. 11 and 13, and only the A_f temperature is changed by annealing, we may surmise that it is the phase-transformation phenomenon which underpins the gradual buckling process and its subsequent recovery by superelasticity (Fig. 11) and shape-memory effect (Fig. 13), respectively.

To examine the effect of the end constraint on the buckling mode of the thin shape-memory shells, we have also tested some samples with the unconstrained boundary condition. Figure 15 shows the load-displacement curve obtained in this manner for a shell of 1.95 L/D that has been annealed at 773 K for 150 min. Figures 16 and 17 show the corresponding buckling modes. The load-displacement curve is similar to that in Fig. 12, while the buckling mode is not quite similar to that in Fig. 13. As seen in Fig. 16, several rings are first formed along the length of the shell during the loading stages (2)–(4), and then this axially symmetric buckling mode changes to a non-symmetric mode similar to that in Fig. 13, leading to a chessboard pattern. The initial ring buckling is not observed in shells of smaller L/D's.

For comparison, a 7075 aluminum shell with a 0.127 mm wall thickness, 4.5 mm outer diameter ($R/t=18$), and 11.8 mm length ($L/D=2.6$) is also tested under an unconstrained end condition. Figure 18 displays the variation of the load with displacement for a displacement-controlled case with a cross-head speed of 6.71×10^{-3} mm/s (nominal strain rate $\approx0.57\times10^{-3}$/s). While the initial response is similar to the other cases, the buckling is drastic,

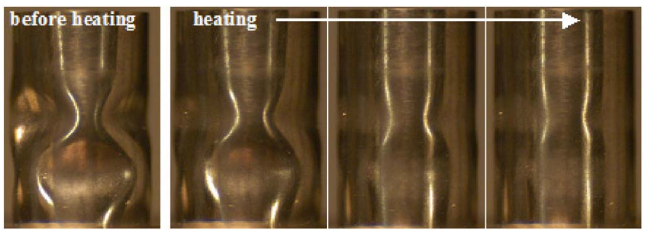


Fig. 14 Photographs of the recovery by heating of the buckled NiTi tube, showing a shape-memory effect

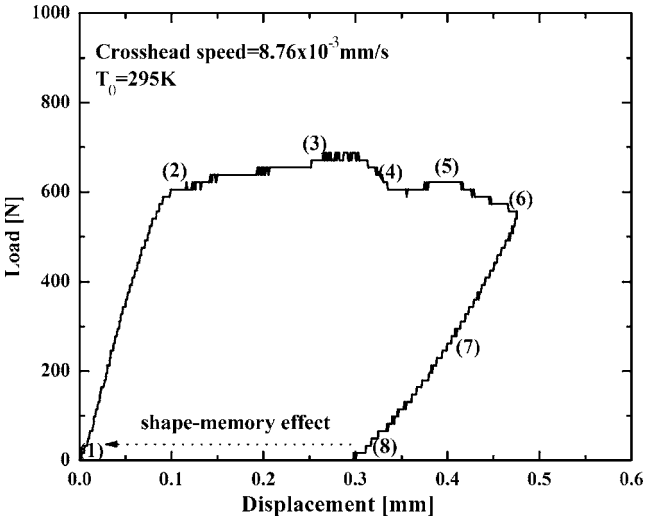


Fig. 15 Variation of load with displacement for a NiTi tube ($R/t=18$) with unconstrained ends, obtained under a displacement-controlled loading with a cross-head speed of 8.76×10^{-3} mm/s (nominal strain rate $\approx10^{-3}$/s); the tube is annealed at 773 K for 150 min, resulting in an A_f temperature higher than 295 K

as in the case of the steel shell. Figure 19 shows the progressive development of the buckling folds, corresponding to the load fluctuation in Fig. 18.

Figure 20 compares the load-displacement curves of the aluminum and NiTi shells. The area under each curve is the absorbed energy. The NiTi shells show greater energy absorption capability than the aluminum shell for a similar total displacement, and they have the capacity to fully recover.

In summary, it is observed that the buckling is sudden and catastrophic for maraging steel and 7075 aluminum shells, resulting in permanent deformations. In contrast, a NiTi shape-memory

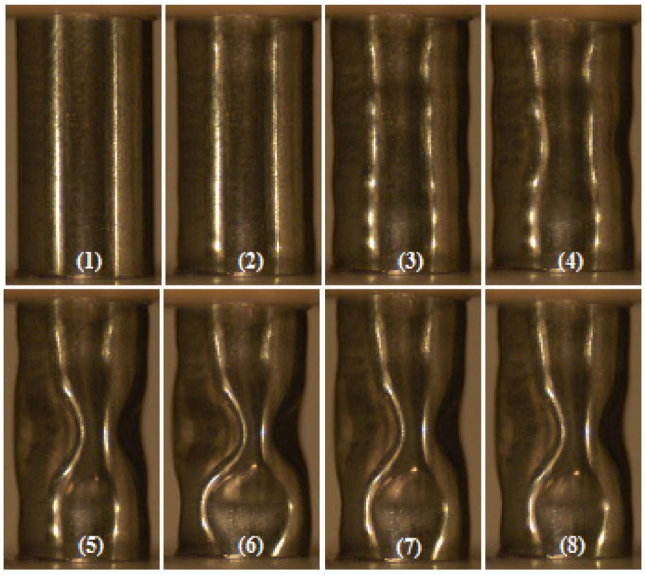


Fig. 16 Photographs of NiTi tube with unconstrained ends, buckling in uniaxial compression under a displacement-controlled loading with a cross-head speed of 8.76×10^{-3} mm/s (nominal strain rate $\approx10^{-3}$/s); numbers correspond to the load stages in Fig. 15, and the tube is annealed at 773 K for 150 min, resulting in an A_f temperature higher than 295 K

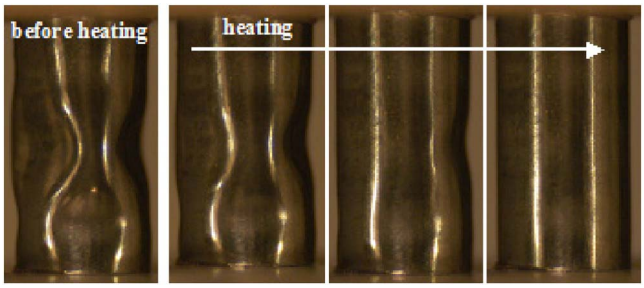


Fig. 17 Photographs of the recovery by heating of the buckled NiTi tube, showing a shape-memory effect

shell of similar dimensions buckles gradually and gracefully, fully recovering at a temperature in the superelastic range or upon heating (i.e., shape-memory effect). A NiTi shell of 1.5 L/D buckles by forming nonsymmetric chessboard folds, whereas with a L/D of 1.95, the initial buckling mode is by the formation of rings which then develop into a chessboard pattern (all shells have a R/t of 18). The NiTi shells are resilient with relatively good energy absorption capabilities.

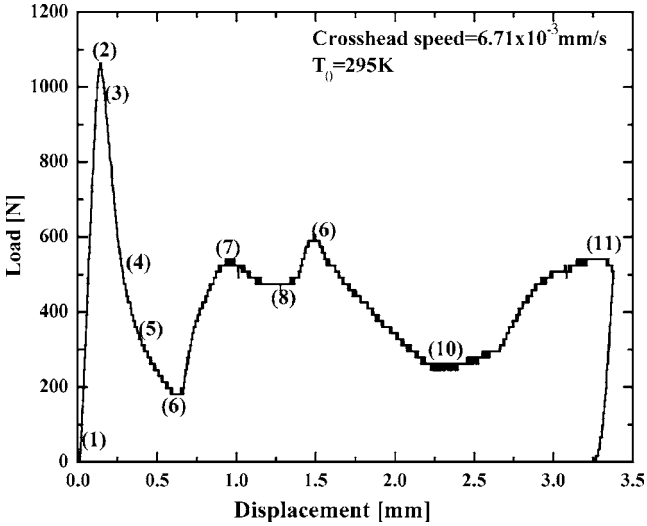


Fig. 18 Variation of load with displacement for an aluminum tube ($R/t=18$) with unconstrained ends, obtained under a displacement-controlled loading with a cross-head speed of 6.71×10^{-3} mm/s (nominal strain rate $\approx0.57\times10^{-3}$/s)

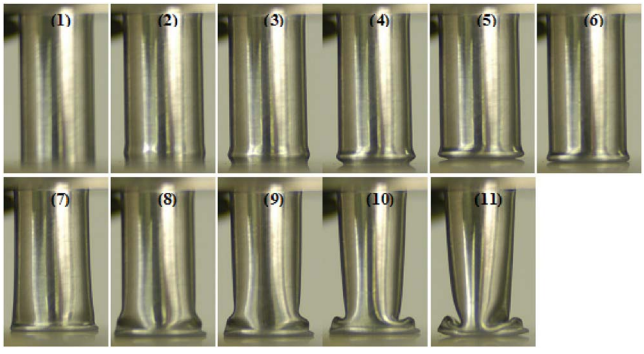


Fig. 19 Photographs of aluminum tube ($R/t=18$) with unconstrained ends, buckling in uniaxial compression under a displacement-controlled loading with a cross-head speed of 6.71×10^{-3} mm/s (nominal strain rate $\approx0.57\times10^{-3}$/s); numbers correspond to the load stages in Fig. 18

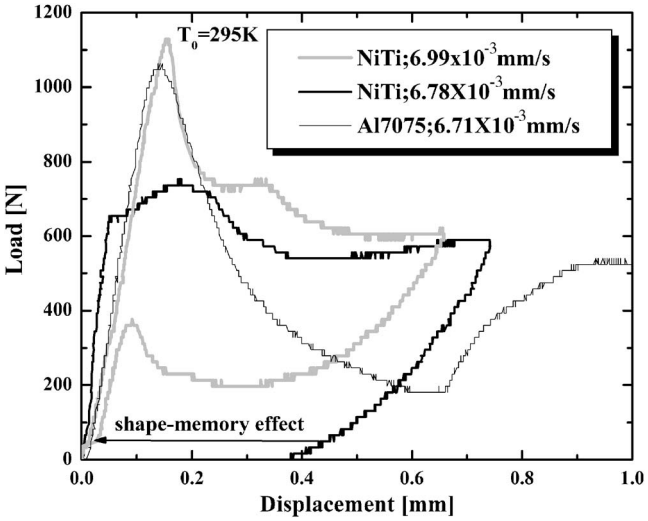


Fig. 20 Comparison of load-displacement relations for the indicated material and loading rate (of the order of 10^{-3}/s)

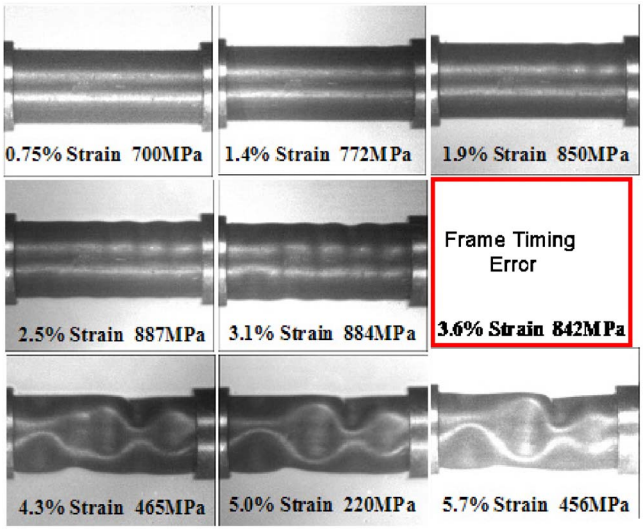


Fig. 22 Photographs of the dynamic buckling of a thin NiTi tube (R/t=18), obtained using mini-Hopkinson bar; L/D is 2.5, and the term "strain" refers to the axial shortening divided by L

3.2 Buckling Under Dynamic Loading Condition. Using the setup shown in Fig. 5, dynamic buckling tests are performed on NiTi cylindrical shells, at a 295 K temperature which exceeds the material's A_f temperature of 281 K. In what follows, the term "stress" refers to the axial load divided by the cross-sectional area of the shell, and the term "nominal strain" or simply "strain" refers to the total axial shortening of the shell normalized by the initial length of the sample. Since the sample buckles, both the strain and stress fields are nonuniform throughout the sample. In Fig. 21, the periodic spikes denote the camera-timing when the corresponding photo was taken (5 μs intervals). The experiment is performed using the mini-Hopkinson bar. The stress-nominal strain curve is also shown in this figure. The sample is a thin NiTi shell of 11.3 mm gauge length, 4.5 mm outer diameter, and 0.125 mm wall thickness (R/t=18; L/D=2.5). The average nominal strain rate was 1300/s. As shown in Fig. 21, the stress is almost constant until the nominal strain reaches approximately 3.6%, after which it rapidly decreases as the shell buckles.

The photos, taken at 5 μs intervals, are shown in Fig. 22. We also note the stress and strain at each picture. At 1.9% strain, the buckling begins by the formation of rings, becoming more pronounced at 2.5% and 3.1% nominal strains, while the stress remains essentially constant. At a nominal strain greater than 3.6%, the rings begin to transform into a nonsymmetrical chessboard pattern, causing a rapid decrease in the axial load. This nonsymmetrical buckling pattern is similar to that observed in the quasi-static case; see Figs. 11 and 13.

We have also examined the dynamic buckling behavior of a thick cylindrical shell of 0.5 mm wall thickness, 16 mm length, and 8 mm outer diameter (R/t=8; L/D=2.0). In this test, CEAM/UCSD's stress-reversal 1/2 in. Hopkinson bar is used [29,32]. Figure 23 exhibits the stress and the camera-timing marks (at 10 μs frame rate) as functions of the nominal strain. The average nominal strain rate is about 600/s. The stress-nominal strain curve has an almost closed hysteresis loop, showing the superelastic property. As shown in Fig. 23, the stress increases with an increase in the axial displacement, even while the sample is buckling. This buckling behavior is different from that of the thinner shell (R/t=8 versus 18). The photos of the buckling process are shown in Fig. 24. The buckling occurs in a ring pattern, and no

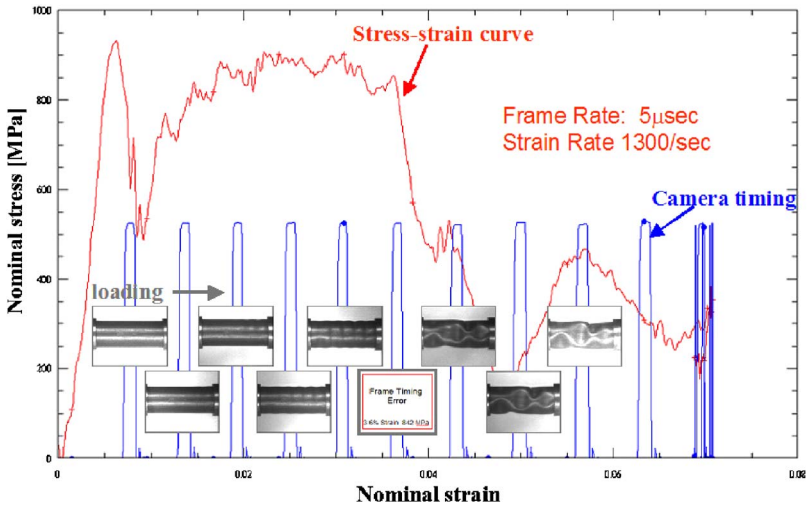


Fig. 21 Variation of stress with nominal strain, obtained using mini-Hopkinson bar; L/D is 2.5

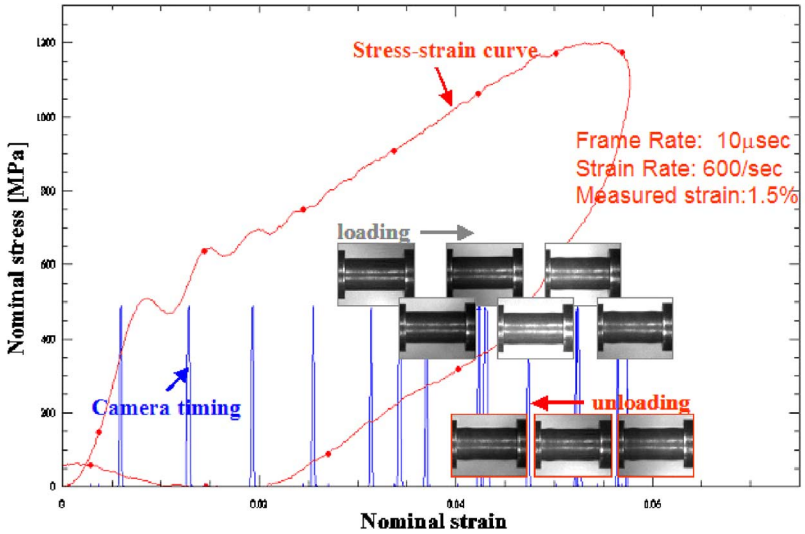


Fig. 23 Variation of stress with nominal strain, obtained using stress-reversal 1/2 in. Hopkinson bar; R/t is 8 and L/D is 2

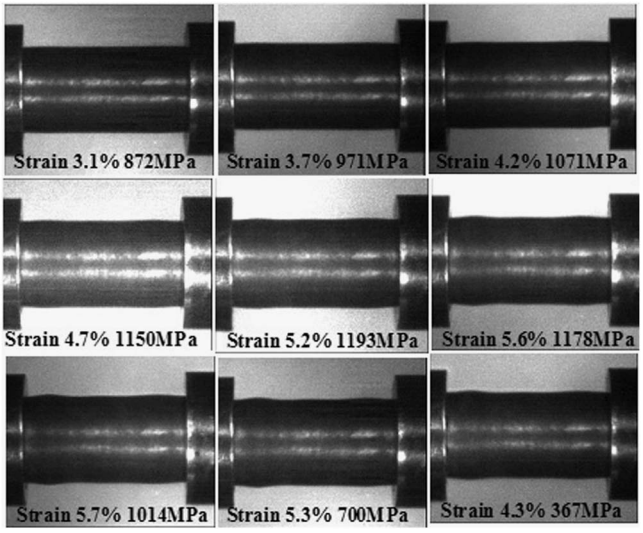


Fig. 24 Photographs of the dynamic buckling of a thick NiTi tube ($R/t=8$), obtained using stress-reversal 1/2 in. Hopkinson bar; L/D is 2, and the term "strain" refers to the axial shortening divided by L

chessboard pattern is formed. There is not a full recovery upon unloading, since the sample has experienced local plastic deformations, leading to a 1.5% residual strain.

To examine the effect of L/D on the buckling of the thick shell, another dynamic test is performed on a shell with 7 mm length, 8 mm outer diameter ($L/D=0.88$), and 0.5 mm wall thickness ($R/t=8$). Figure 25 displays the results, obtained at a nominal strain rate of about 1100/s. The measured residual strain is about 7%. As shown in Fig. 26, the buckling occurs in a ring pattern.

In summary, it is observed that the buckling mode depends on the thickness and R/t of the shell. Under dynamic loading conditions, the buckling of a thin ($R/t=18$) shell with a L/D of 2.5 begins with the formation of axisymmetric rings which then transform into a nonsymmetric chessboard pattern, similar to that for a L/D of 1.95 observed in the quasi-static loading case. A thick shell ($R/t=8$) however buckles in a ring pattern only.

4 Conclusions

Compressive tests are performed on cylindrical shells consisting of a NiTi shape-memory alloy, 7075 aluminum, and maraging steel, to study and contrast their buckling behavior under quasi-

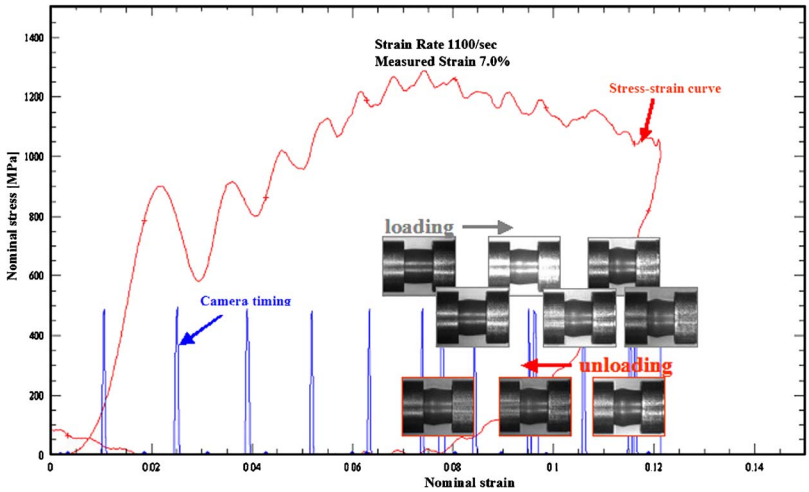


Fig. 25 Variation of stress with nominal strain, obtained using stress-reversal 1/2 in. Hopkinson bar; R/t is 8 and L/D is 0.85

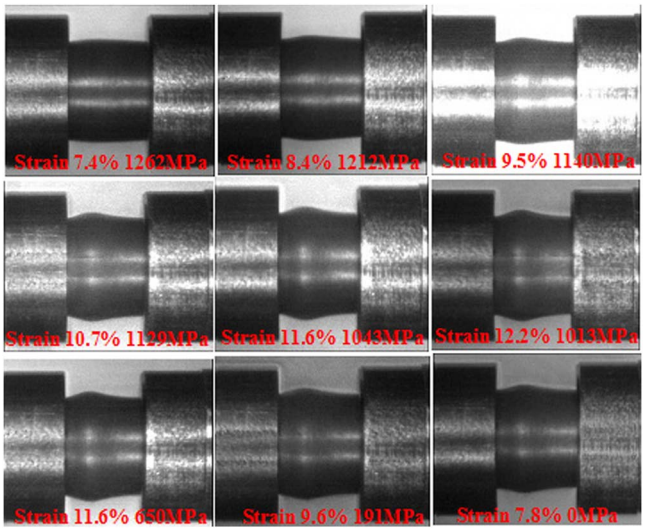


Fig. 26 Photographs of the dynamic buckling of a thick NiTi tube ($R/t=8$), obtained using stress-reversal 1/2 in. Hopkinson bar; L/D is 0.85, and the term "strain" refers to the axial shortening divided by L

static and dynamic loading conditions. For the dynamic cases, CEAM/UCSD's mini and recovery split Hopkinson bar systems are used together with a high-speed camera (Imacon 200), whereas, for the quasi-static cases, an Instron hydraulic testing machine is used together with a special photographic setup to simultaneously capture three images of each sample. It is observed that the NiTi shape-memory shells with an A_f temperature (281 K) less than the test temperature (295 K) buckle gradually and gracefully, fully recovering upon unloading, due to their *superelastic property*. However, under the same test conditions, the NiTi shells that are annealed at 773 K for 150 min and have an A_f temperature greater than 295 K, do not recover spontaneously upon unloading, but do so once heated, due to their *shape-memory effect*. The initial buckling mode of thin ($R/t=18$) shells depends on their L/D ratio. A thin shell of 1.5 L/D buckles in quasi-static loading into a nonsymmetric chessboard pattern. However, a similar thin shell but with L/D of 1.95, first forms axisymmetrical rings which then transform into a nonsymmetric chessboard pattern. A similar buckling mode is observed for a similar thin shell of 2.5 L/D, deformed dynamically. However, thick shells ($R/t=8$) buckled in a ring pattern only. The stress-induced martensite formation in NiTi shape-memory shells appears to have a profound effect on the shells' unstable deformation, and this can be effectively used to mitigate potential catastrophic failure, dissipate considerable amounts of energy, and, most important, obtain remarkable recovery of the buckled structure.

Acknowledgment

This work was supported by ONR (MURI) Grant No. N000140210666 to the University of California, San Diego. One of the reviewers requested that a typical low strain-rate uniaxial stress-strain curve would be useful. We have performed quasi-static experiments on the same shape-memory alloy and the results are summarized in Ref. [33].

References

[1] Duerig, T. W., Pelton, A., and Stockel, D., 1999, "An Overview of Nitinol Medical Applications," Mater. Sci. Eng., A, **237–275**, pp. 149–160.
[2] Van Humbeeck, J., 1999, "Non-Medical Applications of Shape Memory Alloys," Mater. Sci. Eng., A, **237–275**, pp. 134–148.
[3] Van Humbeeck, J., 2001, "Shape Memory Alloys: A Material and a Technology," Adv. Eng. Mater., **3**, pp. 143–156.
[4] Otsuka, K., and Kakeshida, T., 2002, "Science and Technology of Shape-Memory Alloys: New Developments," MRS Bull., pp. 91–100.
[5] Saadat, S., Salichs, J., Noori, M., Davood, H., Bar-on, I., Suzuki, Y., and Masuda, A., 2002, "An Overview of Vibration and Seismic Applications of NiTi Shape Memory Alloy," Smart Mater. Struct., **11**, pp. 218–229.
[6] Miyazaki, S., and Otsuka, K., 1989, "Development of Shape Memory Alloys," ISIJ Int., **29**, pp. 353–377.
[7] Otsuka, K., and Wayman, C. M., 1998, *Shape Memory Materials*, Cambridge University Press, Cambridge, UK.
[8] Otsuka, K., and Ren, X., 1999, "Recent Developments in the Research of Shape Memory Alloys," Intermetallics, **7**, pp. 511–528.
[9] Otsuka, K., and Ren, X., 1999, "Martensitic Transformation in Nonferrous Shape Memory Alloys," Mater. Sci. Eng., A, **273–275**, pp. 89–105.
[10] Lu, T. J., Hutchinson, J. W., and Evans, A. G., 2001, "Optimal Design of a Flexural Actuator," J. Mech. Phys. Solids, **49**, pp. 2071–2093.
[11] McGowan, A. -M. R., Washburn, A. E., Horta, L. G., and Bryant, R. G., 2002, "Recent Results from NASA's Morphing Project," *Smart Structures and Materials: Industrial and Commercial Applications of Smart Structures Technologies*, Proc. SPIE, 4698, pp. 97–111.
[12] Elzey, D. M., Sofla, A. Y. N., and Wadley, H. N. G., 2002, "Shape-Memory-Based Multifunctional Structural Actuator Panels," Smart Structures and Materials: Industrial and Commercial Applications of Smart Structures Technologies, Proc. SPIE, 4698, pp. 192–200.
[13] Thompson, S. P., and Loughlan, J., 2000, "The Control of the Post-Buckling Response in Thin Composite Plates Using Smart Technology," Thin-Walled Struct., **36**, pp. 231–263.
[14] Abramowicz, W., and Jones, N., 1984, "Dynamic Axial Crushing of Circular Tubes," Int. J. Impact Eng., **2**, pp. 263–281.
[15] Jones, N., 1989, *Structural Impact*, Cambridge University Press, Cambridge, UK.
[16] Reid, S. R., 1993, "Plastic Deformation Mechanisms in Axially Compressed Metal Tubes Used as Impact Energy Absorbers," Int. J. Mech. Sci., **35**, pp. 1035–1052.
[17] Singer, J., Arbocz, J., and Weller, T., 1998, *Buckling Experiments: Experimental Methods in Buckling of Thin-Walled Structures*, Wiley, New York.
[18] Seitzberger, M., Rammerstorfer, F. G., Gradinger, R., Degischer, H. P., Blaimschein, M., and Walch, C., 2000, "Experimental Studies on the Quasi-Static Axial Crushing of Steel Columns Filled with Aluminum Foam," Int. J. Solids Struct., **37**, pp. 4125–4147.
[19] Alghamdi, A. A. A., 2001, "Collapsible Impact Energy Absorbers: An Overview," Thin-Walled Struct., **39**, pp. 189–213.
[20] Suzuki, S., Urushiyama, Y., and Taya, M., 2004, "Energy Absorption Material Using Buckling Strength of Shape Memory Alloy Plate," *Smart Structures and Materials: Active Materials: Behavior and Mechanics*, Proc. SPIE, 5387, pp. 218–226.
[21] Ogawa, K., 1988, "Characteristics of Shape Memory Alloy at High Strain Rate," J. Phys. C, **3**, pp. 115–120.
[22] Ogawa, K., 1991, "Dynamic Behavior of Shape Memory Material," J. Phys.: Condens. Matter, **3**, pp. 215–221.
[23] Chen, W. W., Wu, Q., Kang, J. H., and Winfree, N. A., 2001, "Compressive Superelastic Behavior of a NiTi Shape Memory Alloy at Strain Rates of $0.001-750S^{-1}$," Int. J. Solids Struct., **38**, pp. 8989–8998.
[24] Liu, Y., Li, Y., and Ramesh, K. T., 2002, "Rate Dependence of Deformation Mechanisms in Shape-Memory Alloys," Philos. Mag. A, **82**, pp. 2461–2473.
[25] Liu, Y., Li, Y., Xie, Z., and Ramesh, K. T., 2002, "Dynamic Deformation of Shape-Memory Alloys: Evidence of Domino Detwinning?" Philos. Mag. Lett., **82**, pp. 511–517.
[26] Nemat-Nasser, S., Choi, J. Y., Guo, W. -G., Isaacs, J. B., and Taya, M., 2005, "High Strain-Rate, Small Strain Response of a NiTi Shape-Memory Alloy," J. Eng. Mater. Technol., **127**, pp. 1–7.
[27] Nemat-Nasser, S., Choi, J. Y., Guo, W. -G., and Isaacs, J. B., 2005, "Very High Strain-Rate Response of a NiTi Shape-Memory Alloy," Mech. Mater., **37**, pp. 287–298.
[28] Nemat-Nasser, S., and Choi, J. Y., 2005, "Strain Rate Dependence of Deformation Mechanisms in a Ni-Ti-Cr Shape-Memory Alloy," Acta Mater., **53**, pp. 449–454.
[29] Nemat-Nasser, S., Isaacs, J. B., and Starrett, J. E., 1991, "Hopkinson Techniques for Dynamic Recovery Experiments," Proc. R. Soc. London, Ser. A, **435**, pp. 371–391.
[30] Nemat-Nasser, S., and Isaacs, J. B., 1997, "Direct Measurement of Isothermal Flow Stress of Metals at Elevated Temperatures and High Strain Rates with Application to Ta and Ta-W Alloys," Acta Mater., **45**, pp. 907–919.
[31] Follansbee, P. S., 1985, *High Strain Rate Compression Testing*, ASM Handbook: Mechanical Testing, 8, ASM, Materials Park, OH, pp. 198–203.
[32] Nemat-Nasser, S., 2000, *Recovery Hopkinson Bar Technique*. ASM Handbook: Mechanical Testing and Evaluation, 8, ASM, Materials Park, OH, pp. 477–487.
[33] Nemat-Nasser, S., and Guo, W. -G., 2006, "Superelastic and Cyclic Response of NiTi SMA at Various Strain Rates and Temperatures," Mech. Mater. (to be published).

On the Crush Worthiness of a Laterally Confined Bar Under Axial Compression

Herzl Chai
Department of Solid Mechanics, Materials and Systems,
Faculty of Engineering,
Tel-Aviv University,
Tel Aviv 69978, Israel

A combined experimental/analytical work is carried out to elucidate the energy absorption potential of laterally confined bars under monotonically increasing edge displacement. The thickness t and length L of the bar, as well as the wall-to-wall separation distance, h, are systematically varied. Real-time observations show that the deformation of the bar is characterized by progressive buckling and folding, with the fully compacted material exhibiting repetitious cell unit whose wavelength approximately equals four times the bar thickness. The specific crush energy is little sensitive to the thickness of the bar but strongly varies with t/h, the "volume fraction" of the structure, attaining a maximum when $t/h \approx 0.5$. The main sources for energy dissipation are simple compression, plate folding and friction between the bar and the constraining walls, the latter of which dominates for $L/t > 10$. The experimental data are found to be well predicted by simple analytic expressions derived from limit plasticity analysis and incompressible material behavior. The simple configuration studied may shed light on the behavior of more complex structures such as honeycombs, foams, and thin-walled tubes, and may serve as a basis for multi-layer design possessing improved crush energy.
[DOI: 10.1115/1.2047595]

1 Introduction

Synthetic cellular materials manufactured from metals, ceramics, paper, and carbon are used in diverse engineering applications including crush worthiness, heat conductivity, and shock mitigation. Porous materials are also commonly found in nature, e.g., balsa wood, cork, bone, and plant stalks. In addition to their high strength to weight ratio, these structures possess a wealth of geometric and material parameters that offer numerous avenues for optimal design. A comprehensive exposition of the subject can be found in a monograph by Gibson and Ashby [1] and a recent review article by Alghamdi [2]. The primary source for energy absorption in cellular structures, the subject of main interest in this work, is progressive buckling and folding of thin material elements of the structure. The specific behavior may be conveniently classified into two categories, depending on whether the compression load is applied normal to the cell axis (e.g., honeycombs and foams) or along it (e.g., thin-wall tubes). The latter case primarily differs from the former in the additional energy that is absorbed by stretching or compression along the periphery of the fold, and the fact that global buckling may interfere with the inner deformation process.

The response of cellular structures is generally characterized by the mean stress and the densification strain, the product that equals the specific crush energy. Systematic and detailed crush studies on hexagonal honeycombs [3], polymeric foams [4], and balsa wood [5], to name a few, show that the mean stress increases with the relative density or volume fraction of the cellular structure while the opposite trend occurs for the densification strain. In the case of tubes compressed along their axis, the pioneering treatments of the circular and rectangular section tubes by Alexander [6] and Abramowitz and Jones [7] are noted. Bardi et al. [8] carried out careful measurements of the geometric aspects of the fold pattern in circular tubes and identified the merit of various analytical models in terms of the ratio of tube radius and wall thickness. More particularized studies aimed at stabilizing the deformation sequence and improving the crush energy of tube structures include the effects of fillers [9–12] and circumferential grooves [13,14]. Another promising approach for enhancing the crush energy is offered by lateral constraints; compression studies on nestled tubes [15,16] or polymeric foams [4], for example, show that lateral constraints may stabilize the deformation process and significantly increase the collapse load relative to the unconstrained structures.

A common characteristic in all the cell geometries detailed above is that the crush energy is mainly absorbed in filling porosities by means of plastic folding and possibly material stretching. We propose here the laterally confined column or bar configuration illustrated in Fig. 1 as yet another potential concept for energy absorption. In addition to the mechanisms noted above, lateral confinement may offer additional energy absorption through axial compression and friction between the confining walls and the bar. It is envisioned that this simple and instructive configuration may be used as a building block for multi-plate design possessing improved crush performance. Previous studies on the bi-laterally constrained column, conclusively limited to the elastic response [17–20], show that the deformation is characterized by a buckling mode transition process that results from secondary buckling of flattened segments of the bar. It was also pointed out that this configuration may lead to large energy dissipation due to friction between the plate and the confining walls [18]. In this work, the energy absorption characteristics of a polyvinyl chloride (PVC) bar that is confined on all of its four sides and subject to axial compression is studied in situ, with emphasis placed on the deep loading or the plastic deformation regime. Figure 1 shows a schematic of the test specimen. The thickness and length of the bar as well as the gap between two of the four confining walls are systematically varied. The experimental apparatus is detailed in Sec. 2 while the test results are discussed in Sec. 3. A simple analytic model is developed in Sec. 4 to account for energy dis-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received January 18, 2005; final manuscript received March 15, 2005. Review conducted by S. Kyriakides. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

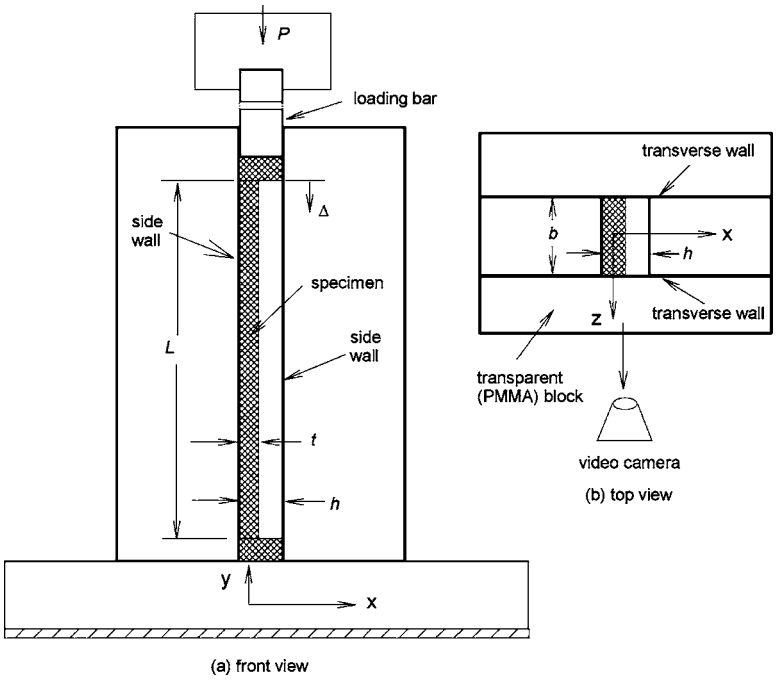


Fig. 1 Schematics of the test fixture and test specimen for a bar confined on all four sides and subject to axial compression

sipation as a function of the system parameters. Finally, Sec. 5 examines the crush performance of the confined bar relative to more traditional cellular structures.

2 Experimental

Tests are carried out to elucidate the energy absorption capability of a laterally confined rectangular bar under axial compression. Figure 1 details the test apparatus. Test samples of length L, thickness t, and width b are cut from flat PVC slabs. The latter is confined by four thick blocks, with two wall surfaces parallel to yz plane ("side walls") and the other two parallel to the xy plane ("transverse walls"). The side walls are spaced a distance h apart while the distance between the transverse walls matches the width of the bar, b. The majority of the tests are carried out with the bar initially resting on one of the two side walls (see Fig. 1). Some tests in which the bar is positioned in the middle of the two side walls are also performed. No distinction between these two configurations is made in the following as the test results are found to be virtually indistinguishable. One of the transverse walls is made of transparent PMMA to allow for in situ observations. The lower and upper edges of the bar are bonded with small pieces of PVC to facilitate a clamped-type support there. The specimens are supported at their lower edge by a steel plate while the upper one is compressed by a rectangular steel rod whose cross section coincides with the confinement area, $b \times h$. The upper part of the loading rod is firmly inserted into the loading head of the testing machine in order to increase its buckling resistance. To reduce geometric imperfections, a high degree of precision in the machining and alignment of all relevant components of the test apparatus is employed. As friction is a leading energy absorption mechanism in this study, the surfaces of the specimen and the confining walls are cleaned prior to testing with an alcohol solvent. Tests are carried out as a function of L and h for three different bar thicknesses, namely, 1, 2, and 4 mm. The analysis of the test data indicates that the ratio t/b may be a significant factor. Therefore, the latter is always maintained fixed, i.e., 0.25. The majority of the tests are carried out keeping the ratio L/t fixed, i.e., 26. Some tests are also carried out as a function of L/t, i.e., from 4 to 30. The volume fraction t/h is varied by varying the wall to wall gap, h, i.e., from 0.12 to 0.9.

The upper edge of the specimen is monotonically compressed using a screw driven testing machine (Instron). The deformation of the specimen is observed through the transparent transverse wall using a video camera. An additional video camera, whose image is superimposed on the main frame, is used to record the load. In all tests, the nominal strain (i.e., end displacement divided by the specimen length) is applied at a fixed rate of about 0.01/s. The axial load, P, and the crosshead displacement are recorded during the tests using a National Instrument data acquisition package. To circumvent the effect of machine compliance and the deformation absorbed in the loading rod and other components of the system, effects that become more and more significant as the ratio t/h increases toward unity, the relative shortening of the specimen ends is directly evaluated from the video image rather than the crosshead displacement. Some tests using solid cylindrical bars are also performed in order to elucidate the post-yield response of the PVC material used, which is a key ingredient in the analysis. Initial tests using uniform section cylinders show a tendency for barreling following yielding. To reduce this undesirable effect, the central part of the cylinders is thinned down in a gradual manner using a circular head cutting tool to form a hyperboloid test section. The radius of the tool is varied from 5 mm to infinity (i.e., uniform section). Tests are performed in either compression or tension. The deformation of these specimens is also observed in real time in order to facilitate concurrent measures of specimen diameter and load. If d_0 and d are the initial and concurrent diameter of the central cross section of the specimen, and F is the applied axial force, then, assuming material incompressibility and a rigid plastic behavior, the true stress and true strain are related by

$$\sigma_t = 4F/\pi d^2, \quad \varepsilon_t = 2\ln(d/d_0) \tag{1}$$

where positive quantities indicate compression.

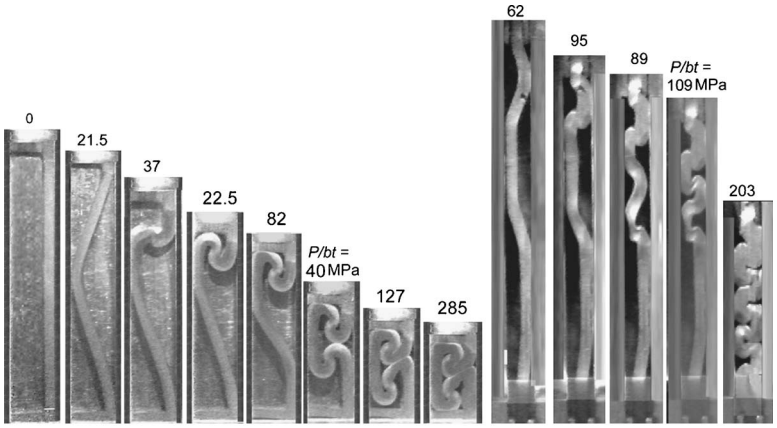


Fig. 2 Video motion pictures showing the deformation of two different laterally confined PVC bars

3 Test Results

We first discuss the material response before moving on to consider the main problem of interest. The test results show that the stress-strain response depends somewhat on the pretest neck radius. However, the yield stress, defined at the time of a noticeable change in the slope of the stress vs. strain curve, is found to be little sensitive to that parameter. The mean values of the yield stress in tension and compression from a number of tests having different initial neck radius are found to be 53 and 83 MPa, respectively, with standard deviation on the order of 5%. (The larger value in compression as compared to tension is a common phenomenon in polymers, attributed to the pressure sensitivity of the yield behavior.) These results will be discussed later in this work.

Let $\sigma = P/bh$ and $\varepsilon = \Delta/L$ be the nominal stress and strain, where P and Δ are the axial load and the axial displacement, respectively. In this way, a plot of σ vs. ε produces the stress-strain behavior of an effective material having a cross-sectional area bh. We further denote $v_f \equiv t/h$ as the volume fraction of the material within the confinement space bhL. Figure 2 shows the deformation sequences for two different bars having a ratio h/t of 4 (*a*) and 2 (*b*). As shown, the first stages in the deformation consist of column buckling, formation of a discrete contact point at the opposing wall, and the spread of the contact area.[1] This is followed by plate folding at the top end, leading to progressive buckling and folding toward the lower end. In the last stage of the deformation, the material is essentially completely compacted. The deformation for relatively large gap ratio (a) is characterized by plate folding while that of the tighter confinement (b) exhibits pronounced shearing. As shown by the solid line curve in Fig. 3, the nominal stress for the case study of Fig. 2(*a*) fluctuates somewhat during the deformation, but generally remains small except near densification, the latter of which occurs at a nominal strain of about 0.75.

As shown in Fig. 2, the material response greatly depends on the volume fraction, t/h. Figure 4 shows fully compacted bars in the unloading state for a number of volume fractions, all pertaining to $L/t = 26$. The stress-strain behavior for some representative cases is given in Fig. 3. As shown, the deformation exhibits a cell-like pattern characterized by plate folding. The number of these cells decreases with increasing h/t, a consequence of the greater room available for deformation. When h/t is decreased from 3, the deformation in a cell seems more associated with shearing than bending. These aspects of the deformation will be discussed further in Sec. 4 in connection with the analytic model. As shown in Fig. 3, the effective stress fluctuates during the de-

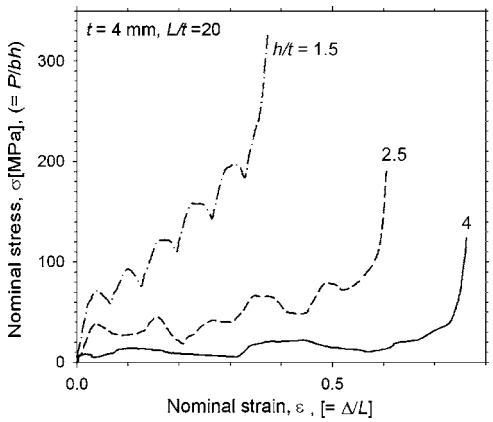


Fig. 3 Nominal stress vs. nominal strain for three different volume fractions, $t = 4$ mm, $L/t = 20$

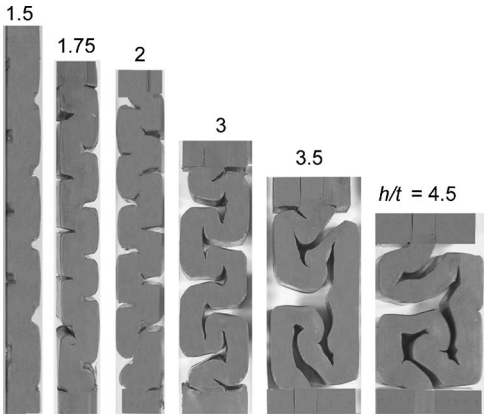


Fig. 4 Fully compacted specimens as they appear after unloading, $t = 4$ mm, $L/t = 26$

[1]The deformation of a bilaterally constrained column in the elastic range was studied in [18]. It was found that the deformation is characterized by a sequential mode transition process that results from secondary buckling. The latter initiates at contact zones formed between the column and the confining walls. This process was found to be inherently asymmetric, which leads to significant scatter in the buckling mode transition loads. An extension of this study to laterally constrained plates [21,22] shows that once a number of buckles are initiated, the buckling pattern in this case tends to approach the 1-D configuration.

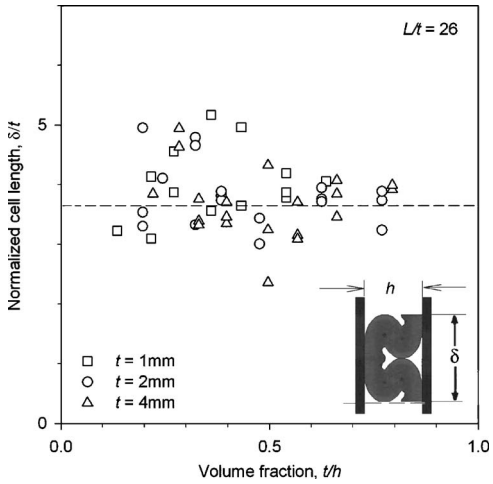


Fig. 5 Normalized cell length vs. volume fraction for a fixed value of L/t(=26); symbols are test results, dashed line is a possible fit

formation, the wavelength of which increases with increasing h/t. As the latter is decreased, the effective stress generally increases. One also observes that the densification strain decreases with h/t. The overall behavior described is very reminiscent of progressive collapse in other cellular structures (e.g., [3,8]).

The geometric aspects of the cell structure are an important ingredient in any analytic treatment. Figure 5 (symbols) details the variation of δ/t with t/h for the three specimen thicknesses tested, where δ denotes the cell wavelength. The data for this plot are generated from results as those shown in Fig. 4, e.g., by dividing the total number of cells in a given specimen by its original length, L. As shown, δ may be reasonably well approximated as

$$\delta = At, \quad A \approx 3.7 \tag{2}$$

Similar results are also found for other values of L/t (i.e., 4 to 30). Another useful quantity that can be extracted from the tests is the number of folds in a single cell, m. This quantity is established by dividing the total number of folds in the entire bar by the number of cells present. The results, shown as symbols in Fig. 6, seem to be well approximated by the dashed curve, given as

$$m = h/t \tag{3}$$

Note that the results shown are limited to relatively small values of t/h; as is apparent from Fig. 4, once the ratio h/t reduces from

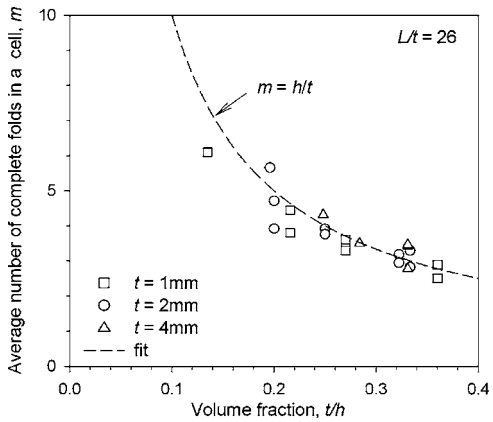


Fig. 6 The average number of folds in a unit cell vs. volume fraction, L/t=26; symbols denote test results, dashed line is a possible fit

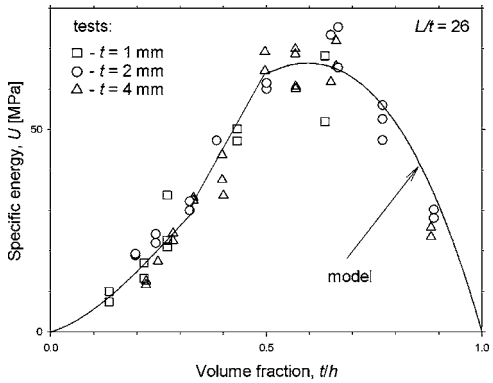


Fig. 7 The specific energy delivered to the bar (i.e., energy per unit confinement volume, bhL) vs. volume fraction for three bar thicknesses, L/t=26; the symbols and the curve denote test results and analytic prediction, respectively

3, the folds are not fully developed, with the deformation associated with shearing rather than bending.

The specific crush energy (i.e., the energy dissipated during the entire deformation history per unit confinement volume, bhL), U, is an important mechanical property. This quantity, given by the area under the stress-strain curves in plots such as those shown in Fig. 3, is shown in Fig. 7 (symbols) as a function of t/h for all three plate thicknesses studied. The variation of the densification strain with t/h for these tests is depicted in Fig. 8 as open symbols. (These data are obtained by dividing the end shortening at densification with the initial length of the bar, with the end shortening determined separately for each test from the video records.) Both plots, corresponding to the case L/t=26, show that the so-normalized results are quite insensitive to the specimen thickness. As shown in Fig. 7, the specific energy increases from zero, reaches a peak value at t/h=0.5–0.7, and steadily declines thereafter. The densification strain steadily declines from unity toward zero as t/h is increased. Another set of tests is carried out to determine the crush energy as a function of the bar length for a fixed values of h/t(=1.5). The results, corresponding to two choices of t (i.e., 2 and 4 mm), are shown as symbols in Fig. 9. As shown, while the specific energy is little insensitive to the bar thickness, it greatly increases with the length of the bar. Clearly, such behavior must be effected by frictional forces.

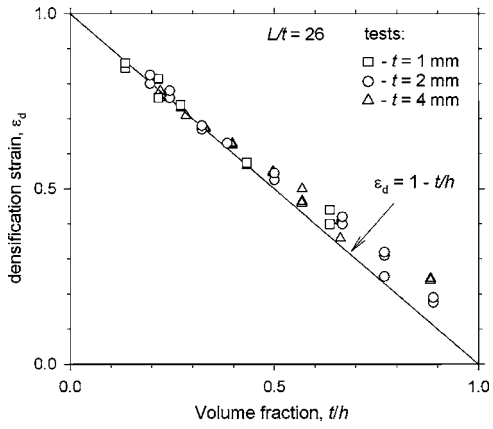


Fig. 8 Densification strain (i.e., end displacement at full stroke divided by the length of the bar) vs. volume fraction. The symbols correspond to the test data of Fig. 7, the solid line is analytic prediction derived based on material incompressibility.

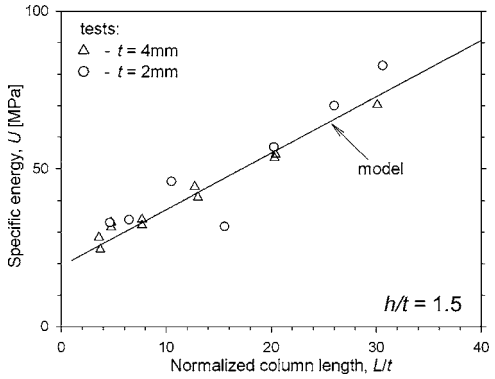


Fig. 9 Specific energy vs. normalized bar length for two different bar thicknesses such that $h/t=1.5$; the symbols and the curve denote test results and analytical prediction, respectively

4 Analytical Model

The external work done by the applied compression force is made up of the strain energy absorbed in the material and the energy dissipated by frictional forces transmitted by the constraining walls. Analytical expressions for these contributions are now derived by neglecting the elastic part of the strain energy and assuming that the material is rate independent, incompressible, and rigid perfectly plastic with yield stress σ_0. Accordingly, the densification strain is given by

$$\varepsilon_d = 1 - t/h \tag{4}$$

It should be noted that this relation provides an upper limit to the densification strain since, as can be seen from the last prints in Figs. 2(*a*) and 2(*b*), there always exist some gaps or cavities in the compacted material. Figure 8 (solid line) shows that Eq. (4) fits the test data reasonably well. (Note that the experimental data are actually a little larger than the theoretical values for relatively large values of t/h. This deviation is possibly due to the highly compressed material pushing out the PMMA confining wall.) Equation (4) was also proven useful as a check on the test data. Initially, the densification strain was determined from the crosshead displacement, which led to large errors for relatively large ratios of t/h. This departure, which is due to the bending of the loading rod and the effect of machine compliance, is largely suppressed once the end shortening is determined directly from the video images.

Figure 10 illustrates the deformation sequence used in the model; the presentation is specified to $n=3$, where n is the number of cells in the bar. The latter, of initial length L, thickness t, and width b, is constrained by four rigid walls, two in the plane of the figures and the other two transverse to it. The side walls are separated by a distance h while the transverse walls fit the bar width, b. The final, fully compacted configuration of the bar, shown in Fig. 10(*c*), consists of a cascade of n identical cells, each of length δ. Let λ denote the length of the original segment of the bar that corresponds to a single compacted cell of length δ. The incompressibility condition and Eq. (2) imply that

$$\lambda t = \delta h \quad \text{or} \quad \lambda = \delta h/t = Ah \tag{5a}$$

$$n = L/\lambda = L/Ah \tag{5b}$$

In evaluating the energy dissipation, it is assumed that the deformation history is composed of two stages. First, each segment of length λ in the undeformed state is compressed to a length δ as shown in Fig. 10(*b*). This process involves energy dissipation by friction and plastic deformation. The final stage involves the redistribution of all individual cells to form the compact configuration shown in Fig. 10(*c*). During this stage, only frictional energy is dissipated. As shown in Fig. 11, the main energy absorption

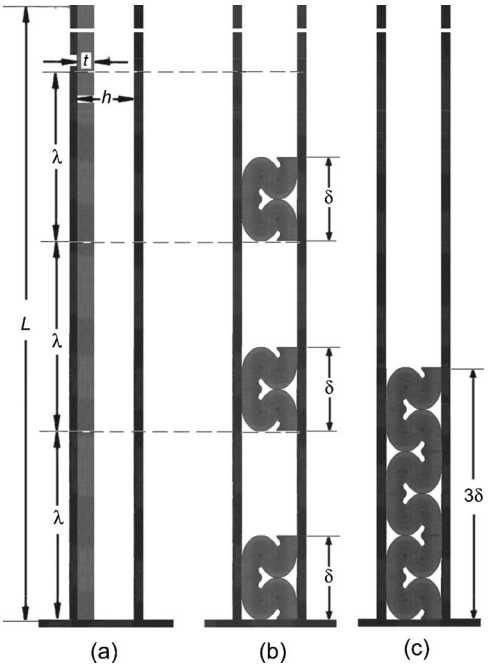


Fig. 10 Deformation sequence used in the analytic model. Illustration is specified to $n=3$, where n is the total number of cells in the bar.

mechanisms are thought to be due to simple compression (*a*), shearing (*b*), folding (*c*), and friction (*d*). In the followings, analytical expressions for each of these contributions are derived. Let

$$U = W/bhL \tag{6}$$

where W is the work done by the external load P and U denotes the specific work. The latter provides a useful baseline for optimizing the energy absorption based on the space available for deformation.

4.1 Compression Model. For sufficiently small clearance h, a uniform state of compression may be assumed [see Fig. 11(*a*)]. Under the previous simplifying assumptions, the specific energy in this case equals $\sigma_0\varepsilon_t$, where $\varepsilon_t=\ln(\lambda/\delta)$ is the true strain in compression. The strain energy for a fully compacted specimen is then given by $W=btL\sigma_0\varepsilon_t$. Making use of Eqs. (2) and (5), one has

$$U_c = \sigma_0(t/h)\ln(h/t), \tag{7}$$

where the subscript "c" indicates compression.

4.2 Shear Model. Consider an initial segment of length λ which undergoes buckling such that the buckle just touches the opposing side wall [see print I of Fig. 11(*b*)]. Let the associated end shortening be η. Assuming no friction, the work done in fully compacting the cell, i.e., from illustration (I) to (II) in Fig. 11(*b*), is given by $(\lambda-\eta-\delta)F_s$, where F_s is the axial load acting on the section of the bar, taken as $tb\tau_0$, where τ_0 is the yield stress in shear. Following [18], $\eta=\frac{1}{2}\int_0^\lambda(v')^2dy$, where v denotes the normal displacement of the bar and the prime sign indicates derivation with respect to the vertical coordinate, y, that extends from the base of the buckle. Substituting $v=(h-t)(1-\cos 2\pi y/\lambda)/2$ and integrating, one has $\eta=(\pi^2/4A)h(1-t/h)^2$ [18]. Making use of the preceding relations, the specific energy due to shearing in a bar of length L containing n cells is thus

$$U_s = (\sigma_0/\sqrt{3})(t/h)(1-t/h)[1-(1-t/h)(\pi/2A)^2] \tag{8}$$

where τ_0 is taken as $\sigma_0/\sqrt{3}$.

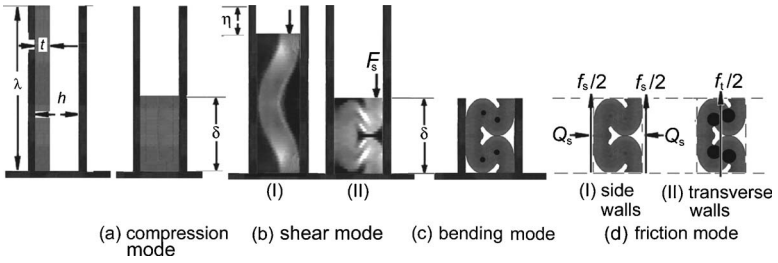


Fig. 11 Illustration of the various modes of energy absorption used in the analytic model: (*a*) compression, (*b*) shearing, (*c*) bending, and (*d*) friction. The dark spots in (*c*) indicate the locations of folds while those in illustration (II) of print (*d*) indicate the assumed contact regions between the bar and the transverse walls.

4.3 Bending Model. Referring to Fig. 11(*c*), it is assumed that the strain energy contained in a fully compacted cell is due to the creation of plastic hinges, the latter of which is indicated as small, solid circles in this print. Invoking plastic limit analysis, the strain energy needed to create a single fold is given by $M_0\pi$, where $M_0=\sigma_0 bt^2/4$ is the bending moment acting at the plastic hinge [1]. As is apparent from Fig. 6, the number of plastic hinges in a single cell, m, may be approximated as h/t. As indicated earlier, this expression is nominally valid for $t/h<1/3$. (Kinematical admissibility suggests that the area occupied by a single fold is $t\lambda/m=At^2$.) For a bar having n cells, the total number of complete folds is nm. Using previous expressions, the specific energy density due to bending is thus

$$U_b = (\pi/4A)\sigma_0 t/h \tag{9}$$

4.4 Friction Model. Friction forces are assumed to develop on all four confining walls. The frictional energy dissipated during the complete deformation process may be broken into two parts. The first is associated with the individual compaction of all cells as shown in Fig. 10(*b*), and the second is due to the downward motion of all cells toward the compact configuration shown in Fig. 10(*c*). Consider first the first mechanism from Fig. 11(*d*), where f_s (illustration I) and f_t (illustration II) denote the total frictional forces exerted on a single cell as it moves downward by the two side walls and the two transverse walls, respectively. The frictional work expanded on compacting n individual cells of initial length λ is thus $n(f_s+f_t)[(\lambda-\delta)/2]$. The second contribution, arising from the translation of all individual cells into the compact configuration shown in Fig. 10(*c*), is given by $(f_s+f_t)(\lambda-\delta)[0+1+2+\cdots+(n-1)]=(f_s+f_t)(\lambda-\delta)n(n-1)/2$. The combined frictional work is thus

$$W_f = (f_s + f_t)(\lambda - \delta)n^2/2 \tag{10}$$

The frictional forces f_s and f_t are now evaluated assuming a Coulomb-type friction law, with μ denoting the friction coefficient between the bar and the confining walls. The normal force Q exerted by the side walls on the cell boundary [see illustration I in Fig. 11(*d*)] must be balanced by the compressive stress acting in the two cell ligaments, i.e., $Q=2tb\sigma_0$, where it is assumed that the compressive stress in each ligament equals the yield stress, σ_0. The total frictional force exerted by the side walls on each cell, f_s, is thus given by $2\mu Q=4\mu tb\sigma_0$. The calculation of the frictional work associated with the transverse walls is a bit more involved. It is assumed that the frictional forces are mainly transmitted over that part of the fold section which undergoes compression, i.e., the four dark spots in illustration II of Fig. 11(*d*). Furthermore, the area for a single contact spot is taken as $t(t/2)=t^2/2$. The total frictional area associated with the two transverse walls for a cell containing m folds is thus mt^2. Assuming again that the normal stress acting over each contact area equals the yield stress of the material, one has $f_t=\mu\sigma_0 mt^2$. Substituting these results in Eq. (10), and making use of preceding relations, one find

$$U_f = (2/A)\mu\sigma_0(L/t)(t/h)^2(1 - t/h)(1 + 0.25h/b) \tag{11}$$

Equation (11) shows that the frictional energy is proportional to the friction coefficient, the yield stress, and the length of the bar. The contribution from the transverse walls is represented by the term $0.25h/b$. The latter vanishes for very wide bars ("plane strain" conditions).

Figure 12 shows the variation of the specific energy, normalized by the yield stress, with the volume fraction t/h for all four energy absorption models, where A is taken as 3.7 (Fig. 5), and, in the case of the frictional energy, $(\mu, L/t, t/b)$ are taken as (0.3, 26, 0.25). Note that the dotted lines in this figure correspond to regions outside the nominally applicable range of the corresponding models. It is apparent that for relatively large volume fractions, the contributions from compression and friction are dominant. One observes that the frictional energy is maximized at $t/h=0.65$.

4.5 Comparison With the Tests. As indicated in the derivation of the analytic models, the frictional model operates over the entire range of t/h while the bending and shearing models nominally apply in the range $t/h<1/3$ and $1/3<t/h<0.5$, respectively. The compression model is nominally limited to $t/h\geq 0.5$ as for smaller ratios the shearing mode becomes operational. Clearly, some interaction among the various deformation modes is expected at the interfaces between these regimes. However, explicit relations in these cases appear to be difficult if not impossible to elucidate. Based on these observations and limitations, the following hybrid relationship is proposed:

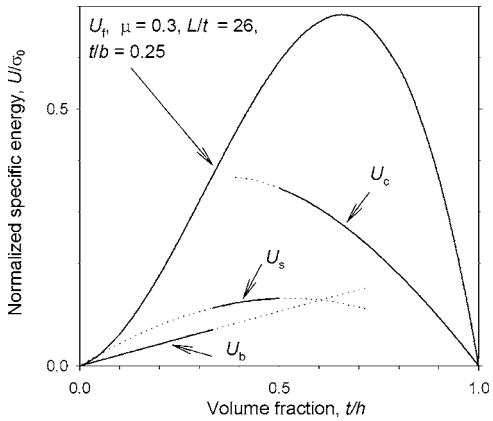


Fig. 12 Normalized specific energy vs. volume fraction for the four energy absorption models considered; dotted lines indicate a regions outside the nominally applicable range for each model. The frictional model is specified to $L/t=26$ and $\mu=0.3$.

$$U = U_f + U_c, \quad t/h \geq 0.5 \tag{12a}$$

$$U = U_b + U_f, \quad t/h \leq 1/3 \tag{12b}$$

$$U = (U_s + U_f) + (U_s - U_c)(h/t - 3), \quad 1/3 < t/h < 0.5 \tag{12c}$$

Equation (12*c*) implies a linear variation of the specific energy between its two extreme end points, i.e., friction plus compression and friction plus shear. It also leads to a small discontinuity in U at $t/h=1/3$, reflecting the difference between U_b and U_s. In order to compare this prediction with the test results, the yield stress and the friction coefficient need be specified. (The cell parameter A is taken as 3.7.) As discussed in Sec. 3, the yield stress in compression and tension differ, being equal to 53 and 83 MPa, respectively. Because buckling and folding involve both types of deformations, we shall take the average of these two values, namely $\sigma_0=68$ MPa. Published values for the friction coefficient of PVC on steel substrates vary from 0.21 to 0.5. In view of this large scatter, the friction coefficient was chosen such as to fit best the experimental data, i.e., 0.3. The analytical prediction from Eq. (12) based on these choices is shown as solid line curves in Figs. 7 and 9. As shown, the analysis seems to agree reasonably well with the test results over the range of parameters studied. The results in Fig. 7 show that the aforementioned discontinuity at $t/h=1/3$ could be removed by replacing U_s with U_b in Eq. (12*c*) without significant loss of accuracy. It also appears that the neglect of the possible interaction among the various deformation modes is justified from a practical viewpoint.

5 Discussions

The crush performance of cellular structures is generally characterized by the mean stress, σ_m, and the relative density or volume fraction, v_f ($=t/h$ in the present application). The specific crush energy is then given by $U=\sigma_m\varepsilon_d=\sigma_m(1-v_f)$. Then, from Eq. (12*b*), the mean stress for the laterally confined bar under the conditions of relatively small volume fraction and $t/b \ll 1$ is

$$\sigma_m/\sigma_0 = 0.54\mu(L/t)v_f^2 + 0.21v_f/(1-v_f). \tag{13}$$

It is interesting to compare this result with those from more common cellular structures such as honeycombs, foams, and hollow tubes. In the case where the cellular material is compressed in a direction normal to the cell axis, the mean stress may be expressed as

$$\sigma_m/\sigma_0 = Bv_f^C \tag{14}$$

where B and C are constants, given as (0.3, 1.5) and (0.28, 2) for open cell foams and hexagonal cell honeycombs, respectively [1]. [Note that the constants for the hexagonal cell honeycomb are based on the relation $v_f=(8/3)t/c$, where c is the distance between adjacent cell walls, with four of the wall cells being of thickness t while the other two of thickness $2t$.] Now consider the case where the loading is applied along the tube axis. For a circular section tube, the mean load, P_m, and the wavelength of the fold, γ, may be approximated as $P_m=11.2\sigma_0 t^2(R/t)^{1/2}$ and $\gamma/R=1.85\sigma(t/R)^{1/2}$ [23], where R and t are the mean tube radius and the wall thickness, respectively. An effective volume fraction value for this configuration may be obtained by considering the tube radius in the fully crushed state, i.e., $R+\gamma/2$, rather than the initial state. This gives $v_f=2\pi Rt/[\pi(R+\gamma/2)^2]$. Making use of this relation in the preceding expressions, and taking $\sigma_m=P_m/[\pi(R+\gamma/2)^2]$, one has

$$\sigma_m/\sigma_0 = 1.26v_f^{1.5}/(1+0.66v_f^{0.5})^2 \tag{15}$$

In the case of a square section tube, the mean load may be approximated as $P_m=12.2\sigma_0 t^2(c/t)^{0.37}$ [7], where t and c are the wall thickness and the wall to wall distance, respectively. Again, the effective volume fraction value for this configuration is constructed by considering the dimension of the section in the fully crushed state, i.e., $c+2H$, where $H=(\pi ct/2)^{0.5}$ [10]. The volume fraction then becomes $v_f=4tc/[c+2H]$, from which

$$\sigma_m/\sigma_0 = 1.3v_f^{1.63}/(1+1.25v_f^{0.5})^2 \tag{16}$$

The variations of the normalized mean stress with volume fraction for the various cellular structures detailed in Eqs. (13)–(16) are shown in Fig. 13. In the case of the laterally constrained bar, results are plotted for two choices of the friction coefficients, i.e., $\mu=0$ and 0.3, with $L/t=26$; note that the case $\mu=0$ corresponds to the bending model alone, and it is independent of L/t. For all cellular structures, the mean stress monotonically increases with the volume fraction. The performance of the foam, the square and the circular section tubes, and the frictionless bar is quite similar. The result for honeycomb is a bit lower, but this may be due to the fact that the loading in this case is normal to the tube axis. With friction included, the constrained bar exhibits a significant increase in the crush energy over the more common structures.

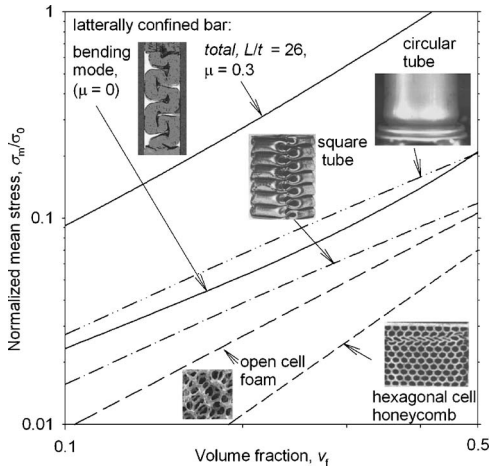


Fig. 13 Normalized mean stress vs. volume fraction for various cellular structures. The data for the circular and square tubes are constructed from results given in [23] and [7], respectively, while those for the honeycomb and foams are produced based on relations given in [1]. The inserts for the square tube, the circular tube, and the honeycomb structure are taken from [7,3,16], in that order.

6 Summary and Conclusions

A combined experimental/analytical effort is carried out to elucidate the energy absorption characteristics of laterally confined bars under monotonically increasing edge displacement. Real-time observations show that the deformation of the bar is characterized by progressive buckling and folding. In the context of crush worthiness the main contribution of the confinement is in the solicitation of further energy dissipation by means of axial deformation between adjacent folds as well as by friction between the bar and the walls. The overall deformation consists of repeating cells whose wavelength approximately equals four times the bar thickness. The specific energy delivered to the bar during the complete deformation process seems to be little sensitive to the thickness of the bar while strongly varying with the volume fraction, t/h.

Simple analytic expressions, based on limit plasticity and incompressible material behavior in the post-yield regime, are developed. The main energy absorption mechanisms are simple compression, bending arising from plate folding, and friction. The compression model is limited to a relatively large volume fraction while the bending model nominally applies for $v_f<1/3$. The combination of friction and compression provides good predictions of the actual behavior over the range $0.5<v_f<1$. For a less dense structure, a combination of the friction and bending models takes

over. For frictionless surfaces and a small volume fraction, the energy dissipation is mostly due to plate folding, similarly to the behavior observed for common cellular structures such as honeycombs, foams, and tubes. Frictional effects greatly improve the crush worthiness of the confined bar, with the friction contribution linearly increasing with the friction coefficient or the length of the bar. With friction included, the specific energy is maximized at a volume fraction in the neighborhood of 0.5–0.7. The lower end will probably be more useful because it entails a greater densification strain.

The constrained bar seem to provide a simple and instructive vehicle for understanding the crush performance of more complex configurations. This configuration may be used as a building block in a multi-plate structure to achieve improved crush worthiness over a large working space. The implementation of this in practice does not appear straightforward, however.

References

[1] Gibson, L. J., and Ashby, M. F., 1997, *Cellular Solids: Structure and Properties*, 2nd ed. Cambridge University Press, Cambridge, UK.

[2] Alghamdi, A. A. A., 2001, "Collapsible Impact Energy Absorbers: An Overview," Thin-Walled Struct., **39**, pp. 189–213.

[3] Papka, S. D., and Kyriakides, S., 1994, "In-Plane Compressive Response and Crushing of Honeycomb," J. Mech. Phys. Solids, **42**, pp. 1499–1532.

[4] Subhash, G., and Liu, Q., 2004, "Crushability Maps for Structural Polymeric Foams in Uniaxial Loading Under Rigid Confinement," Exp. Mech., **44**, pp. 289–294.

[5] Vural, M., and Ravichandran, G., 2003, "Dynamic Response and Energy Dissipation Characteristics of Balsa Wood: Experiments and Analysis," Int. J. Solids Struct., **40**, pp. 2147–2170.

[6] Alexander, J. M., 1960, "An Approximate Analysis of the Collapse of Thin Cylindrical Shells Tube Subject to Axial Compressive Forces," J. Mech. Eng. Sci., **13**, pp. 10–15.

[7] Abramowitz, W., and Jones, N., 1984, "Dynamic Axial Crushing of Square Tubes," Int. J. Impact Eng., **2**, pp. 179–208.

[8] Bardi, F. C., Yun, H. D., and Kyriakides, S., 2003, "On the Axisymmetric Progressive Crushing of Circular Tubes Under Axial Compression," Int. J. Solids Struct., **40**, pp. 3137–3155.

[9] Reddy, T. Y., and Wall, R. J., 1988, "Axial Compression of Foam-Filled Thin-Walled Circular Tubes," Int. J. Impact Eng., **7**, pp. 151–166.

[10] Reddy, T. Y., and Al-Hassan, S. T. S., 1993, "Axial Crushing of Wood-Filled Square Metal Tubes," Int. J. Mech. Sci., **35**, pp. 231–246.

[11] Seitzbergera, M., Rammerstorfera, F. G., Gradingerb, R., Degischerc, H. P., Blaimscheind, M., and Walchd, C., 2000, "Experimental Studies on the Quasi-Static Axial Crushing of Steel Columns Filled With Aluminum Foam," Int. J. Solids Struct., **37**, pp. 4125–4147.

[12] Singace, A. A., 2000, "Collapse Behaviour of Plastic Tubes Filled With Wood Sawdust," Thin-Walled Struct., **37**, pp. 163–187.

[13] Dipaolo, B. P., Monteiro, P. J. M., and Gronsky, R., 2004, "Quasi-Static Axial Crush Response of a Thin-Wall, Stainless Steel Box Component," Int. J. Solids Struct., **41**, pp. 3707–3733.

[14] Hosseinipour, S. J., and Daneshi, G. H., 2003, "Energy Absorption and Mean Crushing Load of Thin-Walled Grooved Tubes Under Axial Compression," Thin-Walled Struct., **41**, pp. 31–46.

[15] Reddy, T. Y., and Reid, S. R., 1979, "Lateral Compression of Tubes and Tube-Systems With Side Constraints," Int. J. Mech. Sci., **21**, pp. 187–199.

[16] Shim, V. P.-W, and Stronge, W. J., 1986, "Lateral Crushing in Tightly Packed Arrays of Thin-Walled Metal Tubes," Int. J. Mech. Sci., **28**, pp. 709–728.

[17] Chateau, X., and Nguyen, Q. S., 1991, "Buckling of Elastic Structures in Unilateral Contact With or Without Friction," Eur. J. Mech. A/Solids, **1**, pp. 71–78.

[18] Chai, H., 1998, "The Post-Buckling Behavior of a Bi-Laterally Constrained Column," J. Mech. Phys. Solids, **46**, pp. 1155–1181.

[19] Holmes, P., Domokos, G., Schmitt, J., and Szeber'enyi, I., 1999, "Constrained Euler Buckling: An Interplay of Computation and Analysis," Comput. Methods Appl. Mech. Eng., **170**, pp. 175–207.

[20] Pocueau, A., and Roman, B., 2004, "Uniqueness of Solutions for Constrained Elastica," Physica D, **192**, pp. 161–186.

[21] Chai, H., 2001, "Contact Buckling and Postbuckling of Thin Rectangular Plates," J. Mech. Phys. Solids, **49**, pp. 209–230.

[22] Chai, H., 2002, "On the postbuckling Behavior of Bilaterally Constrained Plates," Int. J. Solids Struct., **39**, pp. 2911–2926.

[23] Wierzbicki, T., Bhat, S. U., Abramowitz, W., and Brodkin, D., 1992, "Alexander Revised—A Two Folding Elements Model of Progressive Crushing of Tubes," Int. J. Solids Struct., **29**, pp. 3269–3288.

An Experimental Investigation of the Motion of Flexible Strings: Whirling

Bisen Lin

K. Ravi-Chandar[1]
Fellow ASME
e-mail: kravi@mail.utexas.edu

Center for Mechanics of Solids,
Structures and Materials,
Department of Aerospace Engineering and
Engineering Mechanics,
The University of Texas at Austin,
Austin, TX 78712-1085

Whirling of strings has been studied both theoretically and experimentally for many decades. According to linear theory, a heavy string can exhibit steady-state whirl only at its natural frequencies which form a discrete spectrum. The nonlinear theory, however, suggests that a string can undergo steady whirl at any frequency larger than the fundamental frequency and further that for each frequency between the nth *and the* $(n+1)$th *eigenvalue, there exist n distinct whirling modes. Quantitative experimental observations on such whirling have never been reported, although anecdotal evidence suggests the possibility of such whirl. In this paper, we examine the whirling of a string with negligible bending stiffness through experiments utilizing a stereo-vision imaging system. It is shown that steady motion exists only when the string whirls at its natural frequencies and that whirling motions for other frequencies exhibit rich dynamics that needs further exploration.* [DOI: 10.1115/1.2172270]

1 Introduction

Studies of the vibrations of strings have a long history; Bernoulli, Euler, Kirchhoff, Rayleigh and many others have explored various aspects of this problem. Routh [1] provides a summary of the analysis of strings and cables in great detail and provides a historical context. There is a vast recent literature on the transverse vibration of strings [2–9]. Most of these investigators focus on the problem of a taut string held between two points under a constant tension and set in motion to generate transverse deflections of the string. Even as early as 1858 observations on the appearance of nonplanar whirling under planar excitation was noted by Rayleigh. When forced by transverse harmonic motion far below the natural frequency of the linearized equations of motion, a string exhibits transverse oscillations that are confined to the plane of excitation. However, as the natural frequency is approached, and the amplitude of the transverse motion exceeds a critical level, the transverse motion of the string becomes unstable to perturbations; the resulting motion is a whirling motion where the string deflects not only in the plane of excitation, but also perpendicular to it. With further increases in the frequency, the amplitude of the whirling motion grows first and then diminishes abruptly; well beyond the natural frequency, the string resumes planar motion with smaller amplitudes. This behavior has been examined analytically by numerous investigators. Narasimha [6] pointed out that even under small amplitude motion, the equations governing the motion are inherently nonlinear. He showed that neglecting the axial stretch and changes in tension along the length of the wire are not justifiable. While this was implicit in the earlier analysis (see [1]), Narasimha [6] made the representation of the length and tension changes explicit and showed further that planar motion was unstable at a high enough amplitude. Tufillaro [10] developed a simple single-mode model for the string and examined the planar and nonplanar motions; periodic, quasi-periodic, and chaotic motions were observed in this model. O'Reilly and Holmes [8] and Nayfeh et al. [9] examined the dynamics of this whirling motion by the method of multiscales. These models are able to capture the essence of the problem: that planar motions are unstable near the natural frequencies of the linearized problem. In all these investigations, the primary motion was the string deflection in the plane of the driving force; the whirling or out-of-plane motion was the result of instabilities in the system.

The problem we consider in this paper is quite different—we examine the dynamics of a string attached to the base of a rotating motor at one end, hanging under self-weight and free of constraint at the other end. This problem was first considered by Bernoulli; solutions of the linearized problem in terms of Bessel functions were described by Fourier. Routh [1] describes many special cases of the hanging string or chain problem. Kolodner [11] provided a complete mathematical model for this *nonlinear* problem associated with the whirling of an inextensible, flexible string. While Kolodner predicted the qualitative behavior of steady-state solutions, he did not obtain a complete solution to the problem. Bessel's functions of the first kind of order zero were already known to represent the whirling shape of the string corresponding to the linearized problem [1]. Caughey [12] performed a perturbation analysis of the small amplitude oscillations near the natural frequencies of the linearized problem and found stable and unstable regimes of the solutions. Caughey [13] also presented an analysis of whirling of a string fixed at one end and constrained to lie on the axis of whirling at the other end; note that this constraint requires a nonzero tension at this end. Under these conditions, he obtained closed form solutions for the modes of free whirling of the string. Investigating the stability of the solutions, he demonstrated that the perturbed motions were Lyapunov stable: points of the string that are initially within a certain distance from the steady-state mode shapes continue to remain within this distance. The nonlinear eigenvalue problem associated with steady-state whirling of the string results in a two-point boundary value problem; solution of the two-point boundary value problem can be obtained by suitable numerical methods, such as the shooting method, the finite difference method, and so on. Coomer et al. [14] have recently revisited this problem and obtained such numerical solutions with MAPLE; we will obtain the solution with MATLAB.

In contrast to the taut string problem, very little experimental work has been reported in the literature concerning whirling. Western [15] used his wife's necklace chain as an example of an

[1]To whom correspondence should be addressed

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received September 2, 2005; final manuscript received December 13, 2005. Review conducted by R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Enginnering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

inextensible and flexible string; he was able to show that the natural modes of the rotating vertical chain corresponding to small amplitudes resembled the zeroth-order Bessel function of the first kind, $J_0(x)$; this is, in fact, a simple and beautiful experimental demonstration of Bessel's function. Careful examination of his photographs indicates that beyond the second mode, the natural modes may not have been achieved in the experiment. Howell and Triantafyllou [16] have addressed the related problem of a hanging chain under in-plane harmonic excitation; through asymptotic analysis and numerical simulations they found that at large excitation amplitudes, stable as well as unstable steady-state solutions exist in both two and three dimensions. Coomer et al. [14] used flash photography to identify the number of nodes at any given frequency of rotation and indicated that the number of nodes varied even under the same rotational frequency suggesting that the string possesses multiple solutions at each rotating frequency. In this sense, the experimental research on whirling of strings has been qualitative in nature. We report the results of a careful, *quantitative* experimental evaluation of the problem of whirling of a string. A review of the formulation and solution of the equations of motion is provided in Sec. 2 to set the stage for our presentation. The experimental tasks of identifying an appropriate string that satisfies the assumptions of the theory, the assembly of the apparatus to provide appropriate control over the imposed twirling motion, and the determination of the spatial variation of the string during the experiment are described in detail in Sec. 3. In Sec. 4, we describe the experimental results and then provide a comparison to the linear and nonlinear theories of whirling. A number of interesting deviations from the steady-state solution assumed in the theoretical formulation are also described in this section.

2 Review of Theory of Whirling Strings

Consider a string of length l, free at one end and attached at the other end to the shaft of a rotating motor. The only internal force that the string is capable of sustaining is assumed to be the tension which is perpendicular to each cross section. The string is assumed to exhibit negligible resistance to bending; we will examine this more closely when the characteristics of the string are described. If the cross-section possesses two axes of symmetry, the problem of torsion is decoupled from the remaining equations and all that is required is that the twist per unit length of the string be constant [17]. This is the setting in which Kolodner considered the nonlinear problem of a rotating string; we will review this formulation in this section.

2.1 Equations of Motion. Let $\mathbf{r}(s,t)$ be the position vector of any point of the string; s is the arc length measured from the free end of the string and t represents time.

$$\mathbf{r}(s,t) = X_i \hat{\mathbf{n}}_i = x_i \hat{\mathbf{b}}_i \tag{1}$$

where $\hat{\mathbf{n}}_i (i=1,2,3)$ is the Newtonian reference frame, $\hat{\mathbf{b}}_i$ is the current frame associated with the rotating axis, and X_i and x_i are the components of $\mathbf{r}$ in the two frames, respectively. A sketch of the deformed shape of the string and the frames is shown in Fig. 1. The only forces acting on any segment of the string as it executes a rotational motion imposed on it about the axis $\hat{\mathbf{n}}_3$ are the body force on the string and the internal tension in the string. The body force per unit length is taken to be $\mathbf{f}(s,t)$. The tension in the string must always be along the tangent to the string and is represented as $\mathbf{T}(s,t)$. Therefore the tension can be written as:

$$\mathbf{T}(s,t) = T(s,t)\hat{\boldsymbol{\tau}}(s,t) = \frac{T(s,t)}{\nu(s,t)} \frac{\partial \mathbf{r}(s,t)}{\partial s} \tag{2}$$

where $\hat{\boldsymbol{\tau}}(s,t)$ is the unit vector tangent to the string, $T(s,t)$ is the magnitude of the tension, and ν is the stretch in string.

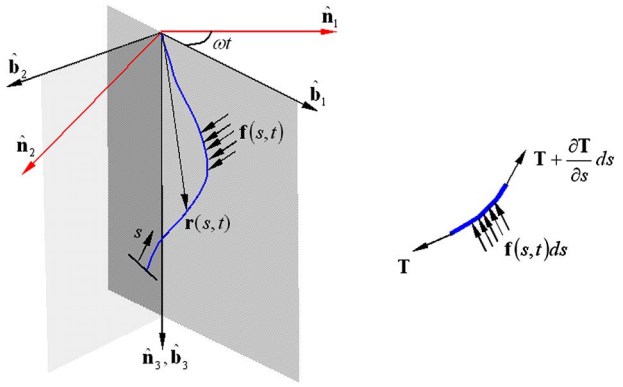


Fig. 1 Motion configuration of the string

$$\nu = \left| \frac{\partial \mathbf{r}(s,t)}{\partial s} \right| \tag{3}$$

We shall consider the string to be inextensible with the result that $\nu = 1$; while this inextensibility is not essential, it makes the problems simpler. Antman and Reeken [18] have examined the problem of whirling while drawing simultaneously. In the experiments described in the present work, the extension of the string was shown to be negligible through direct measurements.

Consider the free-body-diagram of a small segment of the string, as shown in Fig. 1. The general equations of motion for this small segment can be expressed as:

$$\frac{\partial \mathbf{T}(s,t)}{\partial s} + \mathbf{f}(s,t) = \rho \frac{\partial^2 \mathbf{r}(s,t)}{\partial t^2} \tag{4}$$

where ρ is the mass density per unit length of the string. The string is assumed to be set in motion in air, but the speed of motion is considered to be low enough to neglect air resistance; therefore, the only component of body force on the string is its own weight given by:

$$\mathbf{f} = \rho g \hat{\mathbf{n}}_3 = \rho g \hat{\mathbf{b}}_3 \tag{5}$$

where g is the acceleration due to gravity. Equation (4) can now be rewritten in component form in the $\hat{\mathbf{n}}$ frame:

$$\rho \ddot{X}_1 = (TX_1')'$$

$$\rho \ddot{X}_2 = (TX_2')'$$

$$\rho \ddot{X}_3 = \rho g + (TX_3')' \tag{6}$$

The inextensibility constraint is also written in terms of components as:

$$X_1'^2 + X_2'^2 + X_3'^2 = 1 \tag{7}$$

Equations (6) and (7) are the governing differential equations for the three-dimensional motion of the string.

2.2 Nonlinear Eigenvalue Problem. Kolodner [11] and Caughey [12] investigated steady-state solutions to the whirling problem posed above assuming that the following steady-state motion exists:

$$X_1(s,t) = x(s)\cos \omega t$$

$$X_2(s,t) = x(s)\sin \omega t$$

$$X_3(s,t) = z(s)$$

$$T(s,t) = T(s) \tag{8}$$

It is readily seen that Eq. (8) represents motion of the string in a plane, with the plane rotating about the $\hat{\mathbf{n}}_3$ axis at the frequency ω. $x(s)$ is the amplitude in this vertical plane. Introducing Eq. (8) into the governing differential equations of motion in Eqs. (6) and (7), we obtain:

$$(Tx')' + \rho\omega^2 x = 0$$

$$(Tz')' + \rho g = 0$$

$$x'^2 + z'^2 = 1 \tag{9}$$

The necessary boundary conditions are the following: the tension is zero at the free end $s=0$: $T(0)=0$, and the string is fixed at $s=l$: $x(l)=z(l)=0$. Through a change of variable $u(s)=Tx'/\rho g$, the above equations can be simplified as follows

$$u''(s) + \frac{\lambda}{l}\frac{u(s)}{\sqrt{u^2(s)+s^2}} = 0 \tag{10}$$

where $\lambda=\omega^2 l/g$. The boundary conditions can now be written in terms of the transformed variable u as follows:

$$u(0) = u'(l) = 0 \tag{11}$$

Equations (10) and (11) can be nondimensionalized with $\hat{s}=s/l$ and $\hat{u}=u/l$, to yield:

$$\frac{d^2\hat{u}}{d\hat{s}^2} + \lambda\frac{\hat{u}}{\sqrt{\hat{u}^2+\hat{s}^2}} = 0 \tag{12}$$

$$\hat{u}(0) = 0, \quad \frac{d\hat{u}(1)}{d\hat{s}} = 0 \tag{13}$$

Equation (12) together with Eq. (13) constitutes a nonlinear eigenvalue problem. It can also be treated as a nonlinear two-point boundary value problem for prescribed ω and solved numerically by using finite difference or shooting methods. Once we have the solution $\hat{u}(\hat{s})$, other components of the solution for such a steady-state whirling can be obtained. In particular,

$$T(s) = \rho g\sqrt{u^2(s)+s^2} \tag{14}$$

$$x(s) = -\frac{g}{\omega^2}u' = -\int_s^l \frac{u(\xi)}{\sqrt{u^2(\xi)+\xi^2}}\,d\xi \tag{15}$$

$$z(s) = \int_s^l \frac{\xi}{\sqrt{u^2(\xi)+\xi^2}}\,d\xi \tag{16}$$

We will first examine the linearized problem and then discuss the nonlinear problem.

2.3 Small Amplitude Whirling. For $\hat{u} \ll \hat{s}$, the above nonlinear eigenvalue problem can be reduced to a linear eigenvalue problem:

$$\hat{s}\frac{d^2\hat{u}}{d\hat{s}^2} + \lambda\hat{u} = 0 \tag{17}$$

Equation (17) is a Bessel equation. The solution to the linearized problem can be written as:

$$\hat{u}(\hat{s}) = \sqrt{\hat{s}}AJ_1(2\sqrt{\lambda\hat{s}}) + \sqrt{\hat{s}}BY_1(2\sqrt{\lambda\hat{s}}) \tag{18}$$

where J_1 and Y_1 are the Bessel functions of the first and second kinds of order one. From the first of the boundary conditions in Eq. (13), we see that the second term must be discarded since $Y_1(\hat{s})\to\infty$ as $\hat{s}\to 0^+$. Also, A has nontrivial values only at discrete whirling frequencies (eigenfrequencies of the system) given by the zeros of J_1. The general solution can then be written as the superposition of the eigenmodes corresponding to the eigenfrequencies:

$$\hat{u}(\hat{s}) = \sum_{k=1}^{\infty} \sqrt{\hat{s}}A_k J_1(2\sqrt{\lambda_k\hat{s}}) \tag{19}$$

where A_k are the amplitudes to be determined and $J_0(2\sqrt{\lambda_k})=0$ provides the roots λ_k. Finally, the solution for linearized problem for the fundamental mode $k=1$ can be written in terms of the string coordinates as:

$$z(s) \approx l - s$$

$$x(s) = -\frac{g}{\omega^2}u' \approx -\frac{lA}{\sqrt{\lambda_1}}J_0\left(2\sqrt{\lambda_1\frac{s}{l}}\right) \tag{20}$$

The main upshot of this analysis is that *the string can whirl at steady state only at fixed frequencies*: $\omega_k=\sqrt{\lambda_k g/l}$. Casual observation of string twirling, however, suggests that the string can indeed whirl at any frequency; this observation prompted Kolodner first, and others later, to look for steady-state solutions from the complete equation of motion in Eqs. (12) and (13).

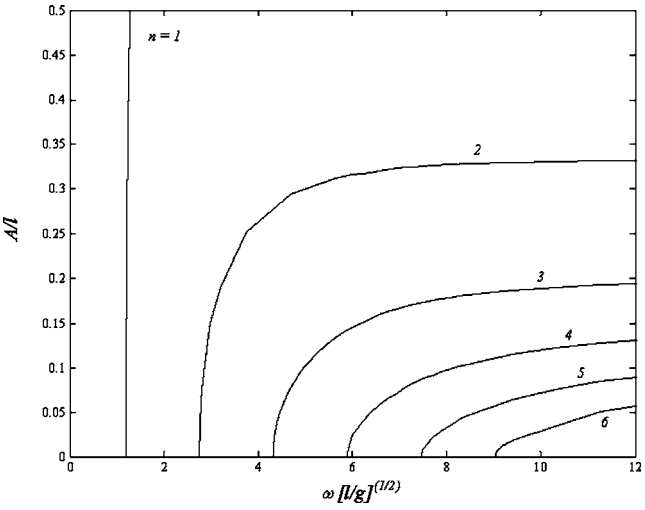


Fig. 2 Variation of the tip amplitude with frequency of whirling for the different modes. The first mode has a horizontal asymptote at $A/l\sim 1$ and hence is not shown completely.

2.4 Large Amplitude Whirling. When $\hat{u}$ is not small, the nonlinear eigenvalue problem in Eqs. (12) and (13) must be solved. Kolodner [11] showed that whenever $\omega_k<\omega<\omega_{k+1}$, steady-state whirling can occur at any of the k eigenmodes. Caughey [13] examined this problem thoroughly, but with the end $s=0$ constrained to remain along the initial vertical line; this requires a nonzero tension to appear at this end. Caughey [12] showed that steady whirling cannot occur for $\omega<\omega_1$, that unique periodic solutions exist for arbitrary rotational speed, and that the solutions are stable. He also obtained closed-form solutions to the nonlinear problem in terms of Jacobi elliptic functions. Very recently, Coomer et al. [14] showed that the two-point boundary value problem in Eqs. (12) and (13) could be solved using the *shooting method* implemented in MAPLE. We determined the solution to the whirling problem in MATLAB. In order to obtain these solutions to the nonlinear problem, we start the numerical solution procedure with different initial guesses for the shape of the string; in particular the mode shape of the linearized eigenvalue problem can be used as an initial guess. For a prescribed initial tip amplitude, the program searches for the eigenvalue at which the solution is obtained. In Fig. 2, the amplitude at the tip of the string A corresponding to each frequency ω is shown. The corresponding

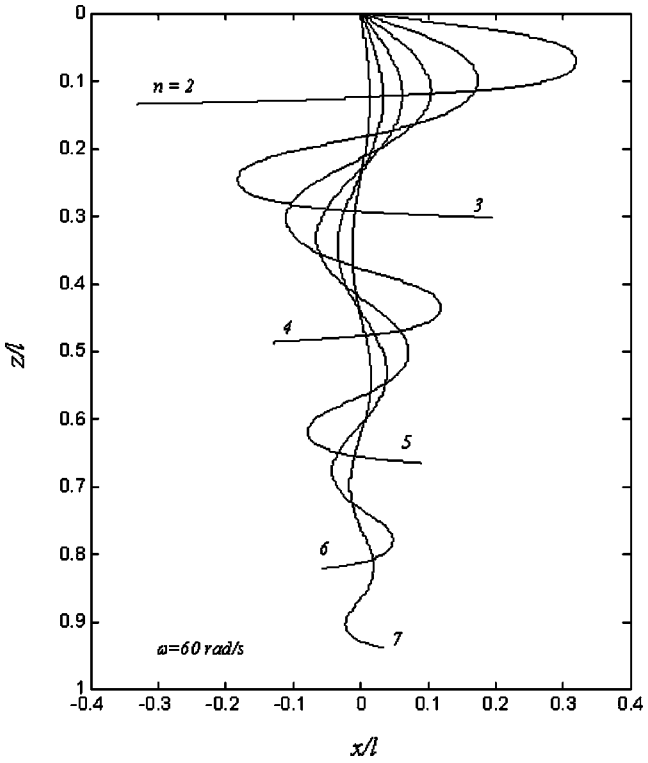


Fig. 3 Mode shapes determined from the nonlinear analysis for modes two to seven; mode shapes are plotted corresponding to a frequency of 60 rad/s. The first mode appears nearly horizontal and hence is not shown in this figure.

mode shapes for each frequency are shown in Fig. 3. Clearly, for any frequency $\omega_k < \omega < \omega_{k+1}$, there are k mode shapes and corresponding amplitudes. The variation of the tension in the string along its length normalized by the weight of the string $\rho g l$ is shown in Fig. 4. The tension takes on its maximum value, significantly larger than the static value for the lower modes, at the nodes and is equal to the static value at points with a vertical tangent. The main result of this exercise is that for the nonlinear problem, *at each frequency of whirling, there are a number of discrete modes at which steady-state whirling can occur*. The amplitude of whirling is determined by the mode. In a reasonably

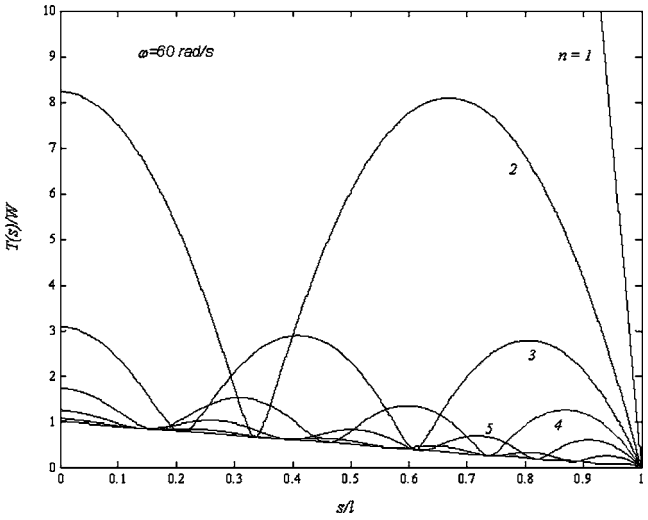


Fig. 4 Variation in the tension (normalized by the weight) in the string along its length for different modes

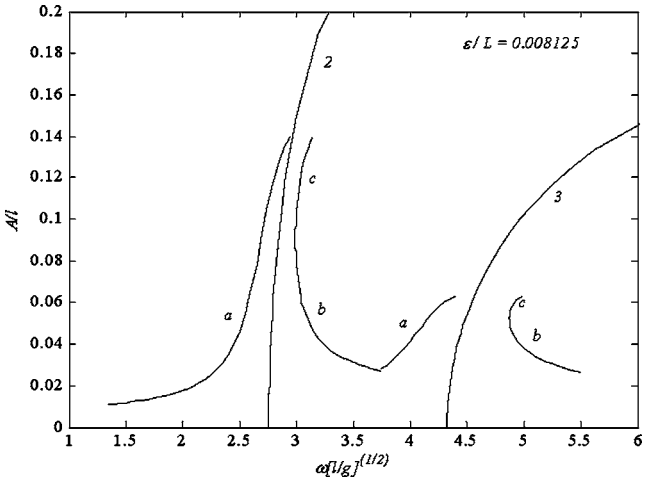


Fig. 5 Variation in the amplitude with the frequency for forced motion of the string near the second and third natural frequencies. The stable branches may be identified as in the simple case of the Duffing equation.

exhaustive search into the literature, we have not been able to find quantitative observations on the whirling motion of strings. Coomer et al. [14] point out that the generation of pure whirling modes is a nontrivial task.

2.5 Forced Whirling. Caughey [12] considered approximate nonlinear solutions for the forced whirling of the string; he further examined the stability of such motions. We summarize his results for use in interpretation of our experimental results. Consider that the rotational motion is imposed at the upper end of the string by fixing it at a distance ε from the axis of rotation. Then, the solution to the forced nonlinear problem near any natural frequency ω_i can be written as

$$x(s) = \varepsilon + A_i J_0(2\sqrt{\lambda_i s/l}) \tag{21}$$

Typically, a summation of the normal modes would be used to represent the shape under forced oscillations, but since the resulting nonlinear equations are quite difficult to solve, Caughey suggested considering only the ith mode when the driving frequency is in the neighborhood of ω_i; this places a restriction on the magnitude A_i, but this is not of major consequence here. Eliminating z between the second and third of Eqs. (9), the tension in the string may then be written as:

$$T(s) = \frac{\rho g s}{\sqrt{1-[x'(s)]^2}} = \frac{\rho g s}{\sqrt{1-A_i^2[J_0'(2\sqrt{\lambda_i s/l})]^2}} \tag{22}$$

where the prime denotes differentiation with respect to s. In order to determine the amplitude A_i, we first substitute the above into the differential equation, then multiply by $J_0(2\sqrt{\lambda_i s/l})$ and integrate with respect to s over $[0,l]$. The first term of the resulting equation is integrated by parts, and after substitution of limits and a change of variable to $\xi = \sqrt{s/l}$, we get

$$\frac{A_i}{l}\left[\int_0^1 \xi J_0^2(2\sqrt{\lambda_i}\xi)d\xi - \frac{\omega_i^2}{\omega^2}\int_0^1 \left(\frac{\xi J_1^2(2\sqrt{\lambda_i}\xi)}{\sqrt{1-(A_i^2\lambda_i/l^2)[J_1(2\sqrt{\lambda_i}\xi)/\xi]^2}}\right)d\xi\right] + \frac{\varepsilon}{l}\int_0^1 \xi J_0(2\sqrt{\lambda_i}\xi)\,d\xi = 0 \tag{23}$$

which is a nonlinear equation for A_i; the first and third terms can be integrated explicitly and then Eq. (23) can be solved numeri-

Table 1 Material and structural properties of steel, nylon, and rubber

Material		Steel	Nylon	Rubber
Density, ρ	kg/m^3	7800	1100	1130
Modulus of elasticity, E	GPa	200	2.8	0.0036
String diameter, d	m	0.00060	0.00060	0.00060
String length, l	m	0.40	0.40	0.40
Bending rigidity, EI	N m^2	1.27E−03	1.78E−05	2.29E−08
Axial stiffness, EA	N	5.65E+04	7.92E+02	1.02
Axial wave speed, $c=\sqrt{(E/\rho)}$	m/s	5064	1595	56
Whirling resonance	rad/s	5.95, 13.66, 21.42, 29.18, 36.95		

cally for A_i at each λ_i. This solution is shown in Fig. 5 for frequencies in the neighborhood of the second and third natural frequencies. This solution is not unlike the simple nonlinear response of the Duffing equation and hence one would expect similarities such as unstable branches and hysteretic behavior. Caughey [12] investigated the stability of these solution branches and found that *a* and *b* are stable while *c* was unstable. Upon increasing the frequency, the amplitude will follow branch *a* and at some level of amplitude—dictated by inherent damping—jump to branch *b*, while upon decreasing the frequency, the amplitude will follow the branch *b* until some frequency and then jump to branch *a*. Of course, this applies only for small amplitude whirling due to the approximate nature of the solution used in Eq. (21).

3 Experimental Setup

There are three main ingredients in bringing to realization the whirling string problem. The first ingredient is the identification of a string with appropriate mechanical condition that can transmit the motion imposed at one end to the rest of the string but can rightly be considered an inextensible and flexible string. The second is the ability to excite a pure whirling motion in the string. This requires a speed-controlled motor with a high degree of control as well as a proper balancing of the rotating parts in order to eliminate any spurious driving forces in other directions. The third is the ability to obtain quantitative measurements of the amplitude of the motion. We describe the details of the experimental arrangement in this section.

3.1 String. It is unfortunate that the terminology "heavy chain" was used in the early papers. It might seem that a chain would be ideal in the sense that it can exhibit the required torsional stiffness in order to transmit the applied torque but at the same time exhibit almost negligible bending stiffness. However, a chain (or even a twined or braided thread or string) introduces hysteretic behavior in the torque-twist response that is not quite suitable for this particular demonstration. Western [15] was able to demonstrate some elements of this problem with a necklace chain only because he was in the region of small amplitude whirling as will become clear from our experimental results. Furthermore, the string does not have to be "heavy;" the mass density of the string must be nonzero in order to provide the gravitational and inertial forces assumed in the model; other than that, it plays no role whatsoever in determining the motion of the string. In our quest for the ideal "string," we examined tightly braided cotton and nylon threads, twines, nylon fishing line, cosmetic jewelry chains, piano wire, rubber cord extracted from elastic waist bands, etc. Material and structural properties are shown in Table 1 for steel, nylon, and rubber. The natural frequencies of the linearized whirling modes are also indicated in this table. For these thin strings, the bending stiffness is very small. Comparison of the bending energy to the stretching energy requires a length scale. If we take the length scale to be about $l/10$, such that there are 10 bends over the length of the string, then we can estimate that the ratio of bending energy to stretching energy is roughly 10^{-4}. Thus, for a first approximation it appears that we may ignore bending contributions for any of the materials considered. The experiments reported here were performed with a rubber string with properties shown in Table 1.

3.2 Whirling Apparatus. In order to provide the proper boundary conditions for the whirling string, a 12 V dc motor (Pittman Express Model 8322S001) with an encoder and a speed controller (JR Kerr Automation Engineering PIC-SERVO board) were used. This motor provides a continuous torque of 0.011 Nm at speeds below 800 rad/s. The encoder provides 500 counts per revolution in order to enable the speed controller to maintain a constant speed. The encoder on the motor and the controller together are capable of maintaining the speed error to within 15 parts per million. The motor was mounted on a stiff frame on top of a vibration-isolated optical table. A precision-machined sleeve was press fit on the shaft of the motor to ensure proper balance of the rotating mass. The string was inserted into one end of the sleeve through a small hole and fixed in place through two setscrews. The other end of the string was left freely hanging, thereby replicating the boundary conditions assumed in the problem formulation. This arrangement provided good alignment of the loading; the string was provided with a rotational motion along its axis with the magnitude of the lateral motions below the sensitivity of our measuring instruments. The entire assembly was placed in an enclosure in order to eliminate influences from air currents in the room; it was fitted with a view port on one side to enable observations.

3.3 Quantitative Imaging. Quantitative measurement of the string while whirling was obtained using video imaging systems. Three different schemes were used in these experiments. In the first arrangement, in order to obtain the mode shapes, a video camera was used in conjunction with a Frame Grabber board in a *time-averaging* mode. A LabVIEW program was written to acquire images with controllable exposure times, under continuous illumination. The camera was aimed so as to obtain a projection in the X_1-X_3 plane. Considering that the first five natural frequencies of whirling corresponding to the linearized problem are below 6.3 Hz, exposure over 0.25 s assures us of capturing many revolutions of the string; thus, this long time exposure should indicate the mode shape in the x-z coordinate system. In the second arrangement, the string was observed through a mirror mounted at 45 deg to the X_3 axis to obtain a projection of the tip point of the string in the X_1-X_2 plane. Multi-flash images of the tip were obtained in many sequences at different frequencies of whirling and used to identify the trajectory of the tip point.

In the third arrangement, computer stereo vision techniques, well developed in the field of artificial intelligence [19,20] were adapted for measurement of $\mathbf{r}(s,t)$ in a *time-resolved* mode. Such a method has recently been used by Helm et al. [21] to measure the three-dimensional shape and displacements of deformable objects. The principle of this method is quite simple: consider an object point (X_1,X_2,X_3) in three-dimensional space. This point is viewed through two cameras and the corresponding points in the left and right image planes are identified by a digital image correlation algorithm; thus, we get the following projections of the object point: (x_L,y_L) and (x_R,y_R) in the coordinate system of the two cameras. (x_L,y_L) and (x_R,y_R) can be used to reconstruct the object point (X_1,X_2,X_3) based on a model for the camera. Calibration of the system for intrinsic parameters of the camera such as focal lengths, principal point, and correction for lens distortion errors and extrinsic parameters of the geometry of the optical arrangement of the two cameras relative to the object (rotation angles and translation vector of two three-dimensional coordinate systems) needs to be performed. This calibration is accomplished by obtaining photographs of a precise checkerboard pattern in multiple orientations and identifying the corner points of the pattern through a nonlinear least square technique to minimize the

following error function to get the optimum parameters:

$$\text{Error} = \sum_{\text{exp points}} [(x_c - x_{c\_\text{exp}})^2 + (y_c - y_{c\_\text{exp}})^2]^{1/2} \qquad (24)$$

where (x_c, y_c) are the camera coordinates from the coordinates (X_1, X_2, X_3) of each corner of the checkerboard pattern determined by using geometric projection. $(x_{c\_\text{exp}}, y_{c\_\text{exp}})$ are measured from the captured images of the corner of the checkerboard pattern. Details of these procedures can be found in Tsai [22]. Here, we used the calibration toolbox based on MATLAB developed by Bouguet [23]. With our optical system, we were able to obtain a position accuracy of less than 1 mm in all three directions.

In our experiments, two Sony XC-55 monochrome progressive scan cameras, two National Instruments' IMAQ 1409 frame grabbing boards, and the image acquisition software based on LabVIEW 7.0 were used to acquire the motion trajectories of the string simultaneously. The progressive scan cameras allow a scan reset that when combined with the IMAQ 1409 controller function enables the acquisition of images from both cameras at the same time. A flash illumination was synchronized with the image capture in order to freeze the motion of the string in the image. It was also possible to acquire the image with a long time exposure, but with the string illuminated by a stroboscope such that multiple images of the string were obtained in the same frame. Then a three-dimensional algorithm was used to reconstruct the three-dimensional shape and trajectories of the rotating string.

4 Whirling Response of a Rubber String

We now describe the experimental observations. A rubber string (length $l=0.40$ m, diameter $d=0.60$ mm, density $\rho=1130$ kg/m^3, and modulus of elasticity under small stretch $E=3.6$ MPa) was attached to the supporting fixture at the base of the motor as described in Sec. 3. The motor was set in steady-state rotation at progressively increasing frequencies. At each frequency the transients were allowed to dissipate and then the whirling mode shape was captured through the time-averaging video imaging described in Sec. 3.3. In contrast to the solutions to the nonlinear problem discussed in Sec. 2.4, it was possible to generate steady-state whirling motion only at specific frequencies, while nonsteady motion was observed at other frequencies. The steady-state behavior is evaluated first; this is followed by an evaluation of the nonsteady response of the string to periodic excitation. Although only results corresponding to the rubber string are presented here, these observations remain valid regardless of the composition of the string.

4.1 Steady-State Response. In order to eliminate possible hysteretic effects of dependence on initial conditions, at each frequency, the whirling motion was diminished by moving two fingers on either side of the string and then allowed to rebuild naturally. In contrast to the solution discussed above, steady motion of the type assumed in Eq. (8) was observed only at specific frequencies. In Fig. 6, the experimentally obtained mode shapes corresponding to steady state are compared to the Bessel function mode shapes in Eq. (19) corresponding to the linearized eigenvalue problem (dashed lines). The mode $N=1$ exhibited a very small displacement amplitude and hence resonance could not be observed easily. As can be seen from the experimental observations, for the lowest mode shown ($N=2$), the experimentally observed mode shape is close to the Bessel function. But for higher modes ($N=3-5$), the deviation from the linearized theory is apparent. Clearly, we must resort to the nonlinear analysis. Western [15] experimentally showed mode shapes from a necklace chain; from his measurements he obtained the positions of the nodes and showed that they were in good agreement with the nodes identified by the linearized solution. As can be seen from Fig. 6, even though the nodes are almost at the same locations as the Bessel function indicates, the disagreement in the amplitude is quite substantial.

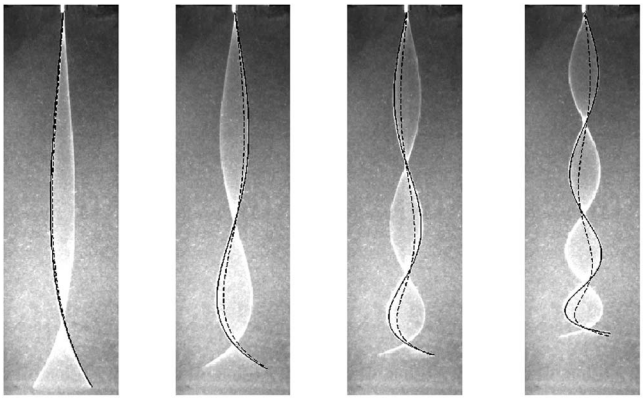


Fig. 6 Time-averaged images of the rubber string corresponding to four steady-state modes are compared with the predictions of the linearized theory (dashed lines) and the nonlinear theory (solid lines)

The nonlinear problem was solved for the parameters of the rubber string in MATLAB as indicated in Sec. 2.4. As described there, Eqs. (12) and (13) possess multiple solutions at any frequency. So, two options are available: either find the whirling mode shape appropriate to the experimentally imposed frequency of rotation or find the appropriate frequency corresponding to a matched mode shape. The first option was discarded since the amplitude determined from Eqs. (12) and (13) at the imposed frequency did not match the experimentally observed amplitudes. On the other hand, searching for an appropriate frequency at which the amplitude of the whirling matched the observed shapes resulted in good agreement as shown by the solid lines in Fig. 6. The frequencies at which the nonlinear solution matched the measured resonance frequencies are given in Table 2; these differ from the experimentally measured resonance frequencies by an increasing amount with increasing mode number. The influence of bending in the rubber string appears to be the primary reason for this discrepancy; this will be explored further in a future contribution.

4.2 Nonsteady Motion of the String. We now turn to an examination of the nonsteady motion of the whirling string. At any frequency between those where steady response was observed, the string exhibited whirling motion with a dominant mode shape very similar to that obtained at the steady states discussed above, but an additional motion in both the radial and axial directions appeared superposed. Selected time-averaged images of the motion are shown in Fig. 7. We note that the images were taken with a constant exposure time and hence, at the higher rotational speeds, more revolutions of the string were captured. From the images in this figure, it is clear that two features of the steady-state solution are violated. First, the nodal points move with time and, second, the vertical position of the string is not time independent as assumed in formulating the steady-state solution. The motion still appears to be repeatable, but with two dif-

Table 2 Comparison of resonance frequencies (ω_n rad/s)

Mode n	Eigenvalues of the linearized problem	Frequency of matching nonlinear solution	Experiment	% difference between nonlinear theory and experiment
2	13.67	13.94	14.47	3.80
3	21.43	23.60	24.91	5.55
4	29.20	34.20	36.89	7.87
5	36.97	47.50	52.22	9.94

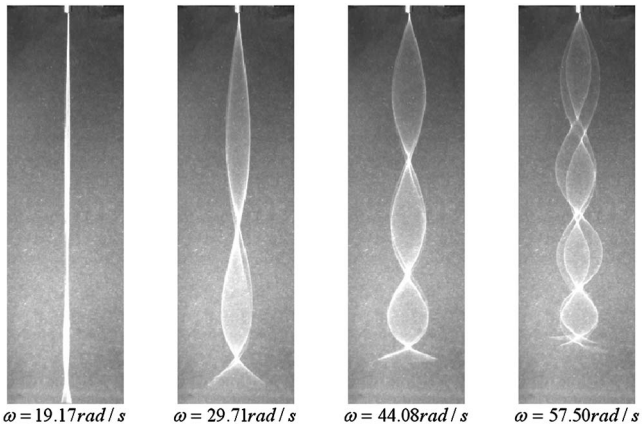


Fig. 7 Time-averaged images of the rubber string corresponding to frequencies in between the different steady modes

ferent periods: the first period is associated with the rotational motion input through the motor and dominated by the mode shapes found earlier. The second periodic motion is associated with variations in the axial motion.

In an effort to quantify the nonsteady motion of the string, the video imaging system was reoriented to the second arrangement discussed in Sec. 3.3. The entire string was painted with black ink, leaving only the tip point visible in the video imaging system. Images of the tip at different frequencies of whirling were then obtained in this arrangement. During each exposure of the camera, the light was pulsed N times (typically ten times) resulting in N distinct images of the tip point. These images were then filtered to eliminate all other points, leaving only the N images of the tip of the string. This imaging scheme was repeated M times (typically 250 images) and all M images were then superposed to obtain a map of the tip in the X_1-X_2 plane; thus, each composite image contains $M \times N$ discrete images of the string tip. Such images, captured under increasing frequencies spanning from $n=2$ to $n=5$, are shown in Fig. 8. A number of observations can be made from these measurements.

- Corresponding to steady-state motion, the locus of the tip points projected in the X_1-X_2 plane is nearly circular; the width of this circle is an indication of the extent of variation of $x(0)$. The small width over which the tip points are dispersed in some cases indicates that $x(0)$ is indeed constant and the motion assumed in Eq. (8) is the one observed. The motion is periodic and the circular orbit is executed at the period of the imposed rotation.
- In other cases, the locus of the tip points projected in the X_1-X_2 plane lies within a circular annulus, but with the extent of this annulus varying significantly with frequency. Thus $X_1(s,t)=x(s)\cos \omega t$ is no longer an appropriate description of the motion. Note also that $X_3(s,t)=z(s)$ is not adequate since time dependence in the axial motion is essential. However, this information is lost in the projection obtained; quantitative stereo-imaging can be used to reveal the nature of this motion as we examine in the next section.
- As the whirling speed is increased from one natural frequency towards another, spectacular patterns are formed by the tip point; this is most evident in the next to the last image in Fig. 8, but if fewer revolutions of the string are captured in the imaging process, patterns become evident at other whirling frequencies as well; this is illustrated through the observations in Fig. 9. Such patterns are a clear indication of multiple periodicities in the phenomenon observed. In particular, Wooley [24] observed patterns in the motion of the flagella of eel sperm that are identical to that shown in Fig. 9(*a*) and called these *flagelloid* curves; while there is a geometric similarity in the patterns, the underlying physical connections need to be developed more carefully; we will explore this in a future contribution.

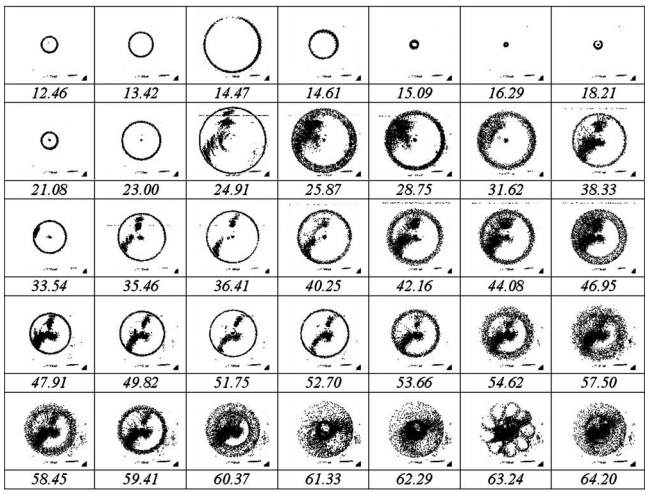


Fig. 8 Locus of motion of the tip point of the string. The dark points in the interior of the annulus are obtained from other points along the length of the string that could not be eliminated in the optical arrangement. (ω: rad/s).

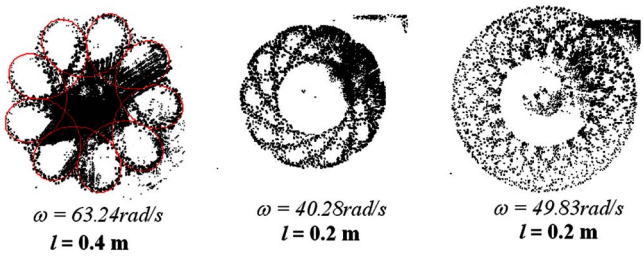


Fig. 9 Patterns formed by the tip point at different rotational frequencies. Solid line in the first image is drawn with $\omega_m/\omega=8$ (see Eq. (26)).

From the images shown in Fig. 8, the amplitude excursion of $X_1(s,t)$ clearly depends nonmonotonically on the frequency of rotation; this is developed further through a plot of the amplitude of the tip $X_1(0,t)$, determined as a function of the frequency of rotation in Fig. 10. In this figure, the frequency of the imposed rotation is normalized by the characteristic frequency $\sqrt{g/l}$. The dotted lines indicate the steady-state solution to the nonlinear eigenvalue problem corresponding to different modes. The solid line indicates measured amplitude excursion at different frequencies as the frequency is increased from the second eigenfrequency; these are lines connecting measurements of the maximum and minimum amplitudes determined from the projection of $X_1(0,t)$ in the X_1-X_2 plane. The dash-dot lines are the corresponding measurements obtained as the frequency is continuously decreased. These measurements reveal the complex nonlinear response of the string to the whirling motion. The main observations are listed below:

- The points R_2–R_5 marked on Fig. 10 correspond to observed steady-state behavior shown in Fig. 6. These appear to lie close to the steady-state solution to the nonlinear problem.
- Beginning at the second mode at R_2, if the frequency of the motor is increased, rather than follow the solution corresponding to the mode $n=2$, the string whirls at a mode that

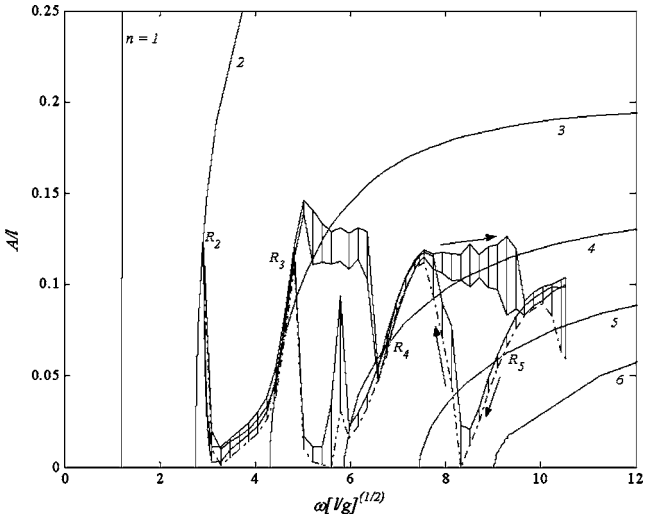


Fig. 10 Variation of the amplitude of the tip point as a function of the driving frequency

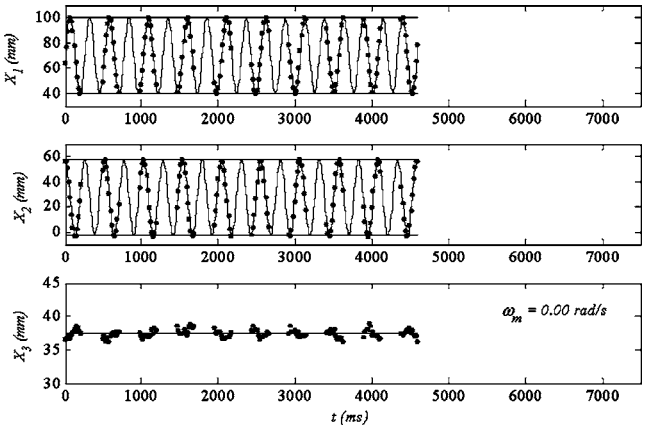


Fig. 11 Time variation of the position of the tip at a driving frequency of 24.91 rad/s

is close to this shape; the amplitude of whirling appears to correspond well with the approximate solution of the forced whirling problem.

- As the driving frequency is increased close to the range where the third mode solution to the nonlinear problem is available, the amplitude increases quickly until steady motion is observed at R_3. At each frequency below R_3, there is still a periodic fluctuation of the amplitude between the vertical lines indicated in Fig. 10, although the magnitude of the fluctuation is small.
- Upon decrease of the frequency from R_3 to R_2, the frequency response follows the same branch and jumps back to the second mode resonance at R_2.
- As the driving frequency is increased above R_3, the string is unable to follow the nonlinear solution corresponding to the branch $n=3$. This is similar to the observation at the second resonance; however, rather than the low amplitude whirling that was observed immediately beyond R_2, now the string maintains the mode shape of $n=3$ with a substantial amplitude together with a superposed time-dependent variation. The amplitude decreases for a while, but is then followed by a gradual increase with the process repeating itself periodically, with a period much larger than the whirling period; such amplitude excursions are marked in Fig. 10 by the vertical lines and their envelope. We refer to this mode as the *trapped mode*; detailed evaluation of the trapped mode is provided in the next section.
- When the frequency is increased to the range where the next mode shape and steady-state nonlinear solution become available, an abrupt transition occurs as the $n=4$ mode shape is taken with a smaller amplitude. Decreasing the frequency from this range towards R_3 results in the string maintaining the $n=4$ mode, but with significantly lower amplitudes; attempted jumps and a successful jump back to the $n=3$ mode are also seen clearly in Fig. 10. This hysteretic behavior is similar to the response of nonlinear systems such as the Duffing spring and the forced response indicated by Caughey (Sec. 2.5).
- Increasing the driving frequency beyond R_4, a similar hysteretic behavior is observed once more. We anticipate that this nonlinear behavior would persist at higher modes, although not all higher modes may be achievable with any given string. In this particular string, with a length of 0.40 m, modes $n>6$ were not attainable even at frequency as large as 65 rad/s.
- If transverse deflection of the string whirling in a trapped mode is diminished by slowly running a cupped palm along the length of the string, the trapped whirling motion is reestablished, suggesting that it is indeed the stable motion for the string.
- Overall, a significant departure from the steady-state solution of the nonlinear equations is observed, at small whirling amplitudes; we believe that this is due to the small bending stiffness of the string (and hence important at higher modes); we will pursue this quantitatively in a future contribution.
- Considering the overall response shown in Fig. 10, it is clear that the peak amplitude of the whirling motion is limited at $0.1l$. The reason for this limit in the whirling amplitude is not clear and needs further investigation; however, it was determined through repeated measurements that the magnitude of the peak amplitude depends on the length of the string.
- We conclude that while steady-state whirling modes are available, the string would not attain such a mode, but be trapped in a mode between two such steady-state solutions.

We now turn to a discussion of the trapped mode of the whirling string.

4.3 Trapped Modes of the String. In an effort to understand the dynamics of the trapped mode response of the string, the stereo-vision system described in Sec. 3.3 was used to obtain the

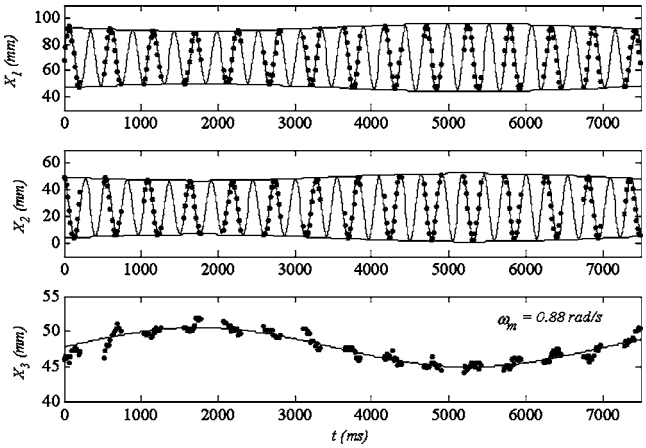


Fig. 12 Time variation of the position of the tip at a driving frequency of 26.83 rad/s

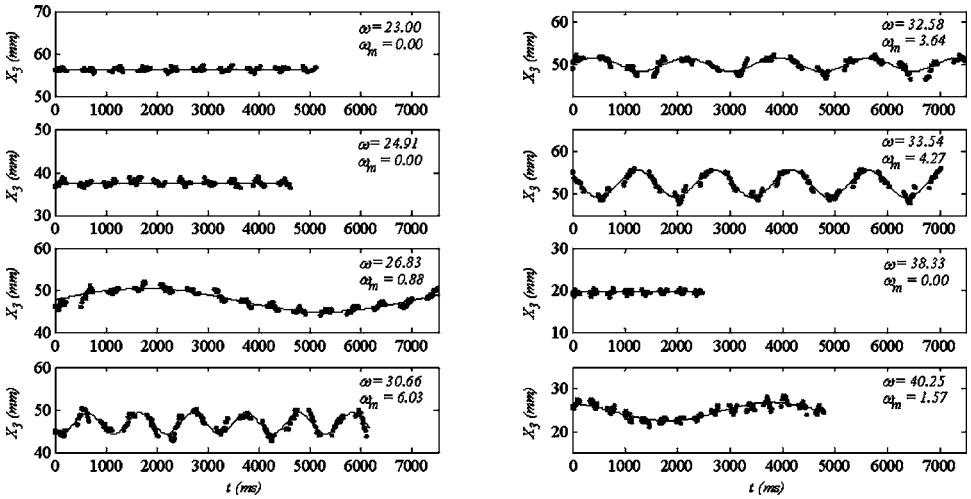


Fig. 13 Time variation of the X_3 position of the tip at different driving frequencies

time variation of the spatial position of the tip point. Multiple stereo pairs of images were obtained over many periods of rotation of the string and then processed to determine $X_1(0,t)$, $X_2(0,t)$, and $X_3(0,t)$. This was then repeated at different whirling frequencies spanning the range from just before the third eigenfrequency to just beyond the fourth eigenfrequency. The variation of $X_1(0,t)$, $X_2(0,t)$, and $X_3(0,t)$ is shown in Figs. 11 and 12 and for two different whirling frequencies; in these figures, the discrete points are samples obtained from the stereo-vision measurements. Figure 11 corresponds to whirling motion at R_3 and provides a clear demonstration of the steady-state motion as well as outlines the sensitivity of the measuring scheme; at resonance, the string exhibits no measurable time dependence in the X_3 direction and the motion in the X_1-X_2 directions is sinusoidal. We recognize that there are higher frequency components to the X_3 motion, but these have not been fully resolved by the measuring system. On the other hand, beyond R_3, when the string enters the trapped modes, the motion in the X_3 direction exhibits a clear sinusoidal motion as shown in Fig. 12; furthermore, the motion in the X_1 -X_2 plane is modulated with the same period (as dictated by the inextensibility condition); in this figure, the data points are samples obtained from the stereo-vision measurements while the solid line corresponds to a modulated sinusoidal motion:

$$X_1(0,t) = x(0)(1 + \varepsilon \cos \omega_m t)\cos \omega t$$

$$X_2(0,t) = x(0)(1 + \varepsilon \cos \omega_m t)\sin \omega t$$

$$X_3(0,t) = z(0)(1 - \varepsilon \cos \omega_m t) \tag{25}$$

With a further increase in the frequency, the motion in the X_3 direction is still periodic, but an increase in the modulation frequency. The time dependence of the motion in the X_3 direction was traced at a number of frequencies from just below R_3 to just above R_4; this is shown in Fig. 13. At R_3, the string maintains a time-independent position $X_3(0,t)=z(0)$ to within the resolution of the measuring scheme ±1 mm. With increasing frequency, a periodic motion is observed: $X_3(0,t)=z(0)f(t)$; this periodic oscillation is sinusoidal when the whirling frequency is close to a resonant frequency, but becomes increasingly distorted and asymmetrical as the whirling frequency is increased away from resonance. The emergence of a second time scale can be observed clearly from the results shown in Fig. 13; this time scale must relate to an interaction between the rotational motion and the axial wave motion that is neglected in the steady-state analysis based on Eq. (8).

The tip motion observed in the experiment suggests that the overall motion of the string could be assumed to be of the following form (in the fixed frame):

$$X_1(s,t) = x_s(s)\cos \omega t + \varepsilon x_m(s) f(t) \cos \omega_m t$$

$$X_2(s,t) = x_s(s)\sin \omega t + \varepsilon y_m(s) f(t) \sin \omega_m t$$

$$X_3(s,t) = z_s(s) - \varepsilon z_m(s) f(t)$$

$$T(s,t) = T_s(s) + \varepsilon T_m(s) f(t) \tag{26}$$

where the subscript s indicates the solution corresponding to the steady state and the subscript m indicates the modulation. An example of fitting the motion in Eq. (25) to the whirling of the string is demonstrated in Fig. 9 with $\omega=1$ and $\omega_m=8$. The equations of motion for the modulation are similar to the original equations of motion in Eq. (6) and unless the modulation is assumed to be small—clearly not justified in light of the results shown in Fig. 13—no reduction in the system of governing equations is obtained. We have yet to obtain the trapped mode solutions analytically or numerically and will address this in a future contribution.

5 Conclusion

That steady-state solutions exist to the problem of a whirling string under gravitational loading has been known for over half a century; however, except for anecdotal evidence and some brief accounts, to our knowledge, careful experimental assessment of the whirling of strings has not been presented in the literature; in this paper, we present such a detailed examination of this problem. Rubber strings are used as models for the whirling string; while the string is elastic, accommodating changes in tension along its length, the total extension was found to be small enough to be negligible. Furthermore, the bending stiffness of the rubber string is also negligibly small. Thus, the strings are assumed to be inextensible and flexible; the resulting equations of motion for the string are geometrically nonlinear. Kolodner [11] showed that multiple steady-state whirling solutions with a constant axial position exist at any whirling frequency beyond the first resonance of the linearized problem. In our experiments, we found three major departures from this result. First, steady-state motions without time variation in the axial motion are observed only at specific frequencies. Second, while multiple steady-state solutions are available at any frequency, the string settles into steady-state trapped modes that exhibit a significant cyclic axial motion; such

trapped modes have been observed and characterized quantitatively for the first time in the present paper. The trapped modes provide interesting time evolution of the mode shapes of the whirling string; in particular, the trajectories in space of the free end of the string present exciting similarities to the trajectories observed in the motion of flagella of sperm, suggesting possibilities for exploring such motions in the string. Finally, the maximum amplitude that the string may attain is limited to a level that depends on the string length. All these discrepancies are generated by axial motions that arise in the string and need to be evaluated further.

Acknowledgment

We are grateful to a referee who brought to our attention the recent works of Clark et al. [25], Lemon and Fraser [26], and Zhu and Rahn [27] that explore the stability of whirling motions; in particular, the analyses in these references have identified a number of unstable regions and bifurcations in the solutions to the problem of whirling in the presence of aerodynamic drag. We hope to pursue quantitative comparisons between their analysis and our experiments in a future contribution.

References

[1] Routh, E. J., 1905, *The Advanced Part of a Treatise on the Dynamics of a System of Rigid Bodies*, 6th ed., McMillan and Co., London, pp. 385–453.
[2] Carrier, G. F., 1945, "On the Non-linear Vibration Problem of the Elastic String," Q. Appl. Math., **3**, pp. 157–165.
[3] Harrison, H., 1948, "Plane and Circular Motion of a String," J. Acoust. Soc. Am., **20**, pp. 874–875.
[4] Oplinger, D. W., 1960, "Frequency Response of a Nonlinear Stretched String," J. Acoust. Soc. Am., **32**, pp. 1529–1538.
[5] Murthy, G. S. S., and Ramakrishna, B. S., 1965, "Nonlinear Character of Resonance in Stretched Strings," J. Acoust. Soc. Am., **38**, pp. 461–471.
[6] Narasimha, R., 1968, "Non-linear Vibration of an Elastic String," J. Sound Vib., **8**, pp. 134–146.
[7] Miles, J., 1984, "Resonant, Nonplanar Motion of a Stretched String," J. Acoust. Soc. Am., **75**, pp. 1505–1510.
[8] O'Reilly, O., and Holmes, P. J., 1992, "Non-linear, Non-Planar and Non-Periodic Vibrations of a String," J. Sound Vib., **153**, pp. 413–435.
[9] Nayfeh, S. A., Nayfeh, A. H., and Mook, D. T., 1995, "Nonlinear Response of a Taut String to Longitudinal and Transverse End Excitation," J. Vib. Control, **1**, pp. 203–207.
[10] Tufillaro, N. B., 1989, "Nonlinear and Chaotic String Vibrations," Am. J. Phys., **57**, pp. 408–414.
[11] Kolodner, I. I., 1955, "Heavy Rotating String—A Nonlinear Eigenvalue Problem," Commun. Pure Appl. Math., **3**, pp. 395–408.
[12] Caughey, T. K., 1958, "Whirling of a Heavy Chain," Proceedings of the 3rd U.S. National Congress of Applied Mechanics, pp. 101–108.
[13] Caughey, T. K., 1969, "Whirling of a Heavy String Under Constant Axial Tension: A Non-linear Eigenvalue Problem," Int. J. Non-Linear Mech., **4**, pp. 61–75.
[14] Coomer, J., Lazarus, M., Ticker, R. W., Kershaw, D., and Tegman, A., 2001, "A Non-linear Eigenvalue Problem Associated With Inextensible Whirling Strings," J. Sound Vib., **239**, pp. 969–982.
[15] Western, A. B., 1980, "Demonstration for Observing $J_0(x)$ on a Resonant Rotating Vertical Chain," Am. J. Phys., **48**, pp. 54–56.
[16] Howell, C. T., and Triantafyllou, M. S., 1993, "Stable and Unstable Nonlinear Resonant Response of Hanging Chains: Theory and Experiment," Proc. R. Soc. London, Ser. A, **440**, pp. 345–364.
[17] Love, A. E. H., 1927, *A Treatise on the Mathematical Theory of Elasticity*, 4th ed., Dover Publications, New York, p. 415.
[18] Antman, S. S., and Reeken, M., 1987, "The Drawing and Whirling of Strings: Singular Global Multiparameter Bifurcation Problems," SIAM J. Math. Anal., **18**, pp. 337–365.
[19] Sobel, I. E., 1974, "On Calibrating Computer Controlled Cameras for Perceiving 3D Scenes," Artif. Intell., **5**, pp. 185–198.
[20] Bouguet, J. Y., and Perona, P., 1999, "3D Photography Using Shadows in Dual-Space Geometry," Int. J. Comput. Vis., **35**, pp. 129–149.
[21] Helm, J. D., McNeil, S. R., and Sutton, M. A., 1996, "Improved Three-Dimensional Image Correlation for Surface Displacement Measurement," Opt. Eng., **35**, pp. 1911–1920.
[22] Tsai, R. Y., 1986, "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ, pp. 364–374.
[23] Bouguet, J. Y., 2001, http://www.vision.caltech.edu/bouguetj/calib_doc/index.html, accessed on 8/30/2003.
[24] Woolley, D. M., 1998, "Studies on the Eel Sperm Flagellum. 2. Kinematics of Normal Motility," Cell Motil. Cytoskeleton, **39**, pp. 233–245.
[25] Clark, J., Fraser, W. B., Rahn, C., and Rajamani, A., 2005, "Limit-Cycle Oscillations of a Heavy Whirling Cable Subject to Aerodynamic Drag," Proc. R. Soc. London, Ser. A, **461**, pp. 875–893.
[26] Lemon, G., and Fraser, W. B., 2001, "Steady-State Bifurcations and Dynamical Stability of a Heavy Whirling Cable Acted on by Aerodynamic Drag," Proc. R. Soc. London, Ser. A, **457**, pp. 1021–1041.
[27] Zhu, F., and Rahn, C. D., 1998, "Stability Analysis of a Circularly Towed Cable-Body System," J. Sound Vib., **217**, pp. 435–452.

Journal of Applied Mechanics

Technical Brief

A Comparative Evaluation of Three Isotropic, Two Property Failure Theories

Richard M. Christensen
Lawrence Livermore National Laboratory,
Livermore, CA 94551
and
Stanford University,
Stanford, CA 94305

Three fundamentally different failure theories for homogeneous and isotropic materials are examined in both the ductile and brittle ranges of behavior. All three theories are calibrated by just two independent failure properties. These three are the Coulomb-Mohr form, the Drucker-Prager form, and a recently derived theory involving a quadratic representation along with a fracture restriction. The three theories are given a detailed comparison and evaluation. The Coulomb-Mohr form and the Drucker-Prager form are found to predict physically unrealistic behavior in some important cases. The present form meets the consistency requirements. [DOI: 10.1115/1.2173007]

1 Introduction

The first formulation of a general failure criterion for homogeneous and isotropic materials was given by Coulomb [1] 232 years ago. Looking back, it was a prescient contribution, and it has remained a significant influence every since its conception. It was put into its present form 91 years ago by Mohr [2], and is now known as the Coulomb-Mohr Theory. The next most prominent and actively used or recognized general theory was suggested by Nadai [3] 55 years ago, and perhaps by others earlier, and put into its present form by Drucker and Prager [4] 53 years ago. It is somewhat surprising that these two long-standing and pre-eminent theories have not received a thorough comparison and joint evaluation. Such will be given here, along with a third theory that has been recently developed, but also has some historical lineage, as will be mentioned later.

The theories of materials failure mentioned above are calibrated by two independent parameters or properties. The present work is confined to this two-parameter level because this is the number of failure parameters that may admit interpretation as readily accessible physical properties. Certainly, the Coulomb-Mohr form was intended to admit such interpretations. One can view failure criteria as the means of technically specifying the limits of the elastic region of behavior for materials. Now linear, isotropic elastic materials require the knowledge of two moduli type properties for the characterization of the elastic region. This suggests that two more properties may also suffice to characterize the failure limit of the isotropic, elastic range. These two groups of two properties each offer a nontrivial correlation with the fact that general states of deformation are fully characterized by two states, those of dilatation and distortion. Taking the state of pure dilatation as allowing a positive volume change up to failure, but no failure under hydrostatic compression for homogeneous and isotropic materials, and taking shear as requiring a second failure property for isotropic materials then leads to a two-property failure specification.

For these reasons the present work considers only two property failure forms. All failure forms with three or more parameters or properties have, at least partially, an empirical basis. Any combination of two independent properties tests then calibrates the failure forms considered here for general use. For example, the dilatational and shear failure properties just discussed could be directly determined by testing. However, by far the most common, most accessible, and most generally useful tests for failure properties are those of uniaxial tension and compression. Uniaxial tension and compression have a certain symmetry as stress state stimuli, while the corresponding magnitudes of the failure responses can have an asymmetry which reflects fundamental material response mechanisms. These failure properties, specified by "*T*" and "*C*," will be used throughout this work.

In the following sections the Coulomb-Mohr and Drucker-Prager theories will be fully compared with each other as well as with a third recently developed theory, Christensen [5]. It is important to effect these comparisons in both the ductile and brittle ranges since any purported general theory must span a broad range of physical effects. In this examining process a variety of stress states will be considered, especially looking for unusual and even physically unacceptable aspects of behavior. It is necessary to begin with clear, concise mathematical statements of the three failure theories.

1.1 Coulomb-Mohr Failure Theory. The procedure for the Coulomb-Mohr theory (C-M) uses the Mohr's circle construction as shown in Fig. 1. On the normal stress, σ, versus shear stress, τ, axes plot the circles for uniaxial tension and compression to failure as shown in Fig. 1. Then two linear failure envelopes are taken as just being tangent to the two circles. In order to evaluate the safety for any given three dimensional state of stress, the maximum and minimum principal stresses are used to form a circle on the σ-τ plot. If this circle is inside the failure envelopes then there is no failure, otherwise there is failure of the material. For example, take a state of simple shear stress. Form a circle with its center at the origin in Fig. 1 and just tangent to the failure envelopes. This circle intersects the τ axis at the values of the shear stress at failure, S, given by

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received May 21, 2004; final manuscript received December 2, 2005. Review conducted by K. M. Liechti. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

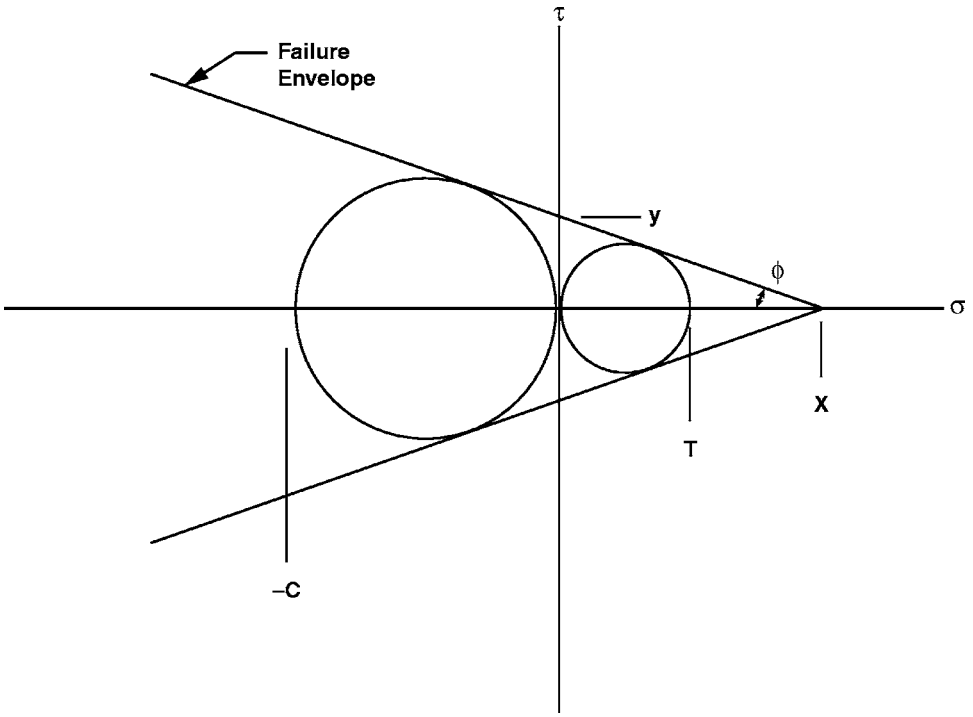


Fig. 1 Coulomb-Mohr failure form

$$S = \frac{TC}{T + C} \tag{1}$$

where T and $-C$ are the failure levels in uniaxial tension and compression. The general characteristics of the construction in Fig. 1 are given by

$$x = \frac{TC}{C - T}$$

$$y = \frac{1}{2}\sqrt{CT} \tag{2}$$

and

$$\sin\phi = \frac{C - T}{C + T}$$

where $T \leqslant C$.

The procedure just outlined for failure assessment from Fig. 1 can be expressed in analytical form. Let σ_1, σ_2, and σ_3 be the principal stresses with σ_1 being the largest and σ_3 being the smallest in an algebraic sense. From relations (1) and (2) and Fig. 1, the failure criterion can be shown to be given by

$$\frac{1}{2}\left(\frac{1}{T} - \frac{1}{C}\right)(\sigma_1 + \sigma_3) + \frac{1}{2}\left(\frac{1}{T} + \frac{1}{C}\right)(\sigma_1 - \sigma_3) \leqslant 1 \tag{3}$$

The first term shows the normal stress effect and the second term the maximum shear stress effect. Relation (3) directly reduces to

$$\frac{\sigma_1}{T} - \frac{\sigma_3}{C} \leqslant 1 \tag{4}$$

This remarkably simple form constitutes the entire two-property C-M failure criterion.

The C-M form was and is very appealing in its simplicity of concept and its generality and ease of use. Simple though it may seem, care still must be taken because it is often misinterpreted. There was great enthusiasm for it in the early 1900s when it came into general use. Then it was shown by von Karmen [6] and Böker [7] to have strong limitations for brittle materials, specifically natural minerals. Much effort has been expended over the years to "correct" and generalize this most simple and direct form of the C-M criterion, but to no special advantage or usefulness. Nevertheless it has continued to be widely used, perhaps because of the lack of a suitable two-property alternative.

In the limit of $T = C$ the C-M form reduces to the maximum shear stress criterion of Tresca. In principal stress space this is given by the infinite cylinder of hexagonal cross section. When $T \neq C$ the C-M form becomes a six-sided pyramid in principal stress space, and this six-sided pyramid has but three-fold symmetry, rather than six-fold symmetry. An analytical form for this criterion in terms of invariants has been given by Schajer [8], although it is far easier to directly use the forms given here. The Coulomb-Mohr theory is used in applications as diverse as from nanoindentation to large scale geophysics.

As mentioned above, the C-M form is known to miss physical reality in a quantitative sense even though it has many of the proper ingredients in its formulation. Now after 232 years, this first developed theory of failure is still the standard by which a more suitable and comprehensive theory must be compared and judged. That is a special tribute to Coulomb and to Mohr.

1.2 Drucker-Prager Failure Theory. The Drucker-Prager [4] theory (D-P) probably was motivated by the desire for a general theory that would reduce to the Mises form in the limit, rather than the Tresca form. A course of action to accomplish this is quite direct. Simply replace the C-M six-sided pyramid in principal stress space by a circular cone.

Following this direction, write the possible failure criterion as

$$a\sigma_{ii} + b\sqrt{s_{ij}s_{ij}} \leqslant 1 \tag{5}$$

where s_{ij} is the deviatoric stress tensor involved in a Mises criterion, but now the term σ_{ii} explicitly brings in the mean normal stress effect which is implicit in the C-M formalism. Parameters a and b are to be determined such that (5) predicts uniaxial stress failure at the tensile value T and the compressive value C. It is easily shown that this then gives (5) as

$$\frac{1}{2}\left(\frac{1}{T}-\frac{1}{C}\right)\sigma_{ii}+\frac{1}{2}\left(\frac{1}{T}+\frac{1}{C}\right)\sqrt{\frac{3}{2}s_{ij}s_{ij}}\leqslant 1 \tag{6}$$

Finally, writing (6) out in terms of components gives

$$\frac{1}{2}\left(\frac{1}{T}-\frac{1}{C}\right)(\sigma_{11}+\sigma_{22}+\sigma_{33})+\frac{1}{2}\left(\frac{1}{T}+\frac{1}{C}\right)\left\{\frac{1}{2}[(\sigma_{11}-\sigma_{22})^2+(\sigma_{22}-\sigma_{33})^2+(\sigma_{33}-\sigma_{11})^2]+3(\sigma_{12}{}^2+\sigma_{23}{}^2+\sigma_{31}{}^2)\right\}^{1/2}\leqslant 1 \tag{7}$$

As shown, only the positive sign is associated with the square roots in (5)–(7).

In principal stress space Eq. (7) is that of a conical failure surface. When $T=C$ the failure surface becomes the circular cylindrical form of Mises.

Like the C-M form the D-P form receives attention and usage, although not to as great an extent. For example, the D-P form was recently used by Wilson [9] in considering metal plasticity.

1.3 Present Failure Theory. Taken directly from Christensen [5], the present failure theory is given by the quadratic form

$$\alpha\left(\frac{\sigma_{ii}}{\kappa}\right)+\frac{3}{2}(1+\alpha)\left(\frac{s_{ij}}{\kappa}\right)\left(\frac{s_{ij}}{\kappa}\right)\leqslant 1 \tag{8}$$

and if under the conditions of use and testing there is the condition $T/C\leqslant 1/2$ (giving the brittle range of behavior) then the following fracture criterion also applies:

$$\sigma_1\leqslant \sigma_{11}{}^T \quad \text{if } \frac{\sigma_{11}{}^T}{|\sigma_{11}{}^C|}<\frac{1}{2} \tag{9}$$

where

$$\kappa=|\sigma_{11}{}^C| \tag{10}$$

$$\alpha=\frac{|\sigma_{11}{}^C|}{\sigma_{11}{}^T}-1 \quad \text{for } \alpha\geqslant 0 \tag{11}$$

and where σ_1 is the largest principal stress.

In component form and in the present notation (8) and (9) become

$$\left(\frac{1}{T}-\frac{1}{C}\right)(\sigma_{11}+\sigma_{22}+\sigma_{33})+\frac{1}{TC}\left\{\frac{1}{2}[(\sigma_{11}-\sigma_{22})^2+(\sigma_{22}-\sigma_{33})^2+(\sigma_{33}-\sigma_{11})^2]+3(\sigma_{12}{}^2+\sigma_{23}{}^2+\sigma_{31}{}^2)\right\}\leqslant 1 \tag{12}$$

and

$$\sigma_1\leqslant T \quad \text{if } \frac{T}{C}<\frac{1}{2} \tag{13}$$

These forms are only applicable to materials for which $T/C\leqslant 1$.

Equation (12) is that of a paraboloid in principal stress space. The key difference between the conical form of (7) for D-P theory and the paraboloidal form of (12) for the present theory is due to the effect of the square root in (7) of the similar terms in each.

The fracture condition (13) provides the equations of three planes in principal stress space. At $T/C=1/2$ these planes are just tangent to the paraboloid. For $T/C<1/2$, these fracture condition planes intersect the paraboloid giving three flattened sections in the otherwise full paraboloid. The size of these fracture cut off zones increases as T/C diminishes. The fracture condition (13) is not appropriate for use when $1/2<T/C\leqslant 1$ because that corresponds to the ductile range of behavior. The nondimensional parameter α in (8) and (11) provided the formalism needed to determine the ductile versus the brittle regions, which is necessary in order to prescribe the fracture criterion (13), see Christensen [5]. The use of the maximum principal stress in the fracture criterion (9) relates to the physical preference for the crack opening mode I rather than mode II in homogeneous materials, Christensen [10]. When $T=C$ the entire criterion reverts to the Mises type circular cylinder in principal stress space.

The paraboloidal part of the present criterion, as with the pyramidal and conical forms, has a considerable history. It goes back to Sleicher [11], with interaction from von Mises [12]. Although it has received less attention than the other two forms, paraboloidal forms certainly have been examined, for example Stassi [13]. The complete formalism in (8)–(13) is a recent development, and the fracture restriction (9) in particular is a vital part of the present criterion in the brittle range.

Now the critical comparison of these three basic, two property failure theories can begin. All results in the following cases follow directly from the three failure formalisms just given.

Table 1 Shear stress at failure

	Shear stress, S/C, at failure		
	$T/C=1$	$T/C=1/2$	$T/C=1/10$
C-M	1/2	0.333	0.091
D-P	$1/\sqrt{3}$	0.385	0.105
Present	$1/\sqrt{3}$	0.408	0.1

1.4 Shear Stress, Equibiaxial and Equitriaxial Stresses. The simple shear stress, S, at failure from the three theories are found to be

$$S=\frac{TC}{T+C},\ \text{C-M} \tag{14}$$

$$S=\frac{2TC}{\sqrt{3}(T+C)},\ \text{D-P} \tag{15}$$

and

$$S=\frac{\sqrt{TC}}{\sqrt{3}} \quad \text{for } \frac{T}{C}\geqslant\frac{1}{3} \tag{16a}$$

$$S=T \quad \text{for } \frac{T}{C}\leqslant\frac{1}{3} \tag{16b}$$

present

The values of the shear stress at failure from (14) to (16) at different values of T/C are shown in Table 1.

Although the differences between the C-M results and the other two are significant, there is nothing unusually extreme in any of these results. It may be noted that the value of S/C at $T/C=1/10$ in the present theory would have been much larger than that in Table 1, were it not for the fracture condition (16*b*).

Turning now to states of equibiaxial stress, the tensile and compressive failure stresses for the three theories are found to be

$$\left.\begin{aligned}\sigma_{2D}^T&=T\\ \sigma_{2D}^C&=-C\end{aligned}\right\}\text{C-M} \tag{17}$$

$$\left.\begin{aligned}\frac{\sigma_{2D}^T}{T}&=\frac{2}{3-\frac{T}{C}}\\ \frac{\sigma_{2D}^C}{C}&=\frac{2}{\frac{C}{T}-3}\end{aligned}\right\}\text{D-P} \tag{18}$$

and

Table 2 Equibiaxial and equitriaxial stresses, $T/C=1/3$

	Tension, $T/C=1/3$		
	σ_{1D}/T	σ_{2D}/T	σ_{3D}/T
C-M	1	1	3/2
D-P	1	3/4	1
Present	1	0.646	1/2
	Compression, $T/C=1/3$		
	σ_{1D}/C	σ_{2D}/C	σ_{3D}/C
C-M	−1	−1	$-\infty$
D-P	−1	$-\infty$	$-\infty$
Present	−1	−1.55	$-\infty$

$$\left.\begin{aligned}\frac{\sigma_{2D}^{T}}{T}&=\left(1-\frac{C}{T}\right)+\sqrt{1-\frac{C}{T}+\frac{C^2}{T^2}}\\ \frac{\sigma_{2D}^{C}}{C}&=-\left(1-\frac{T}{C}\right)-\sqrt{1-\frac{T}{C}+\frac{T^2}{C^2}}\end{aligned}\right\}\text{present} \tag{19}$$

In states of eqitriaxial tension the results are

$$\frac{\sigma_{3D}^{T}}{T}=\frac{1}{1-\dfrac{T}{C}},\ \text{C-M.} \tag{20}$$

$$\frac{\sigma_{3D}^{T}}{T}=\frac{2/3}{1-\dfrac{T}{C}},\ \text{D-P} \tag{21}$$

$$\frac{\sigma_{3D}^{T}}{T}=\frac{1/3}{1-\dfrac{T}{C}},\ \text{present} \tag{22}$$

In equitriaxial compression all three theories give the same result

$$\sigma_{3D}^{C}=-\infty \tag{23}$$

First, note from (20) to (22) for equitriaxial tension, the differences in failure stresses are in the ratios of 1:2:3 for the respective present, D-P and C-M theories irrespective of the values of T/C, and therefore occur in all ductile and brittle ranges. These are very large differences. It is not surprising that many schemes have been proposed for cutting off the apex regions of the C-M and D-P failure surfaces, Paul [14]. To do so, however, requires the insertion of more parameters into these theories, which then would need to be determined.

The equibiaxial and equitriaxial failure stress results are shown in Tables 2 and 3. Table 2 gives results for $T/C=1/3$ which is somewhat into the brittle range while Table 3 is for the limit of completely damaged materials wherein the uniaxial tensile failure stress is negligible compared with the compressive value, $T/C\rightarrow 0$. In general, the three methods give distinctly different results. The C-M method gives the equibiaxial results as being identical with the uniaxial results, the other two methods do not do so. The most unusual and unexpected result is the prediction from D-P theory that the equibiaxial compressive stress at failure is infinitely large at $T/C=1/3$, Table 2. This value follows from σ_{2D}^{C} in Eq. (18) which ceases to be valid for values of T/C less than $T/C=1/3$. The value of $T/C=1/3$ is about that typical of cast iron. It is extremely unlikely that cast iron has an unlimited resistance to equibiaxial compressive stress. This peculiar prediction from the D-P theory suggests looking more fully at biaxial stress states, as is done next.

Table 3 Equibiaxial and equitriaxial stresses, $T/C\rightarrow 0$

	Tension, $T/C\rightarrow 0$		
	σ_{1D}/T	σ_{2D}/T	σ_{3D}/T
C-M	1	1	1
D-P	1	2/3	2/3
Present	1	1/2	1/3
	Compression, $T/C\rightarrow 0$		
	σ_{1D}/C	σ_{2D}/C	σ_{3D}/C
C-M	−1	−1	$-\infty$
D-P	−1	$-\infty$	$-\infty$
Present	−1	−2	$-\infty$

1.5 General Biaxial Stress States. Consider a fully biaxial stress state involving σ_{11} and σ_{22} with all other stress components as vanishing. The C-M form follows directly from the forms given. The biaxial stress states for the D-P and present theories are given by

$$\sigma_{11}{}^2+(C-T)\sigma_{11}-\frac{1}{4}\left(3\frac{C}{T}-2+3\frac{T}{C}\right)\sigma_{11}\sigma_{22}+(C-T)\sigma_{22}+\sigma_{22}{}^2 = TC,\ \text{D-P} \tag{24}$$

and

$$\sigma_{11}{}^2+(C-T)\sigma_{11}-\sigma_{11}\sigma_{22}+(C-T)\sigma_{22}+\sigma_{22}{}^2=TC,\ \text{present} \tag{25}$$

The fracture condition (13) must also be applied along with (25) in the present forms.

The corresponding failure forms are shown in Fig. 2 for the value $T/C=1/3$. If the value of $T/C=1$ were taken, the result would have been the Mises ellipse for the D-P and present forms and the six-sided Tesca figure for the C-M case.

It is seen from Fig. 2 that the D-P theory gives an open-ended form while the C-M and present forms give closed failure surfaces for $T/C=1/3$. Furthermore, the C-M and present theories give closed failure forms for all values of T/C while the D-P forms are open for all $T/C\leqslant 1/3$. This latter characteristic was discussed in the last section and found to be physically unrealistic. This is the first evidence of unrealistic behavior by any of the three failure theories, in this case the D-P theory. The mathematical reason for this behavior is straightforward.

The axis of the D-P cone in principal stress space makes equal angles with the three coordinate axes. At the value of $T/C=1/3$ the cone has a sufficiently large opening angle such that the coordinate planes intersecting the cone produces curves of unlimited extent. This is the biaxial stress failure locus. A cone and a six-sided pyramid may seem quite similar, with the expectation that the C-M and D-P predictions would be only slightly different, similarly to the Mises and Tresca forms of the ductile limit. That reasoning, however, would be incorrect. For values of T/C considerably less than 1, the six-sided pyramid of C-M becomes closer to being a triangular pyramid, and the opening angles of the two forms are different. The results here reveal the two methods to be fundamentally different, and the C-M method is not subject to the same deficiency as that just found for the D-P method.

1.6 Shear Stress Plus Pressure. Shear stress by itself has been considered. Now take shear stress plus a superimposed pressure. Let τ be the shear stress and p be the pressure. The shear stress at failure is given by

$$\tau=\frac{TC}{T+C}+\frac{(C-T)}{(C+T)}p,\ \text{C-M} \tag{26}$$

$$\tau=\frac{2TC}{\sqrt{3}(T+C)}+\frac{\sqrt{3}(C-T)}{(C+T)}p,\ \text{D-P} \tag{27}$$

and

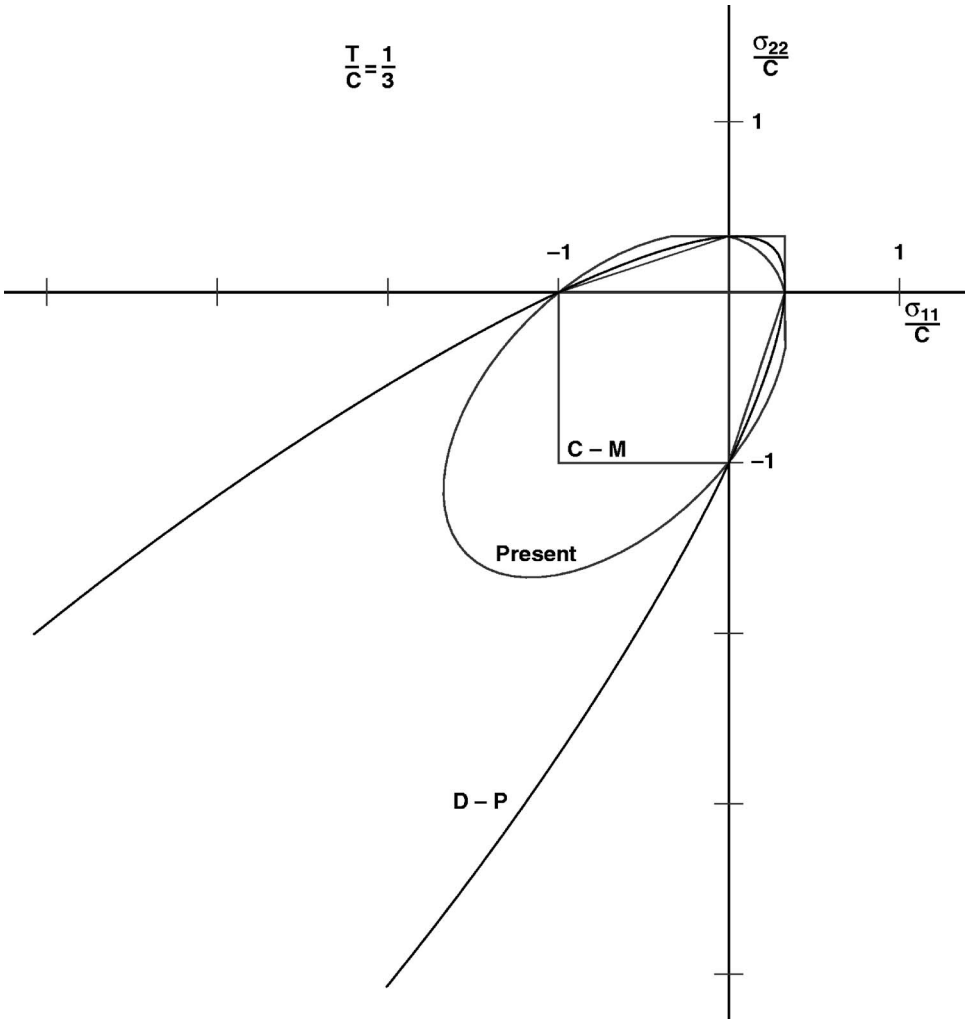


Fig. 2 Biaxial stress state failure, $T/C=1/3$

$$\tau = \sqrt{\frac{TC}{3} + (C-T)p} \quad \text{for } \frac{T}{C} \geqslant \frac{1}{3}, \text{ present} \tag{28}$$

In the present prediction, (28), for $T/C \leqslant 1/3$, it is necessary to consider each case separately to determine whether the fracture criterion (13) applies instead of (28).

For the case of $T/C=1/3$ and $P/C=4$ these results give

$$\left.\begin{aligned} &\frac{\tau}{C} = 2.25, \quad \text{C-M} \\ &\frac{\tau}{C} = 3.75, \quad \text{D-P} \\ &\text{and} \\ &\frac{\tau}{C} = 1.67, \quad \text{present} \end{aligned}\right\} \tag{29}$$

The differences of all three methods are considerable. For larger values of pressure, p, the differences between the present method and the other two become much larger, however, at some level of high pressure the predictions from all three methods must be considered to be out of range. Although these differences are large, they cannot, by themselves, be used to differentiate reasonable from unreasonable behavior for any of the three theories. The forms (26)–(28) and the results (29) do, however, show that the present failure methodology has a fundamentally different character from the other two.

1.7 Some Triaxial Stress States. Now some fully three-dimensional stress states will be considered which will show extreme differences between the three theories. As always, the C-M failure surfaces for any particular stress state follows directly from (3) or (4). The governing forms for the other two theories must be deduced from the forms given: D-P, Eq. (7) and present theory Eqs. (12) and (13).

To begin, two particular stress states will be considered

$$\text{(i)} \quad \sigma_{11} \text{ and } \sigma_{22} = \sigma_{33} \tag{30}$$

and

$$\text{(ii)} \quad \sigma_{33} = \sigma_{11} \text{ and } \sigma_{22} \tag{31}$$

The initial reason for choosing these two stress cases is that the C-M method is not able to distinguish between them. The other two methods will show a strong difference between the two cases.

For the D-P theory, the resulting failure envelopes are given by

D-P, case (i), σ_{11} and $\sigma_{22}=\sigma_{33}$

$$\text{for} \quad \frac{\sigma_{11}}{C} \geqslant \frac{2/3}{\frac{C}{T} - 1} \tag{32}$$

then

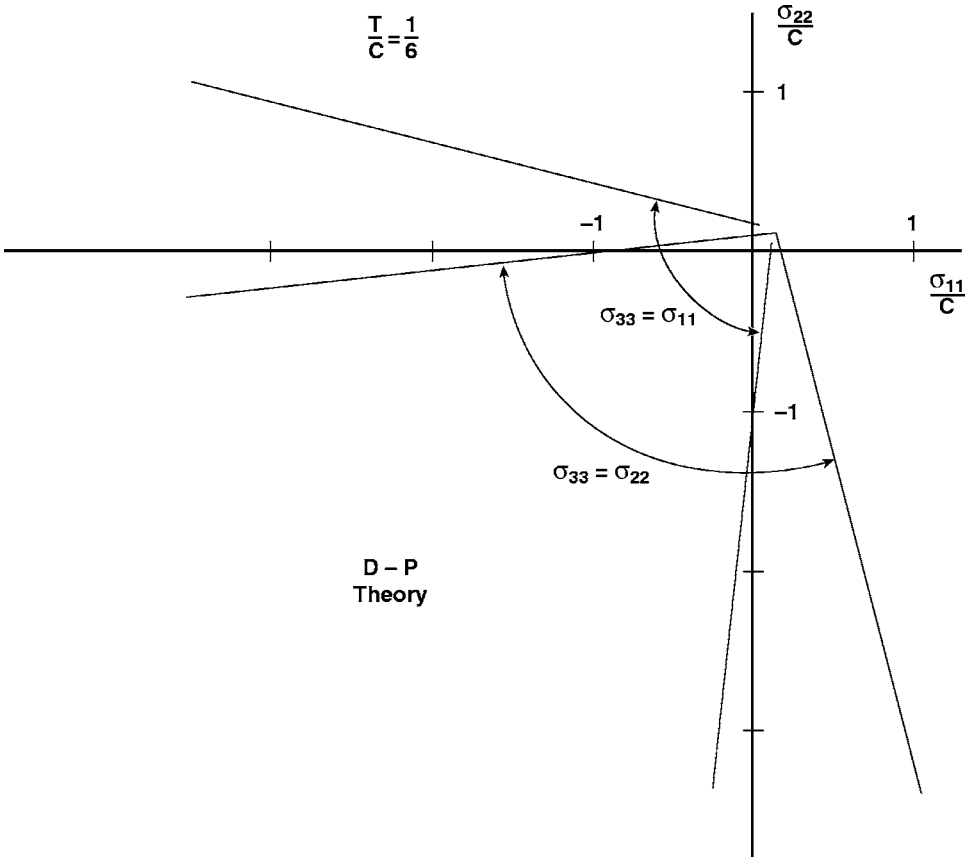


Fig. 3 Stress states σ_{11} vs $\sigma_{22}=\sigma_{33}$ and $\sigma_{11}=\sigma_{33}$ versus σ_{22}, D-P theory

$$\frac{\sigma_{22}}{C}=\frac{2\left(\frac{T}{C}-\frac{\sigma_{11}}{C}\right)}{1-3\frac{T}{C}} \tag{33}$$

and for

$$\frac{\sigma_{11}}{C}\leqslant\frac{2/3}{\frac{C}{T}-1} \tag{34}$$

then

$$\frac{\sigma_{22}}{C}=\frac{2\left(1+\frac{\sigma_{11}}{C}\right)}{3\frac{C}{T}-1} \tag{35}$$

Now for case (ii) the results are found by the interchange of indices 11 and 22 in case (i). This is easily verified by the independent derivation of the forms for case (ii).

D-P, case (ii), $\sigma_{33}=\sigma_{11}$ and σ_{22}

$$\text{for}\quad\frac{\sigma_{22}}{C}\geqslant\frac{2/3}{\frac{C}{T}-1} \tag{36}$$

then

$$\frac{\sigma_{11}}{C}=\frac{2\left(\frac{T}{C}-\frac{\sigma_{22}}{C}\right)}{1-3\frac{T}{C}} \tag{37}$$

and for

$$\frac{\sigma_{22}}{C}\leqslant\frac{2/3}{\frac{C}{T}-1} \tag{38}$$

$$\frac{\sigma_{11}}{C}=\frac{2\left(1+\frac{\sigma_{22}}{C}\right)}{3\frac{T}{C}-1} \tag{39}$$

In the present failure theory it is found that

Present theory, case (i), σ_{11} and $\sigma_{22}=\sigma_{33}$

$$\frac{\sigma_{11}}{C}=-\frac{1}{2}\left(1-\frac{T}{C}\right)+\frac{\sigma_{22}}{C}\pm\sqrt{\frac{1}{4}\left(1-\frac{T}{C}\right)^2-3\left(1-\frac{T}{C}\right)\frac{\sigma_{22}}{C}} \tag{40}$$

Present theory, case (ii), $\sigma_{33}=\sigma_{11}$ and σ_{22}

$$\frac{\sigma_{22}}{C}=-\frac{1}{2}\left(1-\frac{T}{C}\right)+\frac{\sigma_{11}}{C}\pm\sqrt{\frac{1}{4}\left(1+\frac{T}{C}\right)^2-3\left(1-\frac{T}{C}\right)\frac{\sigma_{11}}{C}} \tag{41}$$

The present theory also involves the fracture condition (13).

With the C-M result from Fig. 1 and (4) with (32)–(39) various comparisons can be made. The value $T/C=1/6$ will be used as representing a quite brittle material such as a ceramic. The D-P, present and C-M results are as shown in Figs. 3–5. In Figs. 3 and 4 the D-P and the present methods strongly distinguish between the stress states (i) and (ii) in (30) and (31). The C-M method, Fig. 5, cannot distinguish between these two cases, and this is a major shortcoming for the method. The reason the C-M shows this unusual behavior is that the method only uses the maximum and minimum principal stresses in its formulation. It cannot discrimi-

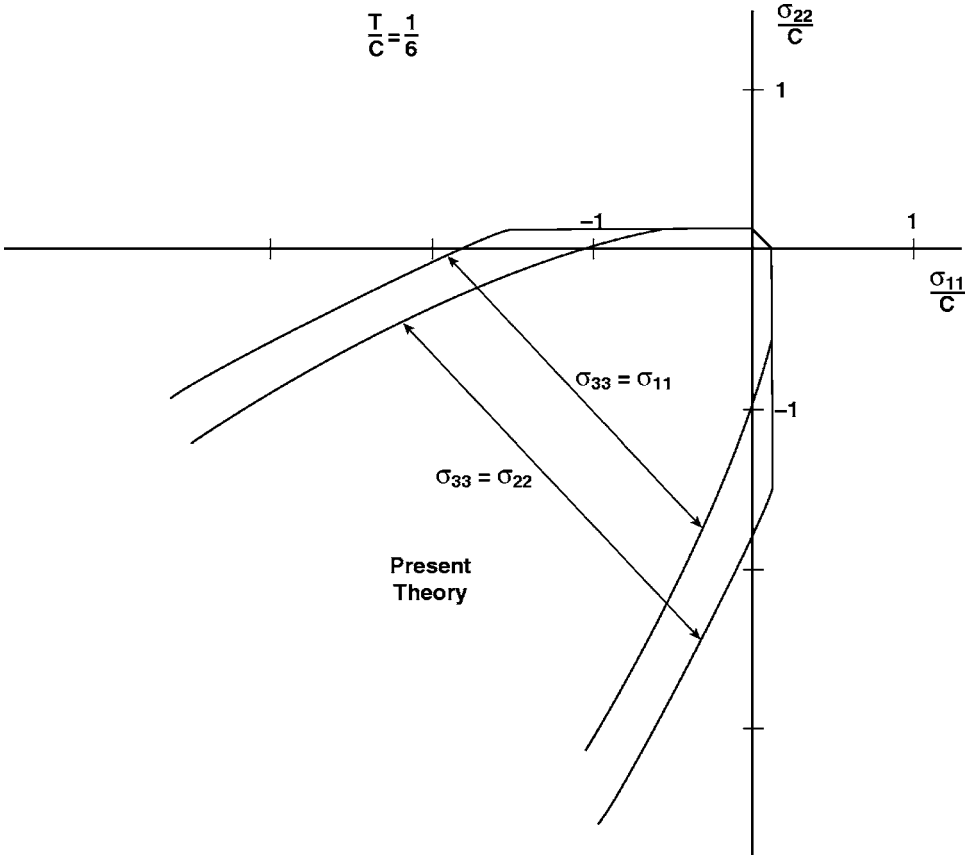


Fig. 4 Stress states σ_{11} versus $\sigma_{22}=\sigma_{33}$ and $\sigma_{11}=\sigma_{33}$ versus σ_{22}, present theory

nate different cases involving wide fluctuations in the intermediate principal stress, as are involved in cases (i) and (ii). The two cases (i) and (ii) must be different when viewed in σ_{11}, σ_{22} space. Stress states near $\sigma_{11}=\sigma_{33}$ will be more tolerable than ones similarly near σ_{22} because of the greater stabilizing effect of the greater hydrostatic compressive stress for $\sigma_{11}=\sigma_{33}$ than for σ_{22}, when both are in the generally compressive region.

The results from the three methods shown in Figs. 3–5 are very widely different from each other. In Fig. 3 it is seen that the D-P method gives an unusually widely opened form for the failure envelopes. The fracture cutoff condition of the present theory is strongly apparent in Fig. 4. A much larger scale than that in Fig. 4 would be needed to see the detail in the tensile region near the origin.

The examples just considered suggest looking at one further related example. Take a case of $T/C=1/10$ which is still in the range of behavior for ceramics. Now take a stress state that is nearly that of uniaxial compression, but with a slight transverse pressure specified by

$$\frac{\sigma_{11}}{C}=-\beta$$

$$\frac{\sigma_{22}}{C}=\frac{\sigma_{33}}{C}=-\frac{\beta}{10} \tag{42}$$

The process now will be to solve for the value of β for failure from all three theories. If there were no transverse pressure, the value of β would simply be 1 for all three theories in uniaxial compression.

For the C-M criterion when stress state (42) is substituted into (4), it gives

$$(0)\beta \leqslant 1, \text{ C-M} \tag{43}$$

where the two terms in (4) cancel each other. For the D-P case when (42) is substituted into (7) the result is

$$\left(\frac{-9}{20}\right)\beta \leqslant 1, \text{ D-P} \tag{44}$$

Thus, in both cases no matter how large is β, failure cannot occur.

Finally, using the stress state (42) in the present theory criterion (12), it is found that

$$\beta=\frac{2}{3}+\frac{\sqrt{46}}{9}, \text{ present.} \tag{45}$$

To summarize then for a material with T/C=1/10, and for the nearly uniaxial stress state (42) the three theories predict

C-M, $\beta \to \infty$, no failure
D-P, $\beta \to \infty$, no failure
Present, $\beta = 1.42$, at failure

The conclusion then is that for the stress state (42) representing only a slight deviation from a uniaxial compressive state, the C-M and D-P theories both predict a drastic shift in behavior actually giving a no failure result while the present theory predicts a failure value that is only slightly changed from that of uniaxial compression. This represents physically unrealistic behavior from both the C-M and D-P theories. Although this unrealistic behavior was demonstrated here for the value $T/C=1/10$, the same behavior can be found for a wide range of T/C values. The reason for this abrupt transition from standard behavior to unacceptable behavior is that the stress state given by (42) lies entirely within the safety zone of both the C-M and D-P criteria, and the transition in going from safety of stress state (42) to failure for uniaxial compression

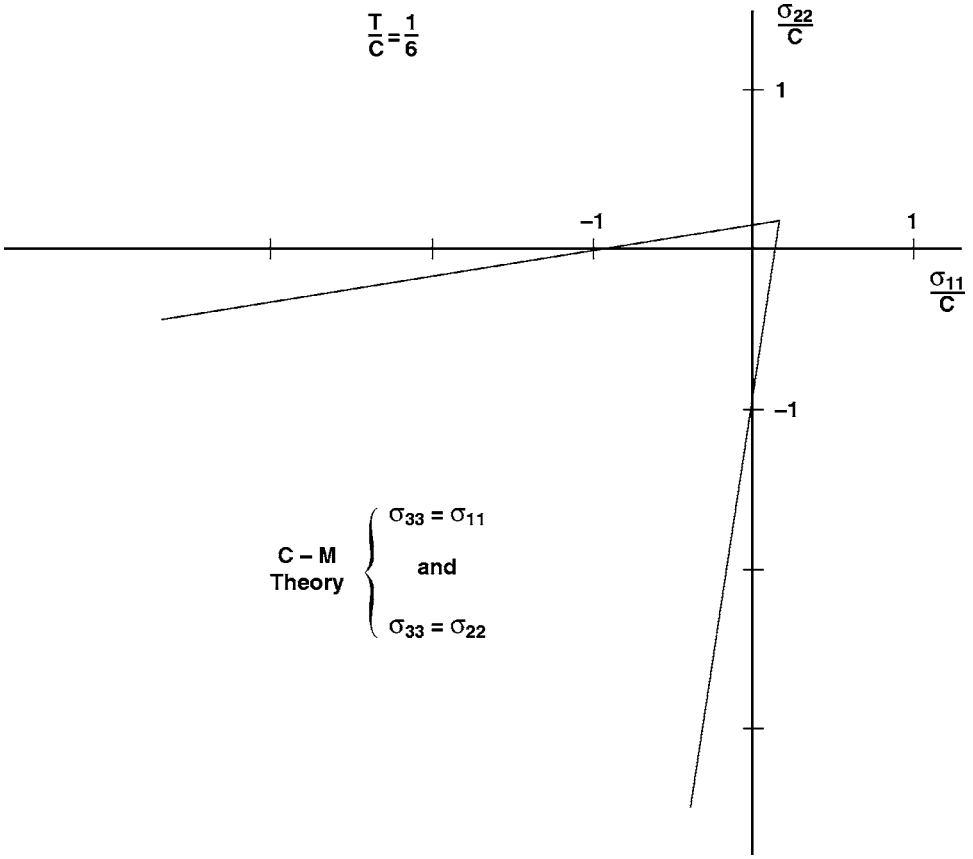


Fig. 5 Stress states σ_{11} vs $\sigma_{22}=\sigma_{33}$ and $\sigma_{11}=\sigma_{33}$ vs σ_{22}, C-M theory

is extremely sensitive to the level of the transverse pressure in the cases of the C-M and D-P criteria.

2 Conclusions

Viewing the Coulomb-Mohr and Drucker-Prager forms as pyramidal and conical surfaces in principal stress space, they both are seen to be forms of first degree with respect to variations in the axial direction, starting at the apex. As shown here, both the C-M and D-P forms give some aspects of unphysical behavior, and in general both are inadequate for the purpose of failure characterization ranging from totally ductile to extremely brittle. The deficiency relates to the first degree form of the two criteria mentioned above resulting in a strong overestimation of tolerable stress in both the very compressive and the very tensile regions of principal stress space, along with other anomalies in the D-P case. For purposes of broad range applications and to correct these deficiencies, it appears necessary to characterize the failure function by a second-degree form (combined with a fracture failure mode). By no means is the present failure theory the only possible form of the type just specified. However, it does appear to be the only one that is presently available possessing these requisites and having general 3D failure calibrated by just two properties for homogeneous and isotropic materials.

Acknowledgment

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48. This work was partially supported by the Office of Naval Research, Dr. Y. D. S. Rajapakse, Program Manager.

References

[1] Coulomb, C. A., 1773, *In Memories de Mathematique et de Physique* (Academie Royal des Sciences par divers sans), 7, pp. 343–382.
[2] Mohr, O., 1914, *Abhandlungen aus dem Gebiete der Technischen*, 2nd ed., Ernst, Berlin.
[3] Nadai, A., 1950, *Theory of Flow and Fracture of Solids, I and II*, McGraw Hill, New York.
[4] Drucker, D. C., and Prager, W., 1952, "Soil Mechanics and Plastic Analysis or Limit Design," Q. Appl. Math., **10**, pp. 157–165.
[5] Christensen, R. M., 2004, "A Two Property Yield, Failure (Fracture) Criterion for Homogeneous, Isotropic Materials," J. Eng. Mater. Technol., **126**, pp. 45–52.
[6] von Karman, T., 1912, "Festigkeits versuche unter allseitigem Druck," Mitt. Forschungsarbeit Gebiete Ingeneeures, **118**, pp. 37–68.
[7] Böker, R., 1915, "Die Mechanik der Bleibenden Formänderung in Kristallinish Aufgebauten Körpen," Mitteilungen Forschungsarbeit auf dem Gebiete Ingeniurwesens, **24**, pp. 1–51.
[8] Schajer, G. S., 1998, "Mohr-Coulomb Failure Criterion Expressed in Terms of Stress Invariants," J. Appl. Mech., **65**, pp. 1066–1068.
[9] Wilson, C. D., 2002, "A Critical Reexamination of Classical Metal Plasticity," J. Appl. Mech., **69**, pp. 63–68.
[10] Christensen, R. M., 2003, "The Brittle Versus the Ductile Nature of Fracture Modes I and II," Int. J. Fract., **123**, pp. L157–L164.
[11] Schleicher, F., 1926, "Der Spannongszustand an der Flieffgrenze Plastizitätsbedingung," Z. Angew. Math. Mech., **6**, pp. 199–216.
[12] von Mises, R., 1026, footnote to reference by Schleicher, Ref. [11].
[13] Stassi, F., 1967, "Flow and Fracture of Materials According to a New Limiting Condition of Yielding," Meccanica, **3**, pp. 178–195.
[14] Paul, B., "Macroscopic Criteria for Plastic Flow and Brittle Fracture," *Fracture*, H. Liebowitz ed., Academic Press, New York, Vol. II, pp. 313–496.

The Effect of Self-Assembled Monolayers on Interfacial Fracture

Alberto W. Mello

Kenneth M. Liechti[1]

Research Center for the Mechanics of Solids, Structures and Materials, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, TX 78712

This paper describes a series of experiments and analyses that were used to examine crack growth near sapphire/epoxy interfaces. Adhesion of the epoxy to the sapphire was enhanced by coating the sapphire with mixtures of two silane coupling agents that form self-assembled monolayers. A new biaxial loading device was used to conduct a series of mixed-mode fracture experiments. Crack opening interferometry, atomic force microscopy, and angle-resolved X-ray photoelectron spectroscopy allowed cohesive zone sizes, fracture surface topographies, and loci of fracture to be established. The experiments were complemented by finite element analyses that accounted for the rate- and pressure-dependent yielding of the epoxy. The analyses also made use of traction-separation laws to represent the various interphases that were produced by the mixed monolayers. The intrinsic toughness (defined as the area underneath the traction-separation curve) of the bare sapphire interfaces was independent of mode-mix and lower than values from previous experiments with glass/epoxy and quartz/epoxy specimens. The increase in overall toughness with mode-mix was completely accounted for by viscoplastic dissipation in the epoxy outside the cohesive zone. The minimum toughness of the coated sapphire interfaces was about five times higher than the mode-mix independent intrinsic toughness of the uncoated specimens. The increase in overall toughness with mode-mix was almost completely accounted for by increases in the intrinsic toughness as the traction-separation law varied with mode-mix. As a result, viscoplastic dissipation outside the cohesive zone was minimal. Atomic force fractography and X-ray photoelectron spectroscopy indicated that the crack growth mechanisms and the loci of fracture in the coated and uncoated specimens were quite different. [DOI: 10.1115/1.1940662]

1 Introduction

In studies of mixed-mode fracture near glass/epoxy interfaces, Swadener and Liechti [1] found that the increase in toughness with increasing shear was mirrored by an increase in the viscoplastic dissipation in the epoxy in a region outside the fracture process zone. The steady-state toughness envelope was shifted vertically from the viscoplastic dissipation by a quantity that was identified as the intrinsic toughness of the interface. Because the viscoplastic dissipation was essentially zero under mode 1 conditions, the values of intrinsic and steady-state toughness were the same there. The intrinsic toughness was about $2\ J/m^2$ for this interface, making it about 20 times higher than the thermodynamic work of adhesion between glass and epoxy, as determined from contact angle and contact mechanics measurements and analyses [2]. This difference was accounted for by first noting the formation of highly localized ridges on the epoxy fracture surface. In addition, angle-resolved X-ray photoelectron spectroscopic analyses of the glass fracture surfaces indicated that the cracks actually grew in the epoxy anywhere from 0.5 to 3 nm from the glass surface. This would be well within the so-called interphase region in the epoxy where the properties differ from those of the bulk material [3–6]. Given the amorphous nature of the glass and the disorder in the interphase region, the present study was motivated by the desire to examine interfacial fracture under more carefully controlled conditions. As a result, sapphire substrates have been used along with mixtures of self-assembled monolayers with a view to providing adhesion control between the substrate and a thin layer of epoxy.

Polymeric self-assembled monolayers (SAMs) generally form covalent bonds with the surface on which they are deposited. Self-assembly is usually a consequence of dense packing [7] of the molecules on the substrate. As a result, the head group at the other end of the chain may have a range of interactions with the material deposited on top of the monolayer. Depending on the materials involved, the interactions may be strong (covalent, ionic, hydrogen bonding, or polar) or weak (nonpolar) in nature. One form of adhesion control can be achieved by making use of mixed SAMs where the head groups of each SAM have very different interactions with the material above [8,9].

Zhuk et al. [8] used methyl (CH_3) and carboxy (COOH) terminal groups on 15-carbon alkanethiols to control adhesion between gold and epoxy. The thermodynamic work of adhesion of epoxy on the coated surface was linearly proportional to the $COOH/CH_3$ fraction in solution up to about 80% and was constant thereafter. A series of superlayer fracture experiments revealed that the interfacial fracture toughness increased strongly with the thermodynamic work of adhesion. The rate of toughening increased with the work of adhesion, suggesting that more and more plastic dissipation was excited in the epoxy layer. Kent et al. [9] used mixed monolayers of dodecyltrichlorosilane (DTS) and bromo-undecyltrichlorosilane (BrUTS) to control adhesion between silicon and epoxy. Both make strong covalent bonds with the silicon. The methyl terminal group on the DTS again makes weak, nonpolar interactions with the epoxy. The authors indicate that the BrUTS forms an alkyl ammonium bromide compound with the amine crosslinker that was used to cure the EPON 828 epoxy. As a result, an ionic bond was achieved with the epoxy through $R\text{-}NH_2^+Br^-R'$ bonds. Between 10% and 20% bromine ter-

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received January 13, 2004; final manuscript received October 8, 2004. Editor: R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

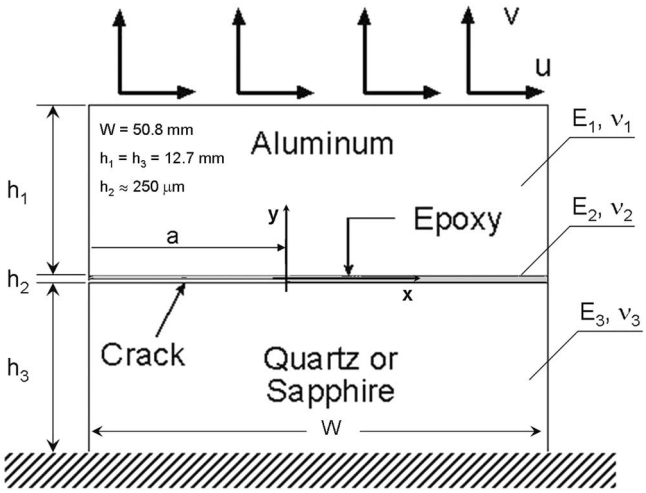


Fig. 1 Sandwich specimen

mination, there was a strong increase in the tensile and shear strengths of the silicon/epoxy interface as determined by cruciform and napkin ring shear experiments. Asymmetric double cantilever experiments were used to determine the toughness of the interface, which increased linearly with the bromine fraction. The linear relationship was ascribed to the linear increase in the thermodynamic work of adhesion with bromine fraction. In contrast to the gold/epoxy experiments, any plastic dissipation effects were apparently the same for all bromine fractions, even though the toughness of the silicon/BrUTS/DTS/epoxy was much higher than that of the gold/COOH/CH3/epoxy interface in moist environments. The mode-mix in the asymmetric double cantilever beam experiments was −8 deg at a reference length of 10 μm, whereas it was about 50 deg in the superlayer experiments.

The objective of this work was to examine the interfacial toughness envelopes of sapphire/epoxy and sapphire/SAM/epoxy interfaces. Fracture and toughening mechanisms were examined via crack-opening interferometry, angle-resolved X-ray photoelectron spectroscopy, atomic force microscopy, and finite element analyses that accounted for the viscoplastic nature of the epoxy and adhesive interactions across the interfaces via traction-separation laws.

2 Experimental Procedures

The loading device and experimental procedures that were used in this work have already been described [10]. Sandwich specimens (Fig. 1) were used in all experiments. They were made of aluminum 2024-T3 bonded to sapphire. The epoxy bond layer thickness h_2 was nominally 250 μm. The exact thickness of the bond layers was measured after each experiment. The epoxy was a Bisphenol A resin cured with amido amine (CIBA-GEIGY products Araldite GY502 and Aradur HY955-1, respectively). The two components were mixed with a weight ratio of 100:40 resin to hardener and placed in a vacuum chamber for about 15 minutes to remove bubbles and solvents. The epoxy mixture was then injected into a mold containing the two substrates that were held at a fixed distance apart [11]. The epoxy was cured for at least seven days at room temperature.

The aluminum face which was to be bonded was polished with 600 grit sand paper and treated with a chromic acid etch for five minutes at room temperature. The surface was then cleaned with acetone. The sapphire, when used as received, was only cleaned with acetone and optical tissue. When used with SAMs, the sapphire was cleaned with a "piranha etch," which is a solution of 98% H_2SO_4 (sulfuric acid) and 30% H_2O_2 (hydrogen peroxide) in volume ratios of 2–4:1. The solution was used at a temperature of 130°C for 15 minutes and provides a well-hydroxylated sapphire surface. The surfaces were then rinsed with de-ionized water and dried in a dry nitrogen stream.

Table 1 Advancing and receding water contact angles on sapphire

Surface coverage	Advancing angle	Receding angle
Clean sapphire	82 deg	43 deg
10% BrUTS, 90% DTS	94 deg	65 deg
55% BrUTS, 45% DTS	101 deg	66 deg
70% BrUTS, 30% DTS	105 deg	81 deg

A mixture of two SAMs was considered for controlling the surface interactions between the epoxy and the substrates. The SAMs were dodecyltrichlorosilane, $CH_3(CH_2)_{11}SiCl_3$, and bromo-undecyltrichlorosilane, $Br(CH_2)_{11}SiCl_3$, designated DTS and BrUTS, respectively. In both cases the Cl reacts with the hydroxylated sapphire surface to form a covalent bond between the molecule and the substrate. The terminal groups of DTS and BrUTS have weak and strong interactions, respectively, with the epoxy. As a result, the degree of adhesion was controlled by altering the ratio of BrUTS to DTS [9]. The ratios considered in this work were 0%, 10%, 55%, and 70% by volume of BrUTS. The reason for mixing DTS and BrUTS was that their chain lengths are quite similar, which discourages island formation. Following the piranha etch, the dried sapphire was subjected to a 15 s flame anneal [12]. The sapphire was then submerged for five hours at 60°C in a hexadecane solution containing a mixture of DTS and BrUTS in a volume ratio of five drops of the SAM for each 20 ml of anhydrous hexadecane. The coated sapphire was sonicated in toluene for 20 minutes to remove clusters and then blown dry with stream of dry nitrogen [11].

There is a direct relationship between the amount of Br in solution and the Br fraction that is deposited on the substrate surface [9]. In the present study, the deposition of the mixed SAMs was checked by measuring the contact angles (Rame-Hart, Inc. goniomiter, model 100-00 115) of the coated surfaces with water. Advancing and receding contact angles were measured by slowly adding and removing water from the drop with a micrometer controlled syringe and recording, respectively, the maximum and minimum contact angles to within ±1 deg. The contact angles increased with increasing amounts of BrUTS (Table 1).

Following a series of experiments at various mode-mixes, the fully debonded specimen was carefully removed from the grips. The aluminum substrate with the epoxy fracture surface was scanned in an atomic force microscope (AFM) for topography mapping. The sapphire fracture surfaces were analyzed with angle-resolved X-ray photoelectron spectroscopy (ARXPS). The substrates had to be sectioned into $12\times7\times4$ mm samples with a clean diamond saw, using de-ionized water as lubricant.

3 Analysis

In this section, we describe the stress and fracture analysis and the constitutive model used for the epoxy. We present closed form solutions for the energy release rate and mode-mix. A rate- and pressure-dependent plasticity model was used to capture the time dependence of the epoxy in the time scale of the experiments. The procedures that were used to extract cohesive zone model parameters from finite element analyses that made reference to the normal crack opening displacement (NCOD) measurements are described. The analyses were also used to examine the viscoplastic dissipation in the epoxy during crack growth.

The energy release rate for the sandwich specimen is given by

Table 2 Elastic material properties

	E (GPa)	μ (GPa)	v
Sapphire	400	222	0.20
Aluminum	70.0	26.3	0.33
Epoxy	2.03	0.66	0.36

$$G = \frac{v^2}{2}\left(\frac{h_1}{\hat{E}_1} + \frac{h_2}{\hat{E}_2} + \frac{h_3}{\hat{E}_3}\right) + \frac{u^2}{2}\left(\frac{h_1}{\mu_1} + \frac{h_2}{\mu_2} + \frac{h_3}{\mu_3}\right), \tag{1}$$

where v and u are the applied displacements normal and tangential to the crack surface, respectively [1]. The h_i are the heights of each component, also identified in Fig. 1. The quantities $\hat{E}_i = 2\mu_i[(1-v_i)/(1-2v_i)]$ are for plane strain and $\hat{E}_i = [2\mu_i/(1-v_i)]$ for plane stress.

The mode-mix angle ψ is given by

$$\psi = \tan^{-1}\frac{K_2}{K_1} + \varepsilon \ln\left(\frac{\hat{l}}{h_2}\right), \tag{2}$$

where $\hat{l}$ is an arbitrary length parameter [13]. The value of $\hat{l}$ was chosen to be 1 μm, in order to allow proper comparison with previous results obtained for glass/epoxy [1] and quartz/epoxy [11] interfaces. The stress intensity factors in Eq. (2) were obtained from

$$K_1 + iK_2 = h_2^{i\varepsilon} e^{i\omega}\left(\frac{E_*}{1-\beta^2}\right)^2 \left[\frac{v}{\sqrt{2}}\left(\frac{h_1}{\hat{E}_1} + \frac{h_2}{\hat{E}_2} + \frac{h_3}{\hat{E}_3}\right)^{-1/2} + i\frac{u}{\sqrt{2}}\left(\frac{h_1}{\mu_1} + \frac{h_2}{\mu_2} + \frac{h_3}{\mu_3}\right)^{-1/2}\right] \tag{3}$$

$$\beta = \frac{\mu_1(\kappa_2 - 1) - \mu_2(\kappa_1 - 1)}{\mu_1(\kappa_2 + 1) + \mu_2(\kappa_1 + 1)} \text{ is a Dundurs parameter} \tag{4}$$

where E_i, μ_i, and v_i are, respectively, the tensile and shear moduli and the Poisson's ratios of the upper (1) and lower (2) materials. The quantities $\kappa_i = 3-4v_i$ for plane strain and $\kappa_i = (3-v_i)/(1+v_i)$ for plane stress and

$$\frac{1}{E_*} = \frac{1}{3}\left(\frac{1-v_1^2}{E_1} + \frac{1-v_2^2}{E_2} + \frac{1-v_3^2}{E_3}\right).$$

The parameter $\varepsilon = \frac{1}{2}\pi \ln[(1-\beta)/(1+\beta)]$ is the bimaterial constant. Its value was $\varepsilon = 0.070$ for the sapphire/epoxy combinations used in this work. The angle ϖ depends on the elastic properties of all three materials and differed [1] from the results for a two-material sandwich [13] by one degree.

The stresses in the sapphire and aluminum were far below yielding for any combination of mode-mix. Consequently, these materials were considered to be linearly elastic in the finite element analyses that follow. The epoxy was subjected to very high stress levels in the vicinity of the crack tip. In addition, as the crack advanced, the plastic strain rate varied ahead of and inside the plastic zone. The mechanical behavior of the epoxy was accounted for using a power law rate-dependent plasticity model. The elastic properties of all the components can be seen in Table 2.

The sapphire used in the experiments was a single crystal with the c axis normal to the surface that became the sapphire/epoxy interface. The 400 GPa given as its modulus in Table 2 is the modulus in the c direction. The smallest modulus in the other directions is 384 GPa. Taking the sapphire to be isotropic in Eq. (1) resulted in an error of 0.7% in energy release rate. The aluminum was considered to be isotropic. The rate dependence of the epoxy [1] is shown in Fig. 2. The viscoplastic model uses the Mises yield surface with associated plastic flow and isotropic hardening. The rate dependence was accounted for by interpolation and extrapolation of data at different rates. This combination allows the hardening curve to accommodate local variations in strain rate. The equivalent plastic strain is given by

$$\varepsilon_e^{pl} = \varepsilon_e^{pl}|_t + \int_t^{t+\Delta t} \sqrt{\tfrac{2}{3}\dot{\varepsilon}^{pl}:\dot{\varepsilon}^{pl}}\, dt, \tag{5}$$

and the equivalent stress is obtained by a power-law-rate dependence:

$$\frac{\sigma_e(\varepsilon_e^{pl}, \dot{\varepsilon}_e^{pl})}{\sigma^0(\varepsilon_e^{pl})} = \left(\frac{\dot{\varepsilon}_e^{pl}}{\dot{\varepsilon}_0^{pl}}\right)^m, \tag{6}$$

where σ^0 is the stress at a base rate, ε_e^{pl} is the equivalent plastic strain, $\dot{\varepsilon}_e^{pl}$ is the equivalent plastic strain rate, and $\dot{\varepsilon}_0^{pl}$ is the base rate at which $\sigma_e = \sigma^0$. The quantity $\sigma_e = \sqrt{\frac{3}{2}(\underline{S}:\underline{S})}$ is the Mises

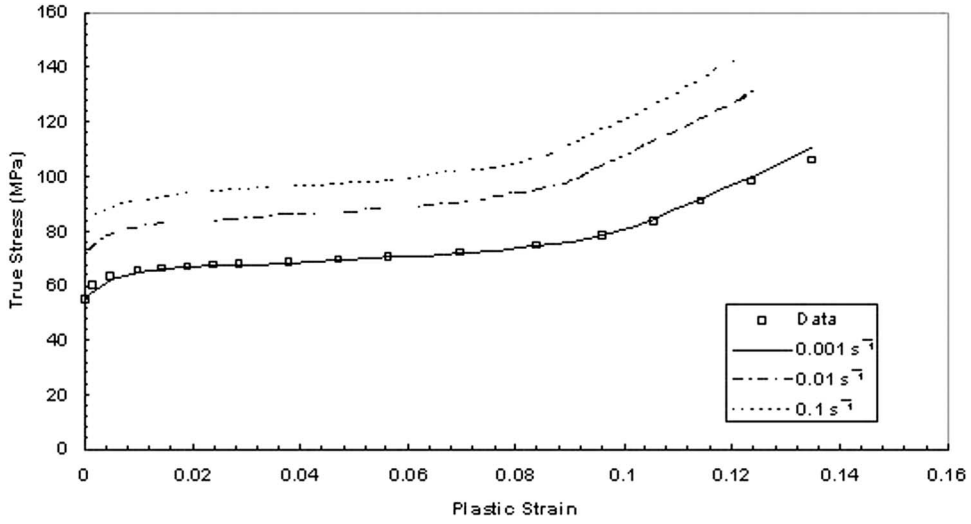


Fig. 2 Epoxy material model response in uniaxial compression [1]

equivalent stress, $\underline{S}=\underline{\sigma}+p\underline{I}$, is the deviatoric stress tensor and $p=-\frac{1}{3}\text{trace}(\underline{\sigma})$ is the hydrostatic stress.

For the current work, the rate dependence in the true uniaxial stress-strain behavior was based on data given in Fig. 2, which has a rate exponent $m=0.076$. The stresses at other rates were interpolated by the ABAQUS® solver.[2]

A plane strain two-dimensional model was chosen for this work based on previous results [1], which showed that the NCOD for a straight crack front were the same as those for a curved crack front as long as the profile was taken perpendicular to the crack front. The finite element analysis was used to obtain solutions for the NCOD, the total (G_{ss}) and plastic dissipation (G_p) components of the energy release rate, for all the specimens used in the study. All the simulations were performed in ABAQUS® 6.2 in a UNIX environment. The NCOD were obtained directly from the output file as the current relative position of the debonded surfaces. The total energy release rate was obtained by means of the J integral. Several contours were defined surrounding the crack tip. The eighth contour and above gave the most consistent results for the J integral. The plastic dissipation was obtained as the plastic work consumed during a period of crack advance, which is given by

$$\Delta W_p = \int_{\Omega} \underline{\sigma} \cdot \Delta \underline{\varepsilon}_p d\Gamma, \tag{7}$$

where Ω is the contour which encircles all elements that deformed plastically. The total value of the plastic work can be obtained directly from the energy dissipated by rate-dependent plastic deformation in each computational step as a function of the crack length. The plastic dissipation component of the energy release rate is given by

$$G_p = \frac{1}{b}\frac{\Delta W_p}{\Delta a}, \tag{8}$$

where b is the width of the specimen and Δa is the crack increment. In fact, W_p calculated in this way includes all the bulk inelastic dissipation during steady-state crack growth, but it does not predict or account for highly localized inelastic deformations that resulted during some forms of crack propagation. It was observed in the experiments with bare sapphire that the crack progression was generally discrete and highly localized. It grew laterally (see next section), forming ridges on the epoxy surface. When the sapphire was coated with mixtures of BrUTS and DTS, ligament formation was the dominant fracture mechanism. In both cases, such localized plastic dissipation was essentially included in the cohesive zone model.

The model for quartz/epoxy/aluminum and bare sapphire/epoxy/ aluminum was meshed with quadrilateral elements [11]. The mesh was highly refined in the vicinity of the crack tip to better capture the nonlinear effects. In this region, the square element size was set to 100 nm. To verify the accuracy and sensitivity of the computation, some other models were run with 50 nm^2 elements in this most critical region. The difference in the NCOD solutions was negligible. The differences in the dissipated plastic energy and the J integral were on the order of 2% and 1%, respectively, over a range of mode-mixes. In order to save computational time and ensure the stability of the computer system, the 100 nm minimum size mesh was chosen for the results presented in this work. This fine mesh occupied a region that was 20 μm along the interface. Crack propagation was permitted for 10 μm, so that any highly stressed region would still be confined to the fine mesh. Each model had a total of 15,750 plane strain elements, with quadratic interpolation (CPE8R). Due to the much higher toughness that was encountered, a coarser meshing scheme was used for the coated sapphire/epoxy/aluminum specimens. In this case, the model had a total of 7473 elements, the minimum size mesh in the vicinity of the crack tip was 1 μm, and the crack was allowed to grow for up to 80 μm inside the refined region.

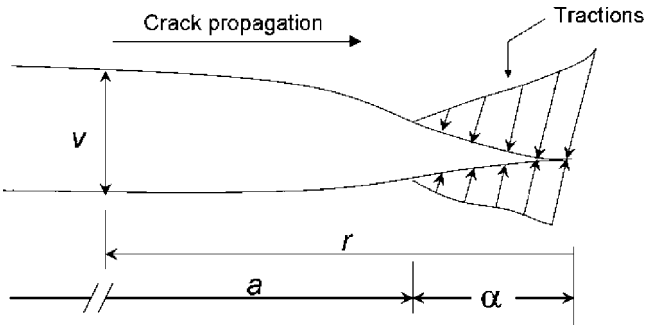


Fig. 3 Cohesive zone model

A cohesive zone model (Fig. 3) was used to represent the fracture process zone in finite element analyses of steadily propagating cracks. In the analyses, the crack speed was taken to be the average of the crack speeds observed in the experiments. In addition, once the crack tip reached a node, the tractions were released linearly with time as the crack advanced. The size of the cohesive zone α was adjusted to match the numerical solution for NCOD with the measured values. The relative displacements of the crack faces were then obtained from the solution. This scheme resulted in an approximately linear relationship between the tractions and relative displacements of the crack surfaces. In the absence of direct measurements, this traction-separation law was adequate for modeling fracture in these experiments, especially since it has been shown that the shape of the traction-separation law does not significantly affect the fracture toughness [1,14]. In order to properly represent steady-state crack propagation, the crack was allowed to grow for several cohesive zone lengths.

The intrinsic toughness Γ_0 is defined [14] as the area underneath the traction-separation curve. It can also be obtained from $\Gamma_0=G_{ss}-G_P$ and has generally been considered to be independent of mode-mix. It corresponded to the minimum value of G_{ss} for glass/epoxy [1,2] and quartz/epoxy [11] interfaces, where there was no plastic dissipation under mode 1 dominant conditions. However, as we will see here, the ligament formation mechanism that arose when sapphire was coated with self-assembled monolayers gave rise to an intrinsic toughness that varied with mode-mix, i.e., $\Gamma_0(\psi)$.

4 Results and Discussion

The crack speed was measured from the video recording of the fringe patterns from the crack opening interferometer [10]. The average crack speed under steady-state conditions was 5.0 μm/s in the experiments with bare sapphire. For the experiments with the SAM coated-sapphire, the average speed was 3.5 μm/s. The crack speed was used as input to the finite element analysis. Its value was particularly important for establishing the strain rate at the crack tip.

In some experiments with low mode-mix, it was possible to observe the crack growing by lateral tunneling at the bare sapphire interfaces. This lateral growth could be seen when the crack length increment that it produced ranged from 1 to 5 μm. Even though this lateral tunneling was observed and later confirmed by AFM analysis (Fig. 10(*a*)), the mathematical model considered an overall steady-state condition with the crack growing in the commonly assumed nondiscrete manner. This mechanism has been explained [13] and confirmed [1]; it provides additional energy to drive the crack through regions of higher fracture toughness [2]. The net result is the same whether the crack propagates by small debonds which spread laterally or grows forward uniformly, but the lateral spreading allows for a crack to propagate through re-

[2]The authors would like to acknowledge the use of the finite element package ABAQUS© under academic license from Hibbitt Karlsson & Sorensen, Inc.

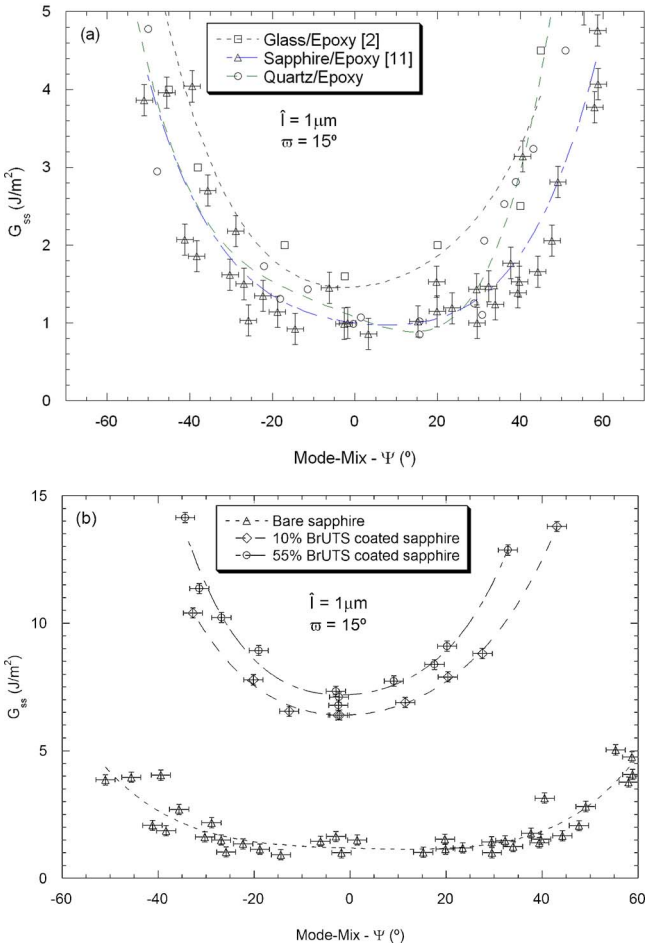


Fig. 4 Comparison of the mixed-mode fracture envelopes of (*a*) glass/epoxy and quartz/epoxy to bare sapphire/epoxy interfaces and (*b*) sapphire/epoxy with and without SAM. The SAMs were composed of 10% BrUTS and 55% BrUTS.

gions with variable fracture toughness. The magnification of the microscope used in the crack opening interferometry was not always sufficient to resolve this type of local behavior. All the measurements of NCOD were taken during steady-state crack propagation from a fully open crack front.

4.1 Toughness Envelopes. The applied displacements during steady-state crack growth were used in equations (1)–(3) for determining energy release rates and phase angles. The results for the current experiments are shown in Fig. 4 where they are compared with previous results for a glass/epoxy interface [1] and a quartz/epoxy interface [10]. All data sets refer to an arbitrary length scale parameter of 1 μm. For the sapphire experiments, there was a higher degree of scatter in the steady-state energy release rate from specimen to specimen. The sapphire surface was more susceptible to slight changes possibly introduced by reuse of specimens. To minimize this scatter, the sapphire was cleaned with piranha etch before reuse. The minimum value of G_{ss} of the bare sapphire/epoxy interface was slightly lower than that of the quartz/epoxy interface (Fig. 4(*a*)). However, for higher mode-mixes, there was an increase in G_{ss}, although the sapphire exhibited a gentler increase in toughness for positive shear. Completely different levels of G_{ss} were obtained when the sapphire was coated with mixed monolayers (Fig. 4(*b*)). It can be seen that the minimum toughness of the coated-sapphire was 4.5 and 5 times greater than that of the bare sapphire for 10% BrUTS and 55% BrUTS, respectively. This increase in toughness is due to the stronger bonding produced by the terminal bromine atoms with the epoxy through specific local interactions [9]. The methyl terminal groups of DTS only bond with the epoxy through nonlocal van der Waals interactions [9]. The higher toughness of the coated sapphire led to a much sharper minimum in the fracture toughness envelope and, based on previous experience [1,10], suggests that a significant amount of plastic dissipation was occurring even under mode 1 conditions. Such changes in the shapes of the fracture toughness envelopes were predicted [14]. The toughness of a specimen made with 70% BrUTS coated sapphire was so high that it was not possible to initiate a single crack at the interface [11]. The adhesion between the sapphire and the epoxy was such that voids nucleated at multiple sites instead. Some void growth also occurred between the epoxy and the aluminum.

4.2 Traction-Separation Laws. As indicated earlier, the scheme that was used to implement a cohesive zone analysis had the cohesive zone size as a free parameter. The proper cohesive zone size was chosen by matching the numerical solutions for the NCODs with the measured values. Figures 5(*a*) and 5(*b*) show the NCODs for the conditions $G_{ss}=2.2\ \mathrm{J/m^2}$ and $\Psi=-29$ deg and $G_{ss}=1.6\ \mathrm{J/m^2}$ and $\Psi=20$ deg , respectively. A cohesive zone size of 0.35 μm was needed to match the measured NCOD. This is consistent with the result for quartz/epoxy interfaces where a cohesive zone size of 0.5 μm provided the best match the measured NCOD [11]. The trials and matching case for a coated-sapphire/epoxy specimen at $G_{ss}=6.4\ \mathrm{J/m^2}$ and $\Psi=-2.6$ deg are shown in Fig. 5(*c*). As can be seen, a cohesive zone size of 5.6 μm provided the best agreement with the measured NCOD. The much larger cohesive zone for the coated sapphire is in agreement with the trend in toughness values.

The traction-separation laws for the analyses that produced the solutions of Figs. 5(*a*) and 5(*b*) for bare sapphire are shown in Fig. 6. It can be seen (Fig. 6(*a*) and 6(*c*)) that the normal N and shear traction S decreased almost linearly with NCOD. In these cases, the normal tractions were higher than the shear values for a given NCOD. This difference disappears when the traction-separation laws are plotted (Figs. 6(*b*) and 6(*d*)) as the vector magnitudes T and δ of the tractions N and S and the NCOD and tangential crack opening displacements (TCOD), respectively. The vector crack opening displacements (VCOD) are δ. As a check on the computations, the intrinsic fracture toughness (area underneath the traction-separation laws) was always the same. The maximum normal traction was about 120 MPa, which was higher than any of the plateau levels shown in Fig. 2. This is due to the fact that the strain rates near the crack tip were higher [11] than any of the values shown there. Nonetheless, it appears that the maximum traction was similar to the yield strength of the epoxy. This was lower than the range of values (from two to eight times the yield strength) that were used in earlier analyses [14] of interfacial crack growth.

In contrast to the results just presented for the bare sapphire specimens, the traction-separation laws for the sapphire specimens coated with 10% BrUTS varied considerably with mode-mix (Fig. 7). First, it can be seen that the traction-separation law for $\Psi=-2.6$ deg was highly nonlinear and had the lowest intrinsic fracture toughness. As the mode-mix increased, the traction-separation law became more linear and the intrinsic toughness increased with increasing mode-mix, something that we had never seen in previous experiments with the glass/epoxy [1] and quartz/epoxy [11] interfaces and the bare sapphire specimens (Fig. 6). In addition, it can be seen that the maximum levels of vector traction for the coated sapphire were about two-thirds of the values that were seen (Figs. 6(*b*) and 6(*d*)) for the bare sapphire. These lower maximum traction levels suggest that very little yielding was occurring outside the cohesive zone. At the same time, the maximum values of VCOD were about ten times larger than the values for the bare sapphire. This is consistent with the much larger cohesive zone size of the coated sapphire specimens (Fig. 5) and seems to

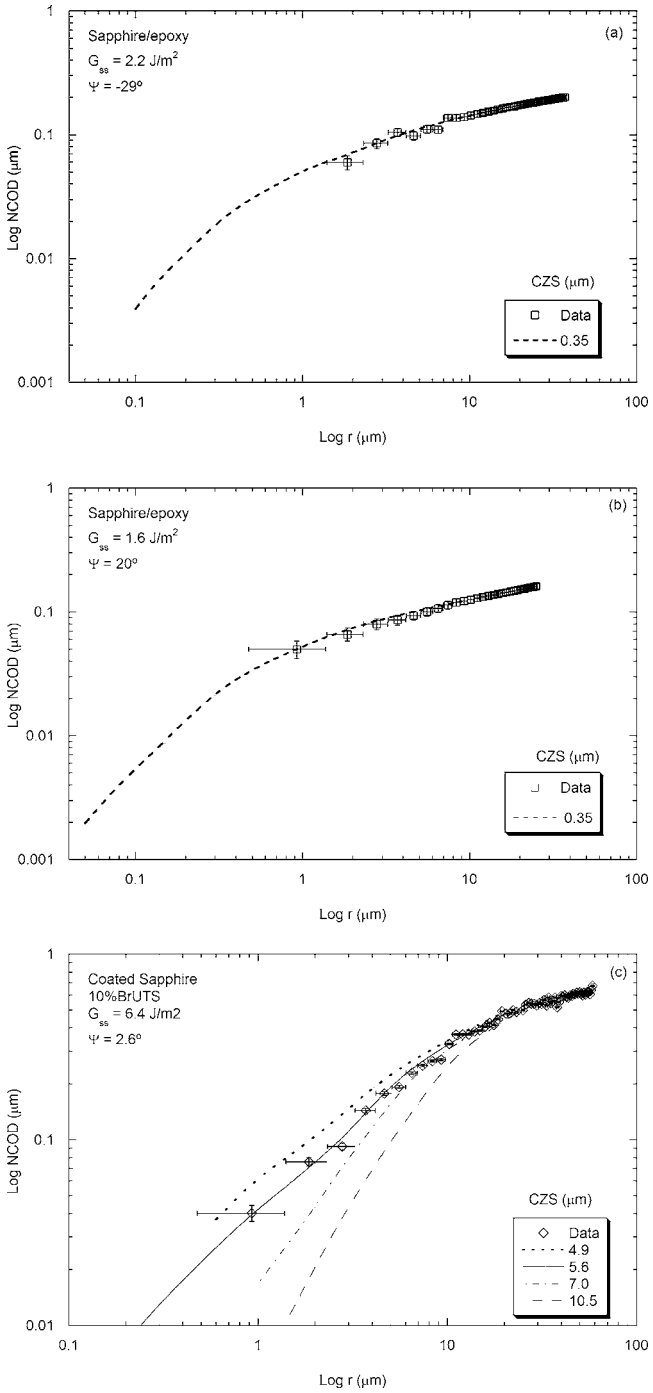


Fig. 5 A comparison of measured NCOD and FEA solutions for (*a*) a bare sapphire/epoxy interface G_{ss}=2.2 J/m^2, Ψ=−29 deg, (*b*) G_{ss}=1.6 J/m^2, Ψ=20 deg, and (*c*) four different cohesive zone sizes for a coated-sapphire/epoxy interface

be linked to the ligament formation and epoxy chain pull-out that was seen on the epoxy fracture surface (Sec. 4.4 below).

4.3 Viscoplastic Dissipation. We also examined the stress distributions, strain rates, and viscoplastic dissipation associated with steady-state crack growth. As the crack advances, a wake of plastic deformation can be seen in the epoxy. All displacements are scaled by 5, to help with visualization. The in-plane principal strain field in the vicinity of the crack tip is shown in Fig. 8(*a*) for a positive (opening) shear load in conjunction with a normal load. A highly strained region can be seen around the crack tip with an orientation in the direction of the applied shear. Figure 8(*b*) shows the strain field for negative (closing) shear. The region of intense strain did not extend into the epoxy as far as it did under positive shear. However, it did extend much further along the wake.

Values of steady-state toughness and plastic dissipation were computed as described in Eqs. (1), (2), (7), and (8). Figure 9(*a*) shows the values of G_{ss} and G_p for the bare sapphire/epoxy interface, as a function of the mode-mix. For $-5 \text{ deg} \leq \Psi \leq 20$ deg, $G_p \approx 0$, which implies that the intrinsic toughness, $\Gamma_0 \approx G_{ss}$ in that range. For larger mode-mixes, the rise in G_{ss} was mirrored by G_p as had been noted for glass/epoxy interfaces [1]. Note that the closed form and numerical solutions for G_{ss} were in good agreement. A similar series of analyses was conducted for the sapphire specimens that were coated with 10% BrUTS. As was noted earlier (Fig. 4(*b*)), this relatively small amount of BrUTS raised the minimum toughness over the intrinsic toughness of the bare sapphire by a factor of 4.5. Our first thought was that this was due to an increase in plastic dissipation, even under mode 1 conditions. However, the results (Fig. 9(*b*)) show that there was very little plastic dissipation outside the cohesive zone over the range of mode-mixes that could be analyzed. It was not possible to obtain convergent solutions for mode-mixes beyond those shown in Fig. 9(*b*). Nonetheless, it is already apparent that G_p did not track G_{ss} as was the case for quartz and bare sapphire. The increase in toughness for the coated sapphire was mainly accounted for by the increase in intrinsic toughness brought about by the BrUTS. As indicated above, this intrinsic toughness was not independent of mode-mix.

4.4 Fracture Mechanisms. The crack opening interferometry revealed some interesting features regarding crack growth mechanisms. The lateral tunneling mechanism of crack growth alluded to earlier was observed in experiments with bare sapphire substrates. It was not detected in the experiments with the coated sapphire substrates. The interferograms suggested that the fracture surfaces were smooth at least down to the resolution of the technique (30 nm). However, ridges on the epoxy fracture surfaces that were about 20 nm high had been observed in previous work with glass/epoxy interfacial fracture [2].

Some epoxy fracture surfaces from the bare-sapphire/epoxy/aluminum specimens were therefore analyzed by AFM. Evidence of lateral tunneling and permanent deformation on the crack surfaces was confirmed by the presence of ridges. For low mode-mixes ($-30 \text{ deg} < \Psi < 30$ deg), the ridges were regularly spaced, continuous, and perpendicular to the crack growth direction (Fig. 10(*a*)). This figure shows a three-dimensional rendering of the topography of a 5 μm by 5 μm region of an epoxy fracture surface at G_{ss}=1.5 J/m^2 and Ψ=−6.1 deg. The direction of crack growth was from the lower left corner to the upper right corner. The height of the ridges was on the order of 5 nm with a spacing of 1 μm. The vertical scale is enhanced to provide a better view of the ridges. The height of these ridges was smaller than that observed for glass/epoxy interfaces. This is consistent with the lower value of intrinsic toughness of the sapphire/epoxy interface. Any plastic dissipation associated with the formation of these ridges was lumped into the intrinsic interfacial toughness [2].

For higher mode-mixes, with associated increases in G_{ss}, the pattern of the ridges changed significantly. Figure 10(*b*) shows a three-dimensional view of a 16 μm by 16 μm scanned area of the epoxy fracture surface in order to reveal the structure. This view was mapped from a location where G_{ss}=4.0 J/m^2 and Ψ=−41 deg. In this case, the ridges were as high as 14 nm with an average height of 10 nm. Similar patterns were observed for high positive mode-mixes. The patterns made by the ridges did not have the same shape as had been seen for the glass/epoxy interface [2]. The ridges in the latter formed interlocking horseshoe shapes. The shapes produced by the bare sapphire/epoxy interface were much more triangular in shape. It is suspected that these ridges on the epoxy fracture surface tracked scratches on a virgin

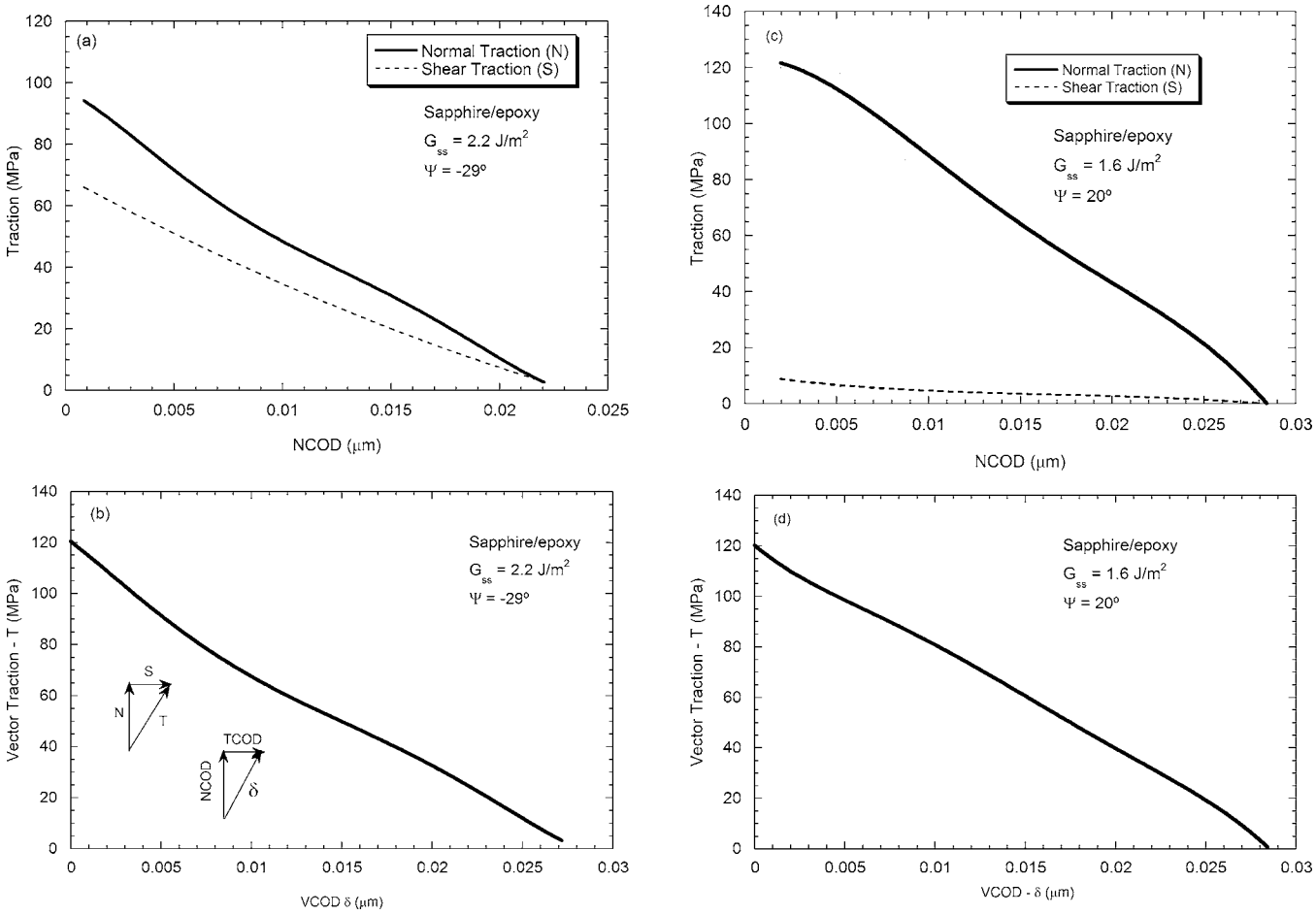


Fig. 6 Traction-separation laws for a sapphire/epoxy interface at G_{ss}=2.2, Ψ=−29 deg J/m^2: (*a*) normal and tangential tractions versus NCOD and (*b*) vector traction-separation law. Similar results for G_{ss}=1.6 J/m^2, Ψ=20 deg: (*c*) normal and tangential tractions versus NCOD and (*d*) vector traction-separation law.

bare sapphire surface (Fig. 10(*c*)). In comparing Figs. 10(*b*) and 10(*c*), it can be seen that the features of an original sapphire surface were much sharper than those that appeared on the epoxy fracture surface. This suggests that some epoxy may have been left behind on the sapphire fracture surfaces.

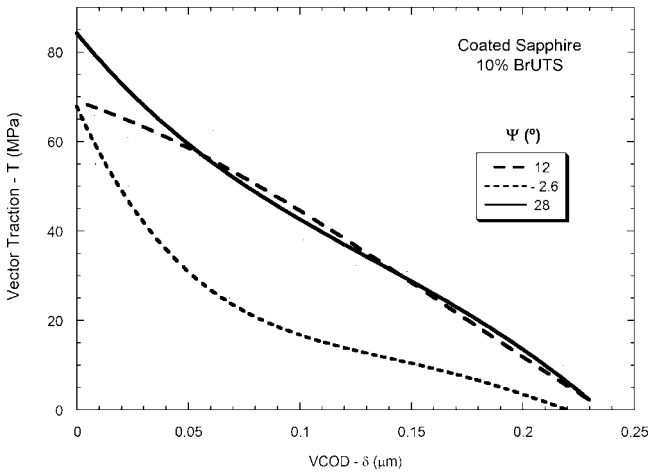


Fig. 7 Traction-separation in the cohesive zone for a 10% BrUTS coated sapphire/epoxy interface for G_{ss}=6.4 J/m^2, Ψ =−2.6 deg; G_{ss}=6.9 J/m^2, Ψ=12 deg; and G_{ss}=8.8 J/m^2, Ψ =28 deg

The epoxy fracture surface from a specimen with sapphire-coated 10% BrUTS under G_{ss}=6.4 J/m^2 and Ψ=−2.2 deg was also analyzed in the AFM. Figure 10(*d*) shows the three-dimensional topography of a 6 by 6 μm^2 region. The same pattern was observed for all the mode-mixes that were considered. It is clear from this figure that no tunneling was present. Instead, peaks can be seen all over the epoxy fracture surface. These peaks probably correspond to broken ligaments. There were approximately 20 ligaments per μm^2. The ligaments had an average height of 5 nm and a base area of 0.04 μm^2. Since the methyl groups only adhere weakly to the epoxy, each ligament could represent connections between the BrUTS and the epoxy. Considering the areal density of the methyl and bromine end groups, it is expected that there should be an average of 12 BrUTS molecules for each ligament. The elongation of each ligament would be due to the locally higher adhesion needed to break the ionic bonds between the epoxy and the BrUTS. Presumably failure then occurred via pullout of epoxy chains. This would tie in with the lower maximum traction levels and the much larger cutoff displacement in the traction separation law (Fig. 7) for the coated sapphire specimens. Even though the height of the ligaments was of the same order as the ridges formed by lateral spreading that was observed in the bare sapphire specimens, the density of the final points of contact before surface disconnection was much higher. It is also interesting to notice that this mechanism led to more stable and slower crack growth.

Portions of the sapphire fracture surfaces from the bare sap-

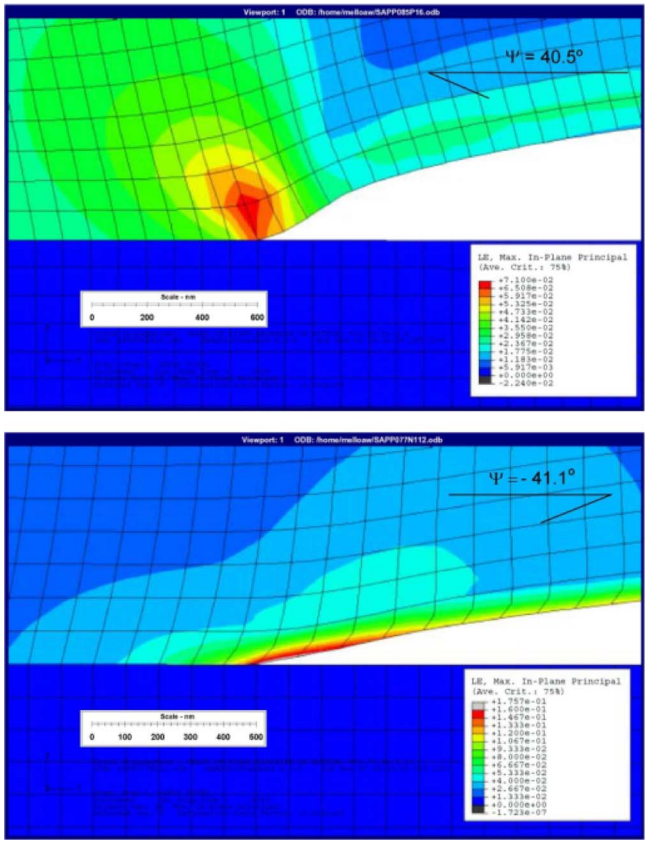


Fig. 8 Maximum in-plane principal strain contours for a growing interfacial crack in a sapphire/epoxy specimen for (*a*) G_{ss} =3.8 J/m^2, Ψ=41 deg, and (*b*) G_{ss}=4.0 J/m^2, Ψ=−42 deg. Displacements scaled by 5.

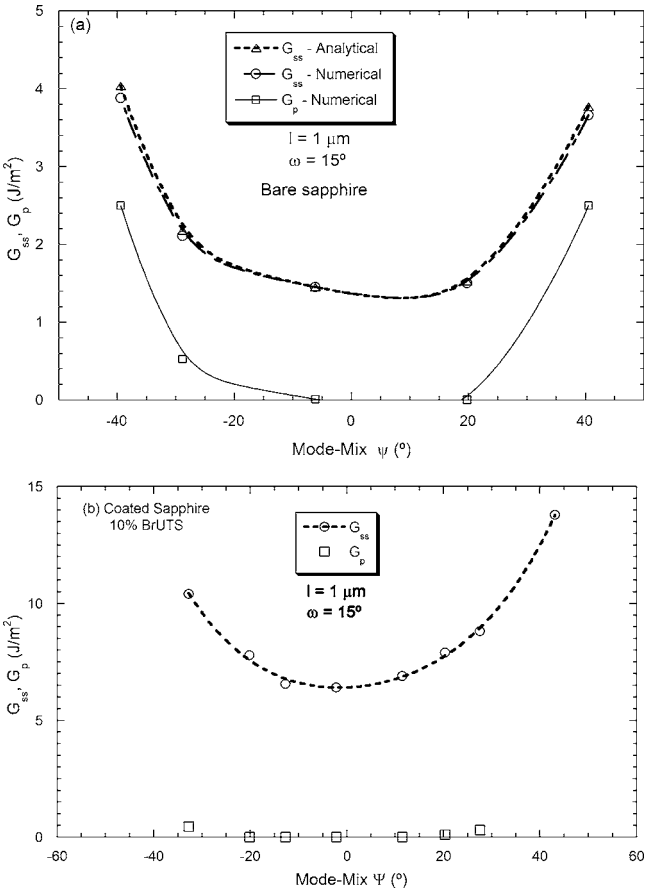


Fig. 9 Viscoplastic dissipation in sapphire/epoxy/aluminum specimens for (*a*) bare sapphire and (*b*) 10% BrUTS coated-sapphire

phire specimens were analyzed by ARXPS. The first one was taken from a region where the crack was subjected to G_{ss} =1.5 J/m^2 and Ψ=−6.2 deg. The second region had been subjected to positive shear at G_{ss}=1.6 J/m^2 and Ψ=19 deg, while the third one was loaded in negative shear at G_{ss}=4.0 J/m^2 at Ψ= −41 deg. ARXPS analysis was performed at three different angles (15 deg, 45 deg, and 75 deg) for each section. Based on material assumptions [11], these angles represent escape depths of approximately 0.6, 1.8, and 2.4 nm, respectively. The total scanned area for each sample was 100×100 μm, which covered a much larger area than the AFM scans.

Using the molecular stoichiometry for sapphire and the expected C–N ratio for epoxy, it was possible, using the depth profile of the atomic concentrations, to determine the relative concentration of epoxy on the sapphire fracture surfaces. It was found [11] that the epoxy was still present at a depth of 2.4 nm, which is the maximum probe depth. As a result, the thickness of the epoxy remaining on the sapphire could actually have been greater. At this probe depth, there was a slight reduction in epoxy remaining on the sapphire as the absolute mode-mix increased. The percentage of epoxy on the fracture surface was 8%, 6%, and 4% for Ψ=−6.2 deg, 19 deg, and −41 deg, respectively. At 1.8 nm, the epoxy concentration was about the same (≈10%) for all three mode-mixes. Closest to the fracture surface ($d_e \leq 0.6$ nm), there was a significant reduction in the amount of epoxy left on the sapphire fracture surface as the absolute value of mode-mix increased. The concentrations were 43%, 34%, and 12% for Ψ= −6.2 deg, 19 deg, and −41 deg, respectively. These results clearly indicate that the crack grew closer to the sapphire surface as the absolute value of the mode-mix increased. For a glass/epoxy interface [1], the glass fracture surface contained 88% epoxy under mode 1 dominant conditions (Ψ=−2 deg). At higher absolute mode-mixes (Ψ=−57 deg) the epoxy coverage of the glass fracture surface was 70%. Thus, the amount of epoxy left on the sapphire fracture surfaces was much lower, particularly at high mode-mixes.

In the case of glass/epoxy interfaces, it was suggested that epoxy strands were able to easily penetrate the gaps in the loose atomic structure of the amorphous glass. Thus, at 3 nm, the epoxy concentration was still about 20%. This was much larger than the average 6% epoxy coverage of the sapphire fracture surface. It is unlikely that epoxy strands could penetrate the more ordered and dense sapphire surfaces. The sapphire coated with 10% BrUTS also was analyzed by ARXPS. In this case, it was of interest to see if any bromine was present on the fracture surfaces. In order to detect the presence of bromine, two samples were analyzed using the largest aperture, larger pass energy, and longer acquisition times (5×) than are normally used in order to increase the sensitivity of the instrument. The binding energy that is most characteristic of bromine (70 eV) overlaps with that of aluminum. For this reason, a secondary peak (189/182 eV) was considered for the survey scan. It was found that there was a slightly lower amount of bromine present in the first 0.6 nm from the fracture surface compared to a newly coated surface. This could indicate that a portion of the bromine had been removed during fracture. However, the complete lack of bromine on the epoxy fracture surface leads to the conclusion that the lower concentration of bromine on the coated sapphire fracture surface was due to the presence of epoxy on that surface which obscured the bromine beyond the reach of the ARXPS probe depth. The epoxy coverage on the coated sapphire fracture surface was higher than the

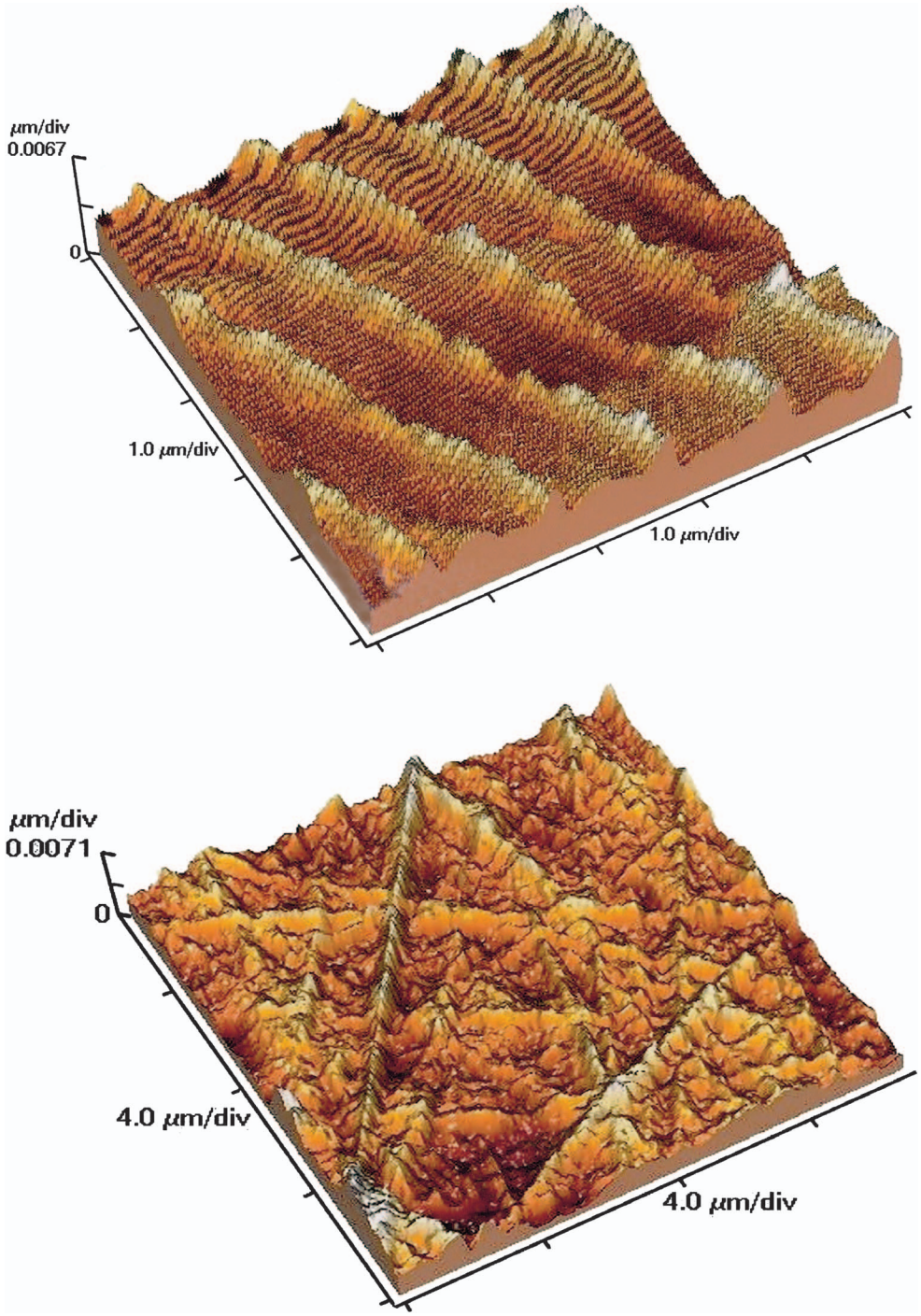


Fig. 10 AFM images of the epoxy fracture surfaces from sapphire/epoxy specimens. The crack was growing from bottom left to top right under (*a*) G_{ss}=1.5 J/m^2, Ψ=−6.1 deg and (*b*) G_{ss}=4.0 J/m^2, Ψ=−41 deg, for the bare sapphire. (*c*) AFM image of a virgin sapphire surface and (*d*) AFM image of the epoxy fracture surface from a 10% BrUTS coated-sapphire specimen, G_{ss}=6.4 J/m^2, Ψ=−2.2 deg.

amount found on the bare sapphire fracture surfaces. In addition, there was no mode-mix dependence to the epoxy coverage when the sapphire was coated. The coverage of epoxy was about 55% close to the surface ($d_e < 0.6$ nm), dropping to about 20% for $d_e > 1.8$ nm.

5 Conclusions

A series of experiments was conducted with sapphire/epoxy/aluminum sandwich specimens. In some experiments, the sapphire was coated with self-assembled monolayers consisting of a mixture of BrUTS (bromo-undecyltrichlorosilane) and DTS (dodecyltrichlorosilane). Cracks were introduced at the sapphire/epoxy interfaces and the specimens were loaded in a specially developed biaxial loading device. Crack opening interferometry was used to measure the NCOD near the crack front. The interfacial toughness during steady-state crack growth was determined from the applied displacements and geometry of the specimen. Plastic dissipation was obtained from finite element analyses that accounted for the viscoplastic nature of the epoxy and the interface between the epoxy and sapphire.

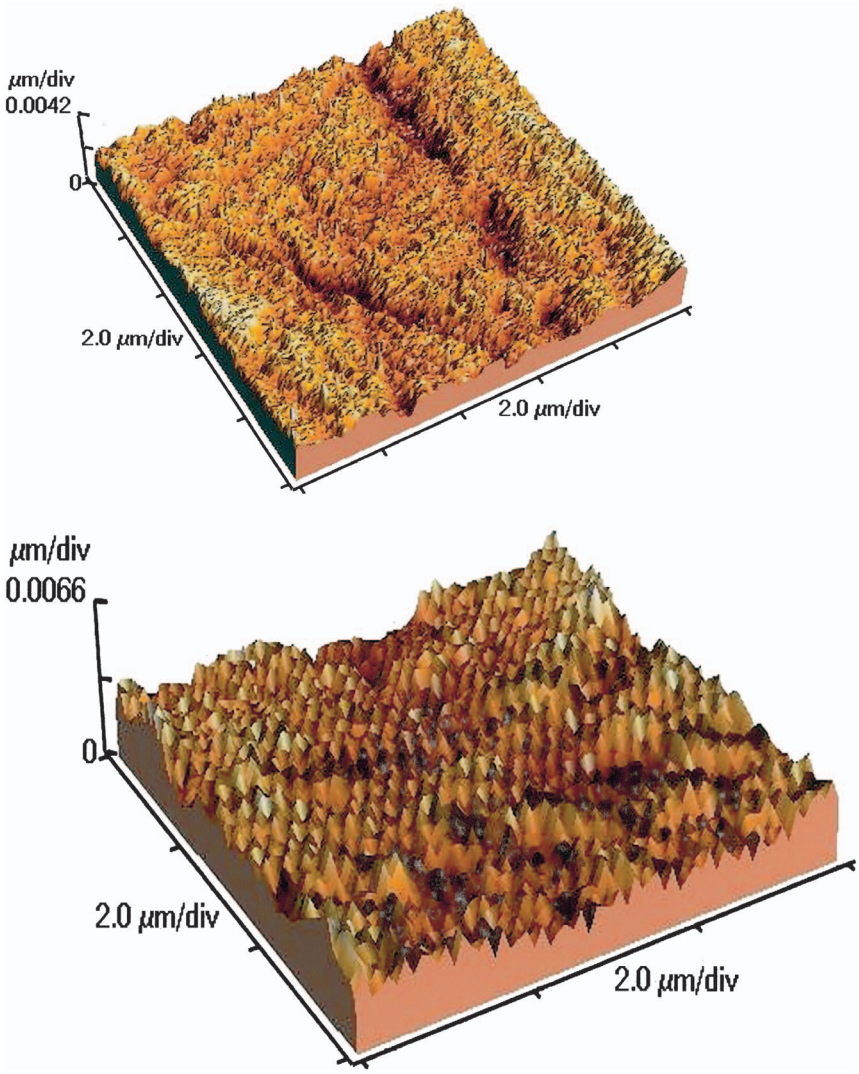


Fig. 10 (***Continued***).

The intrinsic toughness of a bare sapphire/epoxy interface was lower and its fracture envelope was broader than the corresponding glass/epoxy and quartz/epoxy values reported in previous publications. However, there was still a strong increase in toughness for higher mode-mixes.

SAMs composed of different mixtures of BrUTS and DTS were deposited on the sapphire surface using a systematic procedure in order to examine the role of the bromine terminal group on interfacial toughness. The use of monolayers with 10% and 55% BrUTS led, respectively, to an increase in the minimum toughness by factors of 4.5 and 5 compared to bare sapphire. In addition, the toughness envelope of these interfaces had a much sharper minimum. Steady-state crack growth could not be achieved for sapphire coated with a film composed of 70% BrUTS. In this case, void formation along the sapphire/epoxy interface occurred instead of crack growth.

Finite element analyses of steadily propagating cracks that used a viscoplastic model for the epoxy and traction-separation laws for the different interfaces showed that the cohesive zone size of the bare sapphire/epoxy interface was 0.35 μm. For bare sapphire, the traction-separation law between the surfaces inside the cohesive zone was determined to be approximately linear. The finite element analysis showed the presence of a plastic wake along the epoxy interface. The viscoplastic dissipation component of the overall toughness ranged from being negligible for mode 1 dominant loading to being the major component at high mode-mixes.

Similar analyses were conducted on sapphire specimens coated with 10% BrUTS. The cohesive zone size for this interface was about ten times larger than the ones for quartz and bare sapphire interfaces. The maximum tractions were lower and the VCODs at the crack tip were much larger. In addition, the traction-separation laws (and intrinsic toughness) were not independent of mode-mix. To the authors' knowledge, this is the first time such behavior has been observed at this scale. The increasing steady-state toughness with increasing shear component was almost completely accounted for by increases in the intrinsic toughness with very little viscoplastic dissipation outside the fracture process zone for the mode-mixes for which solutions could be obtained.

Analysis of the epoxy fracture surfaces with AFM showed that the crack grew by lateral spreading at the bare-sapphire/epoxy interfaces under mode 1 dominant loading. In this case, regularly spaced 5 nm high ridges were formed, with a distance between ridges of approximately 1 μm. For higher mode-mixes, the ridges were irregular, forming several branches in the crack front. The height of these ridges was about 10 nm. For sapphire coated with 10% BrUTS, AFM showed that the epoxy fracture surface was populated with randomly distributed peaks. These sharp peaks were probably broken ligaments corresponding to regions where

there was a strong bond between the epoxy and the BrUTS. There were about 20 ligaments per μm^2, which had an average height of 5 nm and a base area of 0.04 μm^2.

ARXPS analysis of the debonded bare sapphire surfaces showed that up to 43% of the surface remained covered with epoxy. At a depth of 2.4 nm, the epoxy coverage was about 6%, suggesting that the layer left behind might even have been thicker than 2.4 nm in some locations. For high absolute values of mode-mix, the crack grew closer to the sapphire surface, as indicated by an epoxy coverage of only 12% at $d_e=0.6$ nm. For deeper regions, 1.8 nm$<d_e<$2.4 nm, the thickness of the epoxy was reasonably independent of mode-mix, with coverage ranging from 10% to 6%, respectively. The amount of epoxy left on the sapphire fracture surfaces of coated-change with mode-mix. The portion of the fracture surface covered by epoxy was more than 55% closer the surface and about 20% for 1.8 nm$<d_e<$2.4 nm. Bromine was present in the sapphire fracture surface, further confirming that SAMs caused the crack path to run deeper in the epoxy or caused significant amounts of pull-out of epoxy chains.

While the SAMs considered in this study clearly increased the toughness of the sapphire epoxy interface and the reasons for this have been determined, it is not clear that adhesion control through modulation of the areal density of strongly interacting groups has been achieved. A volume fraction of 10% Br achieved a notable increase in toughness, which was not really further realized for 50% Br. Then, at 75% Br, it was not possible to obtain crack growth. These results suggest that more work is needed to fully understand how SAMs might be used for adhesion control.

Acknowledgment

The authors wish to acknowledge financial support from the National Science Foundation (Grant No. CMS-9978678) and the Brazilian Air Command (AWM). John Harmon was very helpful in establishing the protocol for coating the sapphire with self-assembled monolayers.

References

[1] Swadener, J. G., and Liechti, K. M., 1998, "Asymmetric Shielding Mechanisms in the Mixed-Mode Fracture of a Glass/Epoxy Interface," ASME J. Appl. Mech., **65**, pp. 25–29.

[2] Swadener, J. G., Liechti, K. M., and Lozanne, A. L., 1999, "The Intrinsic Toughness and Adhesion Mechanisms of a Glass/Epoxy Interface," J. Mech. Phys. Solids, **47**, pp. 223–258.

[3] Drzal, L. T., 1986, "The Interphase in Epoxy Composites," Adv. Polym. Sci., **75**, pp. 1–32.

[4] Drzal, L. T., 1990, "The Role of the Fiber/Matrix Interphase on Composite Interfaces," Vacuum, **41**, pp. 1615–1618.

[5] Winter, R. M., and Houston, J. E., 1998, "Interphase Mechanical Properties in an Epoxy-Glass Fiber Composites as Measured by Interfacial Microscopy," *Proceedings of the SEM Spring Conf. on Exp., and Applied Mechanics*, Houston, TX, June 1–3.

[6] Winter, R. M., and Houston, J. E., 1998, "Nanomechanical Properties of the Interphase in Polymer Composites as Measured by Interfacial Force Microscopy," presented at the Mater. Res. Soc. Spring Meeting, San Francisco, CA, April 13–17.

[7] Kojio, K., Ge, S., Takahara, A., and Kajiyama, T., 1998, "Molecular Aggregation State of *n*-Octadecyltrichlorosilane Monolayer Prepared at an Air/Water Interface," Langmuir, **14**, pp. 971–974.

[8] Zhuk, A. V., Evans, A. G., Hutchinson, J. W., and Whitesides, G. M., 1998, "The Adhesion Energy Between Polymer Thin Films and Self-Assembled Monolayers," J. Mater. Res., **13**, pp. 3555–3564.

[9] Kent, M. S., Reedy, E. D., Yim, H., Matheson, A., Sorenson, J., Hall, J., Schubert, K., Tallant, D., Garcia, M., Ohlhausen, T., and Assink, R., 2004, "Using Self-Assembling Monolayers to Study Crack Initiation in Epoxy/Silicon Joints," J. Mater. Res., **19**, pp. 1682–1695.

[10] Mello, A. W., and Liechti, K. M., 2004, "A Piezoelectric Biaxial Loading Device for Interfacial Fracture Experiments," Exper. Mech., **44**, 495–501.

[11] Mello, A. W., 2003, "Mode-Mixed Fracture Experiments on Quartz/Epoxy and Sapphire/Epoxy Interfaces," doctoral dissertation, Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Engineering Mechanics Laboratory Report EMRL 03/11.

[12] Eng, P. J., 2000, "Structure of the Hydrated α-Al_2O_3 0001) Surface," Science, **288**, pp. 1029–1033.

[13] Hutchinson, J. W., and Suo, Z., 1992, "Mixed-Mode Cracking in Layered Materials," Adv. Appl. Mech., **29**, pp. 63–191.

[14] Tvergaard, V., and Hutchinson, J. W., 1993, "The Influence of Plasticity on Mixed-Mode Interface Toughness," J. Mech. Phys. Solids, **41**, pp. 1119–1135.

Youn-Sha Chan
Department of Computer and Mathematical Sciences,
University of Houston—Downtown,
One Main Street,
Houston, TX 77002

Glaucio H. Paulino
Department of Civil and Environmental Engineering,
University of Illinois,
2209 Newmark Laboratory,
205 North Mathews Avenue,
Urbana, IL 61801

Albert C. Fannjiang
Department of Mathematics,
University of California,
Davis, CA 95616

Change of Constitutive Relations due to Interaction Between Strain-Gradient Effect and Material Gradation

For classical elasticity, the constitutive equations (Hooke's law) have the same functional form for both homogeneous and nonhomogeneous materials. However, for strain-gradient elasticity, such is not the case. This paper shows that for strain-gradient elasticity with volumetric and surface energy (Casal's continuum), extra terms appear in the constitutive equations which are associated with the interaction between the material gradation and the nonlocal effect of strain gradient. The corresponding governing partial differential equations are derived and their solutions are discussed.
[DOI: 10.1115/1.2041658]

1 Introduction

In this paper we investigate the constitutive relations for strain-gradient elasticity in both homogeneous and functionally graded materials (FGMs) modeled as nonhomogeneous materials. For classical linear elasticity, the constitutive relations between the Cauchy stresses τ_{ij} and strains ε_{ij} have the same form for both homogeneous and nonhomogeneous materials. That is,

$$\tau_{ij} = \lambda \varepsilon_{kk} \delta_{ij} + 2G\varepsilon_{ij}, \tag{1}$$

in which δ_{ij} is the Kronecker delta; the Lamé moduli λ and G can either be constant,

$$\lambda = \lambda_0 \quad \text{and} \quad G = G_0,$$

or they can be some functions of the material point $\mathbf{x}=(x,y,z)$,

$$\lambda \equiv \lambda(\mathbf{x}) \quad \text{and} \quad G \equiv G(\mathbf{x}).$$

While the form of the constitutive relations is the same for homogeneous or graded materials in classical elasticity, such is not the case for strain-gradient elasticity where extra terms are generated due to the interaction of strain-gradient effect and material gradation. More specifically, for homogeneous materials, the constitutive relations in strain-gradient elasticity are [1,2]

$$\tau_{ij} = \lambda \varepsilon_{kk} \delta_{ij} + 2G\varepsilon_{ij} + 2G\ell' \nu_k \partial_k \varepsilon_{ij}, \tag{2}$$

where ℓ' is a material characteristic length associated with surface energy gradient, $\partial_k = \partial/\partial x_k$ is a differential operator, and ν_k, $\nu_k \nu_k = 1$, $\partial_k \nu_k = 0$, is a director field. For nonhomogeneous materials, one can NOT simply replace the Lamé moduli λ and G in Eq. (2) by the respective functions $\lambda(\mathbf{x})$ and $G(\mathbf{x})$ anymore. The corresponding constitutive equation for nonhomogeneous materials are

$$\begin{aligned}\tau_{ij} = {} & \lambda(\mathbf{x}) \varepsilon_{kk} \delta_{ij} + 2G(\mathbf{x})\varepsilon_{ij} + \ell' \nu_k [\varepsilon_{ll} \partial_k \lambda(\mathbf{x}) + \lambda(\mathbf{x}) \partial_k \varepsilon_{ll}] \delta_{ij} \\ & + 2\ell' \nu_k [\varepsilon_{ij} \partial_k G(\mathbf{x}) + G(\mathbf{x}) \partial_k \varepsilon_{ij}].\end{aligned} \tag{3}$$

Comparing Eqs. (2) and (3), one can observe that there are some extra terms in (3), and those extra terms are essentially the sum of two types of product: the product of the material gradation function [$\lambda(\mathbf{x})$ or $G(\mathbf{x})$] and the gradient of the strains, or the product of the strains and the gradient of the gradation function. It is in this sense that the extra terms are generated by the interaction of strain-gradient effect and material gradation.

Material behavior is often described by differential equations, which are formulated according to the constitutive relations. Thus, the next concern shall be how the change of constitutive equations influences the governing partial differential equations (PDEs). For instance, in classical elasticity (the constitutive relations have the same functional form for both homogeneous and nonhomogeneous materials), the governing PDEs for nonhomogeneous materials are

$$\begin{aligned}& G(\mathbf{x}) \nabla^2 \mathbf{u} + [\lambda(\mathbf{x}) + G(\mathbf{x})] \nabla \nabla \cdot \mathbf{u} + (\nabla \mathbf{u} + \nabla \mathbf{u}^T) \nabla G(\mathbf{x}) \\ & \quad + (\nabla \cdot \mathbf{u}) \nabla \lambda(\mathbf{x}) = 0,\end{aligned} \tag{4}$$

where $\mathbf{u}$ is the displacement vector, and ∇, $\nabla\cdot$, and ∇^2 are the gradient, divergence, and Laplacian operators, respectively. Equation (4) can be considered as a perturbation of the familiar Navier-Cauchy equations for homogeneous materials

$$G_0 \nabla^2 \mathbf{u} + (\lambda_0 + G_0) \nabla \nabla \cdot \mathbf{u} = 0, \tag{5}$$

where G_0 and λ_0 are the Lamé constants. Comparing Eqs. (4) and (5), one can observe that the perturbation brings in only the lower (first) order of differential operators, while the highest (second) order of differential operators has been preserved. As one of the properties of second-order elliptic PDEs, the behavior of the solution mainly depends on the highest order of the differential operators (see [3], Chap. 6). Thus, the solution to PDEs (4) should have similar behavior as the solution to PDEs (5). What is the situation for strain-gradient elasticity? It turns out that for the case of strain-gradient theory applied to FGMs, the change of PDEs is also only pertinent to the lower order differential operators, and the solution to the governing PDEs are still dominated by the highest order differential operators. In order to tell a complete story, we need to derive the governing PDEs from the equilibrium equations, in which the Cauchy stresses τ_{ij}, the couple stresses μ_{kij}, and the total stresses σ_{ij} are all involved. Thus, we need to know all the constitutive relations between strains and each of the stress fields. In this work the derivation of constitutive relations in strain-gradient elasticity relies on the strain-energy density function W.

This paper presents a detailed derivation of the constitutive relations in strain-gradient elasticity and the corresponding govern-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received October 4, 2002; final manuscript received July 14, 2005. Assoc. Editor: H. Gao. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

ing PDEs. The paper is organized as follows. First, the strain-energy density function W is introduced, the constitutive relations are derived from first principles, and some remarks about admissibility of W are made. Then, the governing PDEs of strain-gradient elasticity for anti-plane shear problems and plane state problems are derived. The behavior of the solutions to the corresponding PDEs are discussed. Finally, some concluding remarks are given at the end of the paper.

2 Strain-Energy Density Function

2.1 Elasticity. In classical elasticity, the strain-energy density function has the well-known form

$$W = \tfrac{1}{2}\lambda(\mathbf{x})\varepsilon_{ii}\varepsilon_{jj} + G(\mathbf{x})\varepsilon_{ij}\varepsilon_{ji}, \tag{6}$$

where $\lambda(\mathbf{x})$ and $G(\mathbf{x})$ are the material parameters which are functions of position $\mathbf{x}$, and ε is the small deformation tensor

$$\varepsilon_{ij} = \tfrac{1}{2}(u_{i,j} + u_{j,i}) \tag{7}$$

with $\mathbf{u}$ denoting the displacement vector. The Cauchy stresses are given by Eq. (1), i.e.,

$$\tau_{ij} = \frac{\partial W}{\partial \varepsilon_{ij}} = \lambda(\mathbf{x})\varepsilon_{kk}\delta_{ij} + 2G(\mathbf{x})\varepsilon_{ij}. \tag{8}$$

In the case of homogeneous materials, λ and G are constants (Lamé constants) and the Cauchy stresses, derived from (6), is

$$\tau_{ij} = \frac{\partial W}{\partial \varepsilon_{ij}} = \lambda\varepsilon_{kk}\delta_{ij} + 2G\varepsilon_{ij}. \tag{9}$$

Notice that Eqs. (8) and (9) have the same functional form.

2.2 Gradient Elasticity. Casal's anisotropic grade-2 elasticity theory is used in this paper; as an analogue to the concept of the surface tension of liquid, two material constants, the volume strain-gradient term ℓ and the surface energy strain-gradient term ℓ', were introduced by Casal to characterize the internal and surface capillarity of the solid. The surface energy strain-gradient term ℓ' cannot exist alone (i.e., $\ell=0$ and $\ell'\neq 0$ is not an admissible configuration) because the strain-energy density function needs to be non-negative. The effect of the volume strain-gradient term ℓ is to shield the applied loads leading to crack stiffening, and the effect of the surface energy strain-gradient term ℓ' is to amplify the applied loads leading to crack compliance by increasing the energy release rate of the crack [4]. The ratio $\rho=\ell'/\ell$ has been investigated in detail by Fannjiang et al. [5].

The three-dimensional generalization of Casal's gradient-dependent anisotropic elasticity with volumetric and surface energy for nonhomogeneous materials leads to the following expression for the strain-energy density function:

$$\begin{aligned} W = {} & \tfrac{1}{2}\lambda(\mathbf{x})\varepsilon_{ii}\varepsilon_{jj} + G(\mathbf{x})\varepsilon_{ij}\varepsilon_{ji} + \tfrac{1}{2}\lambda(\mathbf{x})\ell^2(\partial_k\varepsilon_{ii})(\partial_k\varepsilon_{jj}) \\ & + \tfrac{1}{2}\ell'\nu_k\partial_k[\lambda(\mathbf{x})\varepsilon_{ii}\varepsilon_{jj}] + G(\mathbf{x})\ell^2(\partial_k\varepsilon_{ij})(\partial_k\varepsilon_{ji}) \\ & + \ell'\nu_k\partial_k[G(\mathbf{x})\varepsilon_{ij}\varepsilon_{ji}], \quad \ell > 0, \end{aligned} \tag{10}$$

where ℓ and ℓ' are two material characteristic lengths associated with volumetric and surface energy gradient terms, respectively. The terms associated with ℓ' have the meaning of surface energy. It is easy to see that, after integrating W over the material domain Ω and applying the divergence theorem with $\partial_k\nu_k=0$, the terms associated with ℓ' become surface integrals,[1] i.e.,

$$\begin{aligned} & \int_\Omega \left(\frac{1}{2}\ell'\nu_k\partial_k[\lambda(\mathbf{x})\varepsilon_{ii}\varepsilon_{jj}] + \ell'\nu_k\partial_k[G(\mathbf{x})\varepsilon_{ij}\varepsilon_{ji}]\right)dV \\ & = \ell'\int_{\partial\Omega}\left(\frac{1}{2}\lambda(\mathbf{x})(\varepsilon_{ii}\varepsilon_{jj})(\nu_k n_k) + G(\mathbf{x})(\varepsilon_{ij}\varepsilon_{ij})(\nu_k n_k)\right)dS, \end{aligned} \tag{11}$$

where n_k is the outward unit normal to the boundary $\partial\Omega$. By considering the particular case $\nu_k \equiv n_k$, the director field has the same direction as the outward unit normal to the boundary, the surface integral simply becomes

$$\ell'\int_{\partial\Omega}\left(\frac{1}{2}\lambda(\mathbf{x})(\varepsilon_{ii}\varepsilon_{jj}) + G(\mathbf{x})(\varepsilon_{ij}\varepsilon_{ij})\right)dS. \tag{12}$$

By definition, the Cauchy stresses τ_{ij}, the couple stresses μ_{kij}, and the total stresses σ_{ij} are

$$\begin{aligned} \tau_{ij} &= \partial W/\partial\varepsilon_{ij} \\ \mu_{kij} &= \partial W/\partial\varepsilon_{ij,k} \\ \sigma_{ij} &= \tau_{ij} - \partial_k\mu_{kij}. \end{aligned} \tag{13}$$

Using Eqs. (13) and (10), the constitutive equations for functionally graded materials are

$$\begin{aligned} \tau_{ij} = {} & \lambda(\mathbf{x})\varepsilon_{kk}\delta_{ij} + 2G(\mathbf{x})\varepsilon_{ij} + \ell'\nu_k[\varepsilon_{ll}\partial_k\lambda(\mathbf{x}) + \lambda(\mathbf{x})\partial_k\varepsilon_{ll}]\delta_{ij} \\ & + 2\ell'\nu_k[\varepsilon_{ij}\partial_k G(\mathbf{x}) + G(\mathbf{x})\partial_k\varepsilon_{ij}] \end{aligned} \tag{14}$$

$$\mu_{kij} = \ell'\nu_k\lambda(\mathbf{x})\varepsilon_{ll}\delta_{ij} + \ell^2\lambda(\mathbf{x})\partial_k\varepsilon_{ll}\delta_{ij} + 2\ell'\nu_k G(\mathbf{x})\varepsilon_{ij} + 2\ell^2 G(\mathbf{x})\partial_k\varepsilon_{ij} \tag{15}$$

$$\begin{aligned} \sigma_{ij} = {} & \lambda(\mathbf{x})(\varepsilon_{kk} - \ell^2\nabla^2\varepsilon_{kk})\delta_{ij} + 2G(\mathbf{x})(\varepsilon_{ij} - \ell^2\nabla^2\varepsilon_{ij}) \\ & - \ell^2[\partial_k\lambda(\mathbf{x})](\partial_k\varepsilon_{ll})\delta_{ij} - 2\ell^2[\partial_k G(\mathbf{x})](\partial_k\varepsilon_{ij}) \end{aligned} \tag{16}$$

2.3 Remarks. If the material is homogeneous, then the Lamé constants λ and G in Eq. (10) can be placed either before or after the differential operator $\partial_k=\partial/\partial x_k$. However, if the material is nonhomogeneous, then different positions of λ and G in Eq. (10) would lead to different strain-energy density functions. Thus, if one expresses the strain-energy density as

$$\begin{aligned} W_A \equiv {} & \tfrac{1}{2}\lambda(\mathbf{x})\varepsilon_{ii}\varepsilon_{jj} + G(\mathbf{x})\varepsilon_{ij}\varepsilon_{ji} + \tfrac{1}{2}\ell^2\partial_k[\lambda(\mathbf{x})\varepsilon_{ii}](\partial_k\varepsilon_{jj}) \\ & + \tfrac{1}{2}\ell'\nu_k\partial_k[\lambda(\mathbf{x})\varepsilon_{ii}\varepsilon_{jj}] + \ell^2\partial_k[G(\mathbf{x})\varepsilon_{ij}](\partial_k\varepsilon_{ji}) \\ & + \ell'\nu_k\partial_k[G(\mathbf{x})\varepsilon_{ij}\varepsilon_{ji}], \end{aligned} \tag{17}$$

then it is clear that by the product rule of derivative, W_A and W are different. Two other strain-energy density expressions can be obtained by placing $\lambda(\mathbf{x})$ and $G(\mathbf{x})$, the Lamé moduli associated with the surface characteristic length ℓ', in front of the differential operator ∂_k in Eqs. (10) and (17) [6]. We choose to work with W because it gives rise to an energy functional that is always positive-definite regardless of the material inhomogeneities $(\lambda(\mathbf{x}), G(\mathbf{x}))$ and the strain-gradient parameters $\ell, \ell' \geq 0$. When the material inhomogeneities are present and rough [i.e., the derivatives of $\lambda(\mathbf{x})$, $G(\mathbf{x})$ are sufficiently large] the other (three) energy functionals lose positive-definiteness, resulting in negative total energy of possibly arbitrary magnitudes. Thus, in this paper we restricted our consideration to the energy density W and derive the constitutive relations and the corresponding PDEs from it.

3 Plane State Problems

In this section we derive the governing (system of) PDEs of gradient elasticity for a plane problem in functionally graded ma-

[1] To get Eq. (11), one needs to specify the director field in the interior as well, namely, it has to be divergence free. If one allows non-divergence-free director field, then it is possible to have λ and G standing out of the partial derivative in the ℓ' terms of (10) and still representing surface energy.

terials from the strain-energy density function. The process is similar to the one for anti-plane shear case, however the algebra is more involved.

3.1 Constitutive Equations. From the definition of τ_{ij}, μ_{kij}, and σ_{ij} in Eq. (13), we have already obtained the (general plane) constitutive equations of gradient elasticity for FGMs in Eqs. (14)–(16). For homogeneous materials, the constitutive equations are [1,7]

$$\tau_{ij} = \lambda\varepsilon_{kk}\delta_{ij} + 2G\varepsilon_{ij} + \ell' \nu_k \partial_k(\lambda\varepsilon_{ll}\delta_{ij} + 2G\varepsilon_{ij}) \tag{18}$$

$$\mu_{kij} = \lambda\ell^2\partial_k\varepsilon_{ll}\delta_{ij} + 2G\ell'\nu_k\varepsilon_{ij} + \lambda\ell'\nu_k\varepsilon_{ll}\delta_{ij} + 2G\ell^2\partial_k\varepsilon_{ij} \tag{19}$$

$$\sigma_{ij} = \lambda\varepsilon_{kk}\delta_{ij} + 2G\varepsilon_{ij} - \ell^2\nabla^2(\lambda\varepsilon_{kk}\delta_{ij} + 2G\varepsilon_{ij}). \tag{20}$$

Comparing Eqs. (14)–(16) with (18)–(20), one notices that the couple stresses μ_{kij} in (15) and (19) take the same form. However, for the total stresses σ_{ij}, there are more terms in (16) than in (20), and those extra terms will confound the form of the governing (system of) PDEs.

For two-dimensional plane problems, the components of the strain tensor are given by

$$\varepsilon_{xx} = \frac{\partial u}{\partial x}, \quad \varepsilon_{yy} = \frac{\partial v}{\partial y}, \quad \varepsilon_{xy} = \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right), \quad \varepsilon_{xz} = \varepsilon_{yz} = \varepsilon_{zz} = 0. \tag{21}$$

The components of the stress fields for homogeneous materials are [7]

$$\sigma_{xz} = \sigma_{yz} = 0$$

$$\sigma_{xx} = (\lambda + 2G)\varepsilon_{xx} + \lambda\varepsilon_{yy} - (\lambda + 2G)\ell^2\nabla^2\varepsilon_{xx} - \lambda\ell^2\nabla^2\varepsilon_{yy}$$

$$\sigma_{yy} = (\lambda + 2G)\varepsilon_{yy} + \lambda\varepsilon_{xx} - (\lambda + 2G)\ell^2\nabla^2\varepsilon_{yy} - \lambda\ell^2\nabla^2\varepsilon_{xx} \tag{22}$$

$$\sigma_{xy} = \sigma_{yx} = 2G\varepsilon_{xy} - 2G\ell^2\nabla^2\varepsilon_{xy}$$

$$\sigma_{zz} = \lambda(\varepsilon_{xx} + \varepsilon_{yy}) - \lambda\ell^2\nabla^2(\varepsilon_{xx} + \varepsilon_{yy}),$$

and

$$\mu_{xxx} = (\lambda + 2G)\ell^2\partial_x\varepsilon_{xx} + \lambda\ell^2\partial_x\varepsilon_{yy}$$

$$\mu_{yxx} = -(\lambda + 2G)\ell'\varepsilon_{xx} - \lambda\ell'\varepsilon_{yy} + (\lambda + 2G)\ell^2\partial_y\varepsilon_{xx} + \lambda\ell^2\partial_y\varepsilon_{yy}$$

$$\mu_{xyy} = (\lambda + 2G)\ell^2\partial_x\varepsilon_{yy} + \lambda\ell^2\partial_x\varepsilon_{xx} \tag{23}$$

$$\mu_{yyy} = -(\lambda + 2G)\ell'\varepsilon_{yy} - \lambda\ell'\varepsilon_{xx} + (\lambda + 2G)\ell^2\partial_y\varepsilon_{yy} + \lambda\ell^2\partial_y\varepsilon_{xx}$$

$$\mu_{xxy} = \mu_{xyx} = 2G\ell^2\partial_x\varepsilon_{xy}$$

$$\mu_{yyx} = \mu_{yxy} = -2G\ell'\varepsilon_{xy} + 2G\ell^2\partial_y\varepsilon_{xy}.$$

For nonhomogeneous materials, the couple stresses μ_{kij} have the same form as in (23), except that the Lamé constants functional λ and G are not constants, they are functions of (x,y) according to the gradation of the material. The total stresses σ_{ij} have more terms than in (22) and they are

$$\sigma_{xz} = \sigma_{yz} = 0$$

$$\begin{aligned}\sigma_{xx} &= [\lambda(x,y) + 2G(x,y)](1 - \ell^2\nabla^2)\varepsilon_{xx} + \lambda(x,y)(1 - \ell^2\nabla^2)\varepsilon_{yy} \\ &\quad - \ell^2\{[\partial_x\lambda(x,y)]\partial_x(\varepsilon_{xx} + \varepsilon_{yy}) + [\partial_y\lambda(x,y)]\partial_y(\varepsilon_{xx} + \varepsilon_{yy})\} \\ &\quad - 2\ell^2\{[\partial_x G(x,y)]\partial_x\varepsilon_{xx} + [\partial_y G(x,y)]\partial_y\varepsilon_{xx}\}\end{aligned}$$

$$\begin{aligned}\sigma_{yy} &= [\lambda(x,y) + 2G(x,y)](1 - \ell^2\nabla^2)\varepsilon_{yy} + \lambda(x,y)(1 - \ell^2\nabla^2)\varepsilon_{xx} \\ &\quad - \ell^2\{[\partial_x\lambda(x,y)]\partial_x(\varepsilon_{xx} + \varepsilon_{yy}) + [\partial_y\lambda(x,y)]\partial_y(\varepsilon_{xx} + \varepsilon_{yy})\} \\ &\quad - 2\ell^2\{[\partial_x G(x,y)]\partial_x\varepsilon_{yy} + [\partial_y G(x,y)]\partial_y\varepsilon_{yy}\}\end{aligned} \tag{24}$$

$$\begin{aligned}\sigma_{xy} = \sigma_{yx} &= 2G(x,y)(\varepsilon_{xy} - \ell^2\nabla^2\varepsilon_{xy}) - 2\ell^2\{[\partial_x G(x,y)]\partial_x\varepsilon_{xy} \\ &\quad + [\partial_y G(x,y)]\partial_y\varepsilon_{xy}\}\end{aligned}$$

$$\begin{aligned}\sigma_{zz} &= \lambda(x,y)[(\varepsilon_{xx} + \varepsilon_{yy}) - \ell^2\nabla^2(\varepsilon_{xx} + \varepsilon_{yy})] - \ell^2\{[\partial_x\lambda(x,y)]\partial_x(\varepsilon_{xx} \\ &\quad + \varepsilon_{yy}) + [\partial_y\lambda(x,y)]\partial_y(\varepsilon_{xx} + \varepsilon_{yy})\}\end{aligned}$$

3.2 Governing System of PDEs. By imposing the equilibrium equations

$$\frac{\partial\sigma_{xx}}{\partial x} + \frac{\partial\sigma_{xy}}{\partial y} = 0 \quad \text{and} \quad \frac{\partial\sigma_{xy}}{\partial x} + \frac{\partial\sigma_{yy}}{\partial y} = 0, \tag{25}$$

and using Eqs. (21) and (24), one obtains the following system of PDEs:

$$\begin{aligned}&G(x,y)\nabla^2(1 - \ell^2\nabla^2)\mathbf{u} + [\lambda(x,y) + G(x,y)]\nabla(1 - \ell^2\nabla^2)\nabla\cdot\mathbf{u} + [(1 \\ &\quad - \ell^2\nabla^2)(\nabla\mathbf{u} + \nabla\mathbf{u}^T)]\nabla G(x,y) + [(1 - \ell^2\nabla^2)\nabla\cdot\mathbf{u}]\nabla\lambda(x,y) \\ &\quad - \ell^2\left[\left(\nabla\frac{\partial}{\partial x}\mathbf{u}\right)\nabla\frac{\partial G(x,y)}{\partial x} + \left(\nabla\frac{\partial}{\partial y}\mathbf{u}\right)\nabla\frac{\partial G(x,y)}{\partial y}\right. \\ &\quad \left. - \nabla[\nabla\lambda(x,y)\cdot\nabla\nabla\cdot\mathbf{u}]\right] - \ell^2\left(\frac{\partial}{\partial x}[(\nabla\nabla u)\nabla G(x,y)]\right. \\ &\quad \left. + \frac{\partial}{\partial y}[(\nabla\nabla v)\nabla G(x,y)] + (\nabla\nabla^2\mathbf{u})\nabla G(x,y)\right) = 0,\end{aligned} \tag{26}$$

where the boldface $\mathbf{u}$ denotes the displacement vector (u,v). Equation (26) is the most general form. In particular, if the moduli vary as a function of (x,y) and assume the exponential form

$$G \equiv G(x,y) = G_0 e^{\beta x + \gamma y}, \quad \lambda \equiv \lambda(x,y) = \frac{3 - \kappa}{\kappa - 1}G(x,y), \tag{27}$$

then the system of PDEs is

$$\begin{aligned}&\left(1 - \beta\ell^2\frac{\partial}{\partial x} - \gamma\ell^2\frac{\partial}{\partial y} - \ell^2\nabla^2\right)\left((\kappa+1)\frac{\partial^2 u}{\partial x^2} + (\kappa-1)\frac{\partial^2 u}{\partial y^2} + 2\frac{\partial^2 v}{\partial x\partial y}\right. \\ &\quad \left. + \beta(\kappa+1)\frac{\partial u}{\partial x} + \gamma(\kappa-1)\frac{\partial u}{\partial y} + \gamma(\kappa-1)\frac{\partial v}{\partial x} + \beta(3-\kappa)\frac{\partial v}{\partial y}\right) = 0,\end{aligned} \tag{28}$$

$$\begin{aligned}&\left(1 - \beta\ell^2\frac{\partial}{\partial x} - \gamma\ell^2\frac{\partial}{\partial y} - \ell^2\nabla^2\right)\left((\kappa-1)\frac{\partial^2 v}{\partial x^2} + (\kappa+1)\frac{\partial^2 v}{\partial y^2} + 2\frac{\partial^2 u}{\partial x\partial y}\right. \\ &\quad \left. + \gamma(3-\kappa)\frac{\partial u}{\partial x} + \beta(\kappa-1)\frac{\partial u}{\partial y} + \beta(\kappa-1)\frac{\partial v}{\partial x} + \gamma(\kappa+1)\frac{\partial v}{\partial y}\right) = 0,\end{aligned} \tag{29}$$

where $\kappa = 3 - 4\nu$ if plane strain is considered, $\kappa = (3-\nu)/(1+\nu)$ if it is a plane stress problem, and ν is the Poisson's ratio.

If G and λ are constants, then the homogeneous material case is recovered, and the system of PDEs (26) is reduced to

$$(1 - \ell^2\nabla^2)[G\nabla^2\mathbf{u} + (\lambda + G)\nabla\nabla\cdot\mathbf{u}] = 0, \tag{30}$$

which has been studied by Exadaktylos [7]. In the conventional classical linear elasticity (i.e., $\ell \to 0$), the system of PDEs (26) becomes (4). If G and λ take the form in (27), then (4) can be expressed as

$$\begin{aligned}&(\kappa+1)\frac{\partial^2 u}{\partial x^2} + (\kappa-1)\frac{\partial^2 u}{\partial y^2} + 2\frac{\partial^2 v}{\partial x\partial y} + \beta(\kappa+1)\frac{\partial u}{\partial x} + \gamma(\kappa-1)\frac{\partial u}{\partial y} \\ &\quad + \gamma(\kappa-1)\frac{\partial v}{\partial x} + \beta(3-\kappa)\frac{\partial v}{\partial y} = 0\end{aligned} \tag{31}$$

$$(\kappa-1)\frac{\partial^2 v}{\partial x^2}+(\kappa+1)\frac{\partial^2 v}{\partial y^2}+2\frac{\partial^2 u}{\partial x\partial y}+\gamma(3-\kappa)\frac{\partial u}{\partial x}+\beta(\kappa-1)\frac{\partial u}{\partial y}$$
$$+\beta(\kappa-1)\frac{\partial v}{\partial x}+\gamma(\kappa+1)\frac{\partial v}{\partial y}=0. \tag{32}$$

This system (31) and (32) has been studied by Konda and Erdogan [8]; for the homogeneous materials, they can be further simplified to Navier-Cauchy equations (5) for the elastic medium.

4 Anti-Plane Shear

In this section we derive the governing PDE of gradient elasticity for an anti-plane shear problem in functionally graded materials. It is worth mentioning that this type of problem has attracted the attention of several researchers, such as Vardoulakis et al. [2], Fannjiang et al. [5], Georgiadis [9], and Zhang et al. [10].

4.1 Constitutive Equations. In three-dimensional space, the displacement components are defined as

$$u_x\equiv u,\quad u_y\equiv v,\quad u_z\equiv w. \tag{33}$$

As in Eq. (7), strains are defined by

$$\varepsilon_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right), \tag{34}$$

where both the indices i and j run through x, y, and z. The strain-energy density function (for anti-plane shear) is

$$W=\tfrac{1}{2}\lambda\varepsilon_{ii}\varepsilon_{jj}+G\varepsilon_{ij}\varepsilon_{ji}+\ell^2G(\partial_k\varepsilon_{ij})(\partial_k\varepsilon_{ji})+\ell'\nu_k\partial_k(G\varepsilon_{ij}\varepsilon_{ji}). \tag{35}$$

We define the Cauchy stresses τ_{ij}, the couple stresses μ_{kij}, and the total stresses σ_{ij} according to equations in (13). Thus, the constitutive equations of gradient elasticity in anti-plane problems for homogeneous materials can be directly derived as [1,2]

$$\tau_{ij}=\lambda\varepsilon_{kk}\delta_{ij}+2G\varepsilon_{ij}+2G\ell'\nu_k\partial_k\varepsilon_{ij} \tag{36}$$

$$\mu_{kij}=2G\ell'\nu_k\varepsilon_{ij}+2G\ell^2\partial_k\varepsilon_{ij} \tag{37}$$

$$\sigma_{ij}=\lambda\varepsilon_{kk}\delta_{ij}+2G(\varepsilon_{ij}-\ell^2\nabla^2\varepsilon_{ij}). \tag{38}$$

For functionally graded materials the corresponding constitutive equations are

$$\tau_{ij}=\lambda(\mathbf{x})\varepsilon_{kk}\delta_{ij}+2G(\mathbf{x})\varepsilon_{ij}+2\ell'\nu_k[\varepsilon_{ij}\partial_kG(\mathbf{x})+G(\mathbf{x})\partial_k\varepsilon_{ij}] \tag{39}$$

$$\mu_{kij}=2\ell'\nu_kG(\mathbf{x})\varepsilon_{ij}+2\ell^2G(\mathbf{x})\partial_k\varepsilon_{ij} \tag{40}$$

$$\sigma_{ij}=\lambda(\mathbf{x})\varepsilon_{kk}\delta_{ij}+2G(\mathbf{x})(\varepsilon_{ij}-\ell^2\nabla^2\varepsilon_{ij})-2\ell^2[\partial_kG(\mathbf{x})](\partial_k\varepsilon_{ij}). \tag{41}$$

It is worth pointing out that in each of (39) and (41), there is an extra term with respect to (36) and (38), respectively. The extra terms will disappear if there is no material gradation. Thus, for homogeneous materials, Eqs. (39)–(41) will become the same as (36)–(38).

According to the relations in (36)–(38), each component of the stress fields for homogeneous materials can be written as [2]

$$\sigma_{xx}=\sigma_{yy}=\sigma_{zz}=0,\quad \sigma_{xy}=0$$
$$\sigma_{xz}=2G(\varepsilon_{xz}-\ell^2\nabla^2\varepsilon_{xz})\neq 0,\quad \sigma_{yz}=2G(\varepsilon_{yz}-\ell^2\nabla^2\varepsilon_{yz})\neq 0$$
$$\mu_{xxz}=2G\ell^2\partial_x\varepsilon_{xz},\quad \mu_{xyz}=2G\ell^2\partial_x\varepsilon_{yz}$$
$$\mu_{yxz}=2G(\ell^2\partial_y\varepsilon_{xz}-\ell'\varepsilon_{xz}),\quad \mu_{yyz}=2G(\ell^2\partial_y\varepsilon_{yz}-\ell'\varepsilon_{yz}). \tag{42}$$

For FGMs, from the relations in (39)–(41), each component of the stress fields is found to be

$$\sigma_{xx}=\sigma_{yy}=\sigma_{zz}=0,\quad \sigma_{xy}=0$$
$$\sigma_{xz}=2G(x,y)(\varepsilon_{xz}-\ell^2\nabla^2\varepsilon_{xz})-2\ell^2\{[\partial_xG(x,y)](\partial_x\varepsilon_{xz})+[\partial_yG(x,y)](\partial_y\varepsilon_{xz})\}\neq 0$$
$$\sigma_{yz}=2G(x,y)(\varepsilon_{yz}-\ell^2\nabla^2\varepsilon_{yz})-2\ell^2\{[\partial_xG(x,y)](\partial_x\varepsilon_{yz})+[\partial_yG(x,y)](\partial_y\varepsilon_{yz})\}\neq 0 \tag{43}$$
$$\mu_{xxz}=2G(x,y)\ell^2\partial_x\varepsilon_{xz},\quad \mu_{xyz}=2G(x,y)\ell^2\partial_x\varepsilon_{yz}$$
$$\mu_{yxz}=2G(x,y)(\ell^2\partial_y\varepsilon_{xz}-\ell'\varepsilon_{xz}),\quad \mu_{yyz}=2G(x,y)(\ell^2\partial_y\varepsilon_{yz}-\ell'\varepsilon_{yz}).$$

Again, comparing Eqs. (42) and (43), one notices that there are extra terms in the total stresses σ_{ij} of (43) due to the interaction of material gradation and the nonlocal effect of strain gradient. As the equilibrium equation only involves σ_{ij} [see Eq. (46)], the extra terms will complicate the governing PDE(s) a bit more. The couple stresses μ_{kij} in (42) and (43) assume the same form, except that G in (43) is not a constant, but rather a function reflecting the gradation of the material.

4.2 Governing PDE. For an anti-plane problem, the following relations hold:

$$u=0,\quad v=0,\quad w=w(x,y), \tag{44}$$

where u, v, and w denote the displacement components along the axes x, y, and z, respectively. The nontrivial strains are

$$\varepsilon_{xz}=\frac{1}{2}\frac{\partial w}{\partial x},\quad \varepsilon_{yz}=\frac{1}{2}\frac{\partial w}{\partial y}. \tag{45}$$

By imposing the equilibrium equation

$$\frac{\partial\sigma_{xz}}{\partial x}+\frac{\partial\sigma_{yz}}{\partial y}=0 \tag{46}$$

with the expressions σ_{xz} and σ_{yz} in (43), one obtains the following PDE:

$$\nabla G(x,y)\cdot(1-\ell^2\nabla^2)\,\nabla w+G(x,y)(1-\ell^2\nabla^2)\nabla^2w$$
$$-\ell^2\left(\nabla\frac{\partial G(x,y)}{\partial x}\cdot\nabla\frac{\partial w}{\partial x}+\nabla\frac{\partial G(x,y)}{\partial y}\cdot\nabla\frac{\partial w}{\partial y}+\nabla G(x,y)\cdot\nabla\nabla^2w\right)=0. \tag{47}$$

If G is an exponential function of both x and y,

$$G\equiv G(x,y)=G_0e^{\beta x+\gamma y}, \tag{48}$$

then the governing PDE is

$$\left(1-\beta\ell^2\frac{\partial}{\partial x}-\gamma\ell^2\frac{\partial}{\partial y}-\ell^2\nabla^2\right)\left(\nabla^2+\beta\frac{\partial}{\partial x}+\gamma\frac{\partial}{\partial y}\right)w=0. \tag{49}$$

In Table 1 we list the governing PDEs in anti-plane shear problems that correspond to different combinations of parameter ℓ and various material gradation of the shear modulus G.

5 Further Remarks

The conventional continuum mechanics theories have been used adequately when the length scale of the deformation field is much larger than the underlying micro-structure length scale of the material. As the two length scales become comparable, the material behavior at one point tends to be influenced more significantly by the neighboring material points. The criterion for adopting the strain gradient theory should depend on the experimental

Table 1 Governing PDEs in antiplane shear problems

Cases	Governing PDE	References
$\ell=0,\ G\equiv\text{const}$	Laplace equation: $\nabla^2 w=0$	Standard textbooks
$\ell=0,\ G\equiv G(y)=G_0 e^{\gamma y}$	Perturbed Laplace equation: $\left(\nabla^2+\gamma\frac{\partial}{\partial y}\right)w=0$	Erdogan and Ozturk [11]
$\ell=0,\ G\equiv G(x)=G_0 e^{\beta x}$	Perturbed Laplace equation: $\left(\nabla^2+\beta\frac{\partial}{\partial x}\right)w=0$	Chan et al. [12] Erdogan [13]
$\ell\neq 0,\ G\equiv\text{const}$	Helmholtz-Laplace equation: $(1-\ell^2\nabla^2)\nabla^2 w=0$	Vardoulakis et al. [2] Fannjiang et al. [5] Zhang et al. [10]
$\ell\neq 0,\ G\equiv G(y)=G_0 e^{\gamma y}$	$\left(1-\gamma\ell^2\frac{\partial}{\partial y}-\ell^2\nabla^2\right)\left(\nabla^2+\gamma\frac{\partial}{\partial y}\right)w=0$	Paulino et al. [14]
$\ell\neq 0,\ G\equiv G(x)=G_0 e^{\beta x}$	$\left(1-\beta\ell^2\frac{\partial}{\partial x}-\ell^2\nabla^2\right)\left(\nabla^2+\beta\frac{\partial}{\partial x}\right)w=0$	Chan et al. [15]
$\ell\neq 0$, general $G\equiv G(x,y)$	Eq. (47)	Not available

data, and there are many experiments indicating conventional continuum mechanics cannot lead to a satisfactory prediction of the material behavior as the two length scales mentioned above are comparable to each other. Experimental techniques related to strain gradient theory include micro-torsion [16], micro-bending [17], and micro-indentation [18], which can be associated to the parameter ℓ. However, the authors are not aware of experiments associated directly to ℓ', which indicates an area for further research.

The inhomogeneity of materials can be caused by many mechanisms in different length scales, such as the size and distribution of inclusions, the grain size of crystals, and the size of constituent atoms and molecules. Thus a constant ℓ' cannot describe these different length scales. Ideally $\ell'\equiv\ell'(\mathbf{x})$; however, here we consider the gradient parameters ℓ' and ℓ as constants.

6 Conclusion

In the conventional classical linear elasticity, one may derive the governing PDE(s) for nonhomogeneous materials by directly replacing the Lamé constants with the material gradation functions at the level of the constitutive equations. We have shown that this is not the case for strain gradient elasticity because extra terms may arise. These extra terms come from the interaction between the material gradation and the nonlocal effect of the strain gradient. Thus, the constitutive equations for nonhomogeneous materials are different from the ones for homogeneous materials under the consideration of strain gradient elasticity theory (Casal's continuum). The governing PDEs for nonhomogeneous materials are derived by means of the strain energy density function and the corresponding definitions of the stress fields (which have been presented in this paper).

Acknowledgment

Professor Paulino would like to thank the support from the National Science Foundation under Grant No. CMS 0115954 (Mechanics and Materials Program); and from the NASA-Ames "Engineering for Complex Systems Program," and the NASA-Ames Chief Engineer (Dr. Tina Panontin) through Grant No. NAG 2-1424. Professor Fannjiang acknowledges the support from the USA National Science Foundation (NSF) through Grant No. DMS-9971322 and UC Davis Chancellor's Fellowship. Dr. Y.-S. Chan thanks the support from the Applied Mathematical Science Research Program of the Office of Mathematical, Information, and Computational Sciences, U.S. Department of Energy, under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC.; he also thanks the grant (Grant No. W911NF-05-1-0029) from the U.S. Department of Defense, Army Research Office. An Organized Research Award and a Faculty Development Award from University of Houston—Downtown also are acknowledged by Chan.

References

[1] Exadaktylos, G., Vardoulakis, I., and Aifantis, E., 1996, "Cracks in Gradient Elastic Bodies with Surface Energy," Int. J. Fract., **79**(2), pp. 107–119.
[2] Vardoulakis, I., Exadaktylos, G., and Aifantis, E., 1996, "Gradient Elasticity with Surface Energy: Mode-III Crack Problem," Int. J. Solids Struct., **33**(30), pp. 4531–4559.
[3] Evans, L. C., 1998, *Partial Differential Equations*, American Mathematical Society, Providence, RI.
[4] Exadaktylos G., and Vardoulakis I., 1998, "Surface Instability in Gradient Elasticity with Surface Energy," Int. J. Solids Struct., **35**(18), pp. 2251–2281.
[5] Fannjiang, A. C., Chan, Y.-S., and Paulino, G. H., 2002, "Strain-gradient Elasticity for Mode III cracks: A Hypersingular Integrodifferential Equation Approach, " SIAM J. Appl. Math., **62**(3), pp. 1066–1091.
[6] Chan, Y.-S., 2001, "Hypersingular Integrodifferential Equations and Applications to Fracture Mechanics of Homogeneous and Functionally Graded Materials with Strain-Gradient Effects," Ph.D. thesis, University of California, Davis.
[7] Exadaktylos, G., 1998, "Gradient Elasticity with Surface Energy: Mode-I Crack Problem," Int. J. Solids Struct., **35**(5), pp. 421–456.
[8] Konda, N., and Erdogan, F., 1994, "The Mixed Mode Crack Problem in a Nonhomogeneous Elastic Medium," Eng. Fract. Mech., **47**(4), pp. 533–545.
[9] Georgiadis, H. G., 2003, "The Mode III Crack Problem in Microstructured Solids Governed by Dipolar Gradient Elasticity: Static and Dynamic Analysis," ASME J. Appl. Mech., **70**(4), pp. 517–530.
[10] Zhang, L., Huang, Y., Chen, J. Y., and Hwang, K. C., 1998, "The Mode III Full-field Solution in Elastic Materials with Strain Gradient Effects," Int. J. Fract., **92**(4), pp. 325–348.
[11] Erdogan, F., and Ozturk, M., 1992, "Diffusion Problems in Bonded Nonhomogeneous Materials with an Interface Cut," Int. J. Eng. Sci., **30**(10), pp. 1507–1523.
[12] Chan, Y.-S., Paulino, G. H., and Fannjiang, A. C., 2001, "The Crack Problem for Nonhomogeneous Materials Under Antiplane Shear Loading—A Displacement Based Formulation," Int. J. Solids Struct., **38**(17), pp. 2989–3005.
[13] Erdogan, F., 1985, "The Crack Problem for Bonded Nonhomogeneous Materials under Antiplane Ahear Loading," ASME J. Appl. Mech., **52**(4), pp. 823–828.
[14] Paulino, G. H., Chan, Y.-S., and Fannjiang, A. C., 2003, "Gradient Elasticity Theory for Mode III Fracture in Functionally graded Materials—Part I: Crack Perpendicular to the Material Gradation," ASME J. Appl. Mech., **70**(4), pp. 531–542.
[15] Chan, Y.-S., Paulino, G. H., and Fannjiang, A. C., 2006, "Gradient Elasticity Theory for Mode III Fracture in Functionally Graded Materials—Part II: Crack Parallel to the Material Gradation," ASME J. Appl. Mech., accepted for publication.
[16] Fleck, N. A., Muller, G. M., Asby, M. F., and Hutchinson, J. W., 1994, "Strain Gradient Plasticity: Theory and Experiments," Acta Metall. Mater., **42**, pp. 475–487.
[17] Stolken, J. S., and Evans, A. G., 1998, "A Microbend Test Method for the Plasticity Length Scale," Acta Mater., **46**, pp. 5109–5115.
[18] Nix, W. D., 1997, "Elastic and Plastic Properties of Thin Films on Substrates: Nanoindentation Techniques," Mater. Sci. Eng., A, **234/236**, pp. 37–44.

Vector *J*-Integral Analysis of Crack Interaction With Pre-existing Singularities

Lifeng Ma
Department of Engineering Mechanics,
Xi'an Jiaotong University,
Xi'an 710049, China
e-mail: malf@mail.xjtu.edu.cn

Tian Jian Lu
Department of Engineering Mechanics,
Xi'an Jiaotong University,
Xi'an 710049, China
and
Department of Engineering,
University of Cambridge,
Cambridge CB2 1PZ, UK
e-mail: tjlu@mail.xjtu.edu.cn

Alexander M. Korsunsky
Department of Engineering Science,
University of Oxford,
Oxford OX1 3PJ, UK
e-mail: Alexander.korsunsky@eng.ox.ac.uk

In this paper, the mechanics of a semi-infinite crack interacting with near crack-tip singularities (e.g., dislocations) in two-dimensional solids is investigated using the concept of potential energy release rate. The spontaneous relationship between the crack potential energy release rate and the well-known vector conservative integral J_i ($i=1,2$) is derived. It is demonstrated that J_1 and J_2 integrals are equally important in solving crack problems. This allows a more rational criterion to be proposed, based on the criterion of maximum energy release rate, to assess the so-called shielding/amplification effect on the crack tip due to the presence of the singularities. It is shown that the new criterion can not only assess the shielding/amplification effect under pure mode I or mode II remote loading, but also efficiently assess crack-singularity interaction under mixed mode remote loading. Simultaneously, it is found by re-examining the J_i integrals that there exists a simple but universal relation among the three values of the vector J_i integral corresponding separately to the contributions induced from the semi-infinite crack tip, the singularity, and the remote loading. Next, a multi-singularity-crack interaction model is addressed, and the closed-form solution is obtained. Finally, as an example, the problem of a single dislocation interacting with a main crack is solved to demonstrate the validity of the proposed model and the new criterion.
[DOI: 10.1115/1.2165243]

1 Introduction

The following three topics: (i) vector J_i ($i=1,2$) integrals, (ii) crack growth criteria, (iii) crack interacting with multisingularities, are important in both theoretical and applied fracture mechanics. At first view these three topics do not appear to be closely related to each other, but it will be shown later that this is actually not the case. Before discussing their intrinsic relationships, a brief literature review about their histories is presented in the following.

(i) *Vector J_i Integrals.* The well-known conservative vector J_i integrals for two-dimensional solids can be expressed as [1,2]

$$J_i = \int_\Gamma \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds \quad (i = 1,2) \tag{1.1}$$

where Γ denotes an arbitrary curve surrounding a whole crack as shown in Fig. 1(*a*), the variable s indicates the arclength, and $T_j = \sigma_{ji} m_i$ denotes the traction vector on Γ, defined with respect to the outward unit vector $\boldsymbol{m}$ normal to the curve.

Actually, the original concept of conservative integral was introduced by Eshelby [3] to characterize the generalized forces on singularities and/or inhomogeneities in elastic fields. When applied to analyze the condition for crack propagation, the concept of *J*-integral (i.e., the J_1 integral) was introduced by Rice [4] as

$$J = \int_\Gamma (w dx_2 - T_j \partial u_j / \partial x_1 ds) \tag{1.2}$$

where Γ denotes a curve surrounding the crack tip as shown in Fig. 1(*b*). The curve begins at an arbitrary point on the lower crack surface and ends at an arbitrary point on the upper crack surface (i.e., it traverses in the counterclockwise sense; see Fig. 1). Here, it should be pointed out that the coordinate x_1-axis selected in Fig. 1(*b*) actually requires it being parallel to the crack surfaces. Otherwise, one could not arbitrarily choose the integral starting point *A* and ending point *B*. This issue will be discussed in further detail in the following.

The *J*-integral defined in Eq. (1.2) has the physical interpretation of the rate of change of potential energy with respect to the incremental change of crack length along the crack line, i.e., the x_1-axis [4–6]. It has found widespread applications in various aspects of fracture mechanics, which motivates the study of its twin integral, i.e., the J_2 integral

$$J_2 = \int_\Gamma (- w dx_1 - T_j \partial u_j / \partial x_2 ds). \tag{1.3}$$

Unfortunately, J_2 integral does not have the merit that one can arbitrarily choose the starting and ending points like *A* and *B* for J_1 integral in Fig. 1(*b*). Just due to this reason, it has been branded as a path-dependent integral by many researchers and hence little attention to the application of J_2 integral is paid.

(ii) *Crack Growth Criteria.* A variety of criteria have been proposed for crack growth, in particular, for brittle cracks. To predict the growth of a crack subjected to a single mode load such as pure mode I, mode II, or mode III, the corresponding material fracture toughness K_{IC}, K_{IIC}, and K_{IIIC} have been defined. Subsequently, criteria for crack growth under mixed-mode loading within the framework of linear elastic fracture mechanics were proposed. The most widely accepted crack growth criteria include: the maximum tangential stress criterion, the maximum shear stress criterion, the minimum strain energy density theory (*S*-criterion), and the maximum energy release rate (MERR) criterion; see Erdogan and Sih [7], Otsuka et al. [8], Sih [9], Wu [10], Shen and Shen [11], Ma and Korsunsky [12], among others.

On one hand, the validity of crack growth prediction based on a single stress component should be questioned, since fracture ini-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received June 17, 2005; final manuscript received November 21, 2005. Review conducted by Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California–Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

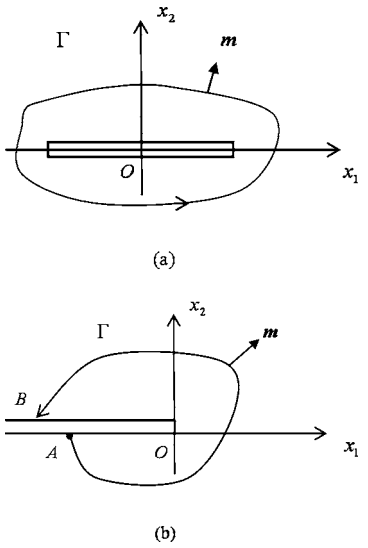


Fig. 1 Curve Γ: (*a*) surrounding a whole crack; (*b*) surrounding a crack tip

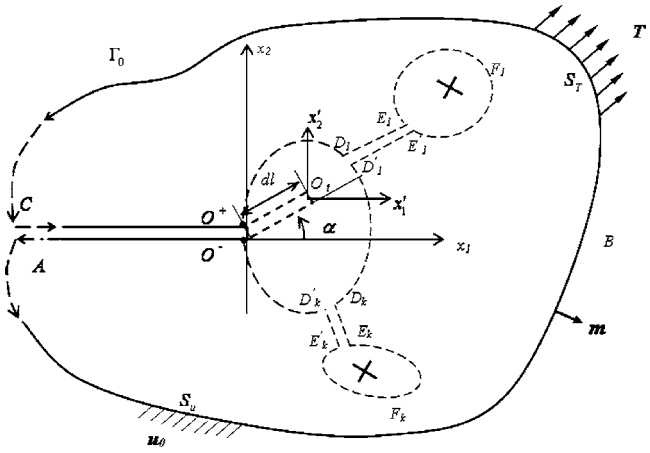


Fig. 3 A two-dimensional body containing a main crack and multisingularities

tiation and propagation may not be governed by a single component out of the six independent stress components, but rather by a combination that may play the role as a critical parameter. On the other hand, the strain energy density criterion (*S*-criterion) requires validation, as well as clarification of its physical basis. In contrast, the use of the MERR-criterion appears to be a logical application of the fundamental principle of fracture mechanics, as illustrated in detail later in this paper.

(iii) *A Semi-infinite Crack Interacting With Multisingularities.* Existing experimental evidence shows that the presence of multisingularities (e.g., dislocations, micro-inclusions, micro-cracks) near a crack tip could lead to the fracture toughness enhancement of brittle materials [13–18]. The two basic approaches employed to investigate this phenomenon are continuum modeling and discrete modeling [19,20]. Whilst the continuum models are beyond the scope of this paper and hence will not be discussed further, discrete modeling requires the study of a macrocrack interacting with singularities [13–17], as shown schematically in Fig. 2. The focus of these studies is mainly concerned with the influence of multisingularities on the stability of the crack tip. If the existence of multisingularities increases its stability, one says that the multisingularities have a "shielding effect" on the macro-crack tip; conversely, they have an "amplification effect." In order to assess this effect, various crack growth criteria have been adopted. Interestingly, in the published literature, the critical stress intensity factors K_{IC}, K_{IIC}, and K_{IIIC} have been more widely used than the other criteria previously mentioned. In this paper, we call the stress intensity factor criterion the SIF-criterion, and the MERR-criterion the *G*-criterion.

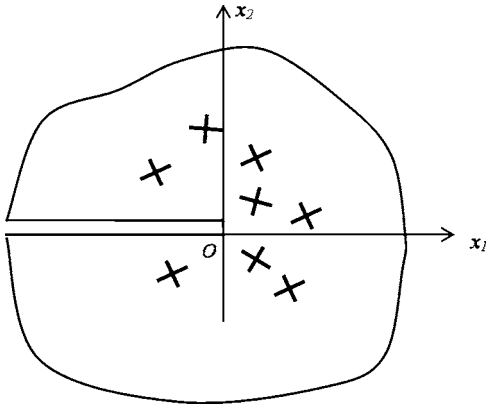


Fig. 2 A semi-infinite crack interacting with multisingularities

The aim of the present study is to establish a consistent approach and to formulate a rational criterion to scale the interaction effect on the macro-crack tip due to singularities. A multisingularity-crack interaction model is particularly investigated. In this procedure, the aforementioned three topics will spontaneously appear, and a new criterion will be proposed to assess the shielding effect. It is anticipated that this criterion can shed some new light on the problem of a macroscopic crack interacting with multisingularities.

The paper is constructed as follows. The relationship between the potential energy release rate and the vector J_i integral for the problem of multisingularity-crack interaction is derived in Sec. 2. Subsequently, the vector J_i integral redistribution due to the presence of multisingularity is investigated, and the *G*-criterion is proposed in Sec. 3. In Sec. 4, a general closed-form solution of the semi-infinite crack interacting with multisingularity is derived. Finally, to illustrate the interaction effect on the macro-crack accessed by the new criterion, an example concerned with a semi-infinite crack interacting with an edge dislocation is studied in Sec. 5.

2 The Relationship Between Potential Energy Release Rate and Vector J_i-Integral

In this section we present the derivation of the conservative vector J_i-integral to establish its relationship with the potential energy release rate for the problem of a semi-infinite crack interacting with multisingularity in two-dimensional solids. The derivation is aimed at the analysis of the interaction effect on the macro-crack tip due to the presence of multisingularities.

Consider the plane problem of a cracked body shown in Fig. 3. The body has perimeter Γ_0 ($\Gamma_0=\overline{O^-ABCO^+}$) enclosing an area *a*. Traction $\boldsymbol{T}$ acts on the part S_T of the boundary, while on the remaining part of the boundary, S_u, displacement boundary conditions are imposed in the form of the displacement vector $\mathbf{u}_0$. The coordinate system can always be chosen such that the origin lies at the crack tip *O*, even when the crack is advancing. It is assumed that the traction and displacement boundary conditions on Γ_0 are fixed, and that O^+ and O^- denote the points reached by approaching point *O* from the upper and lower crack surface, respectively. The potential energy Π of the body is given by

$$\Pi = \int_a w da - \int_{\Gamma_0} T_i u_i ds \tag{2.1}$$

where $w = \int_0^{\epsilon_{ij}} \sigma_{ij} d\varepsilon_{ij}$ is the strain energy density.

Now consider the tip of the crack (which also can be looked at as a singularity) and *all* the singularities undergoing a virtual displacement by an infinitesimal distance *dl* orientated at an arbitrary angle α as shown in Fig. 3. The total energy release rate due to such a virtual displacement can be written as

$$\frac{d\Pi}{dl} = \int_a \frac{dw}{dl} da - \int_{\Gamma_0} T_i \frac{du_i}{dl} ds \quad i = 1,2. \tag{2.2}$$

Since the coordinates are always attached at the macro-crack tip and the perimeter Γ_0 is fixed, it can be shown that

$$\frac{\partial x_i}{\partial l} = -n_i \quad \frac{d}{dl} = \frac{\partial}{\partial l} + \frac{\partial}{\partial x_i}\frac{\partial x_i}{\partial l} = \frac{\partial}{\partial l} - n_i \frac{\partial}{\partial x_i} \tag{2.3}$$

where $n_1 = \cos\alpha$ and $n_2 = \sin\alpha$. Equation (2.2) can be rewritten as

$$\frac{d\Pi}{dl} = \int_a \frac{\partial w}{\partial l} da - \int_a n_i \frac{\partial w}{\partial x_i} da - \int_{\Gamma_0} T_j \left(\frac{\partial u_j}{\partial l} - n_i \frac{\partial u_j}{\partial x_i} \right) ds. \tag{2.4}$$

Noting

$$\frac{\partial w}{\partial l} = \frac{\partial w}{\partial \varepsilon_{ij}} \frac{\partial \varepsilon_{ij}}{\partial l} = \sigma_{ij} \frac{\partial}{\partial x_j} \left(\frac{\partial u_i}{\partial l} \right) \quad \sigma_{ij,j} = 0 \tag{2.5}$$

and the principle of virtual work

$$\int_a \frac{\partial w}{\partial l} da = \int_a \sigma_{ij} \frac{\partial}{\partial x_j} \left(\frac{\partial u_i}{\partial l} \right) da = \int_{\Gamma_0} \sigma_{ij} n_j \left(\frac{\partial u_i}{\partial l} \right) ds = \int_{\Gamma_0} T_i \left(\frac{\partial u_i}{\partial l} \right) ds, \tag{2.6}$$

we obtain the following equation from Eq. (2.4)

$$\frac{d}{dl}\Pi = n_i \left(- \int_a \frac{\partial w}{\partial x_i} da + \int_{\Gamma_0} T_j \frac{\partial u_j}{\partial x_i} ds \right) = n_i \left(- \int_{\Gamma_0} w m_i ds + \int_{\Gamma_0} T_j \frac{\partial u_j}{\partial x_i} ds \right) = -n_i J_i \tag{2.7}$$

where

$$J_i = \int_{\Gamma_0} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds. \tag{2.8}$$

The classical introduction of the *J*-integral by Rice [4] focused the attention on the J_1 component which is responsible for controlling the forward propagation of the crack. However, in the context of multiaxial loading and crack turning it is more appropriate to consider both components of the vector *J*-integral, since that form of this energy integral provides exhaustive information for the application of various crack path determination criteria. We note that J_1 integral path is surrounding not only the crack tip but also all the singularities where one can arbitrarily choose the integral starting and ending points located separately on the upper and lower crack surfaces. Comparatively, the J_2 integral does not have such merit, whose integral must start from O^- to O^+ along a path such as Γ_0 surrounding the crack tip and all singularities shown in Fig. 3. The different property between J_1 and J_2 is attributed completely to the way the coordinates $x_1 - x_2$ are chosen in Fig. 3. For example, if we only rotate the coordinates by 20 deg in clockwise sense, the merit of J_1 vanishes; and if we clockwise rotate the coordinates by 90 deg in Fig. 3, the property of J_1 and J_2 will be exchanged. Therefore, the twin conservative J_1 and J_2 integrals may have an equal importance. This point will be exemplified further in the following derivation.

In this section, the multisingularity-crack interaction problem is investigated and a relationship between potential energy release rate and the vector J_i integral in a general sense is obtained. Built upon this deduction, the vector J_i integral redistribution due to the presence of multisingularity will be analyzed in the next section.

3 Vector J_i Integral Redistribution and *G*-Criteria

3.1 Vector J_i Integral Redistribution due to the Presence of Multisingularity. Let us start with Eq. (2.8). If no singularity is present within the closed contour in Fig. 3, it can be straightforwardly proved that [3]

$$\int_\Gamma \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds = \int_{\Gamma_0} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds + \int_{\Gamma_d} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds = 0 \tag{3.1}$$

where $\Gamma, \Gamma_0, \Gamma_d$ are the surrounding curves in Fig. 3

$$\Gamma = \overrightarrow{O^- ABCO^+ D_1 E_1 F_1 E_1' D_1' D_k E_k F_k E_k' D_k' O^-}$$

$$\Gamma_0 = \overrightarrow{O^- ABCO^+}$$

$$\Gamma_d = \overrightarrow{O^+ D_1 E_1 F_1 E_1' D_1' D_k E_k F_k E_k' D_k' O^-} .$$

It follows from Eq. (3.1) that

$$\int_{\Gamma_0} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds = - \int_{\Gamma_d} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds. \tag{3.2}$$

Letting $D_1 \to D_1', E_1 \to E_1', D_k \to D_k', E_k \to E_k'$ in Fig. 3, one arrives at

$$\int_{\Gamma_0} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds = \int_{C_t} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds + \sum_{k=1}^{N} \int_{C_k} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds \tag{3.3}$$

where

$$C_t = \overrightarrow{O^- D_k D_1 O^+}, \; C_k = \overrightarrow{E_k F_k E_k}$$

in the counterclockwise sense, with $k = 1, 2 \ldots, N$, *N* being the total number of singularities. Then:

$$J_i = \int_{\Gamma_0} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds = \int_{C_t} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds + \sum_{k=1}^{N} \int_{C_k} \left(w m_i - T_j \frac{\partial u_j}{\partial x_i} \right) ds. \tag{3.4}$$

Alternatively, the vector *J* integral can be expressed as

$$J_i = J_{i(t)} + \sum_{k=1}^{N} J_{i(k)} \tag{3.5}$$

where

$$J_{i(t)}=\int_{C_t}\left(wm_i-T_j\frac{\partial u_j}{\partial x_i}\right)ds \qquad J_{i(k)}=\int_{C_k}\left(wm_i-T_j\frac{\partial u_j}{\partial x_i}\right)ds. \tag{3.6}$$

Clearly, during the above manipulations $J_{1(t)}, J_{2(t)}, J_{1(k)}$, and $J_{2(k)}$ appear simultaneously and spontaneously in Eq. (3.5), which may be interpreted separately as the driving forces acting on the crack tip and the singularity k lying along the i-axis direction, with $i=1,2$ [2,3]. In fact, J_1 and J_2 in Eq. (3.5) are the resultant singularity driving forces acting on all singularities (including the crack tip) in the x_1- and x_2-direction, respectively. From another point of view, Eq. (3.5) clearly reveals that the remote loading represented by $J_{i(\infty)}(\equiv J_i)$ is redistributed and divided into $N+1$ parts near the crack tip due to the existence of the singularities. The first part of Eq. (3.5) is distributed at the crack tip (denoted by $J_{i(t)}$), and the other parts are distributed to other singularities.

It should be noted that when the remote loading $J_{i(\infty)}$ is fixed, Eq. (3.5) shows that if $\Sigma_{k=1}^{N}J_{i(k)}>0$, then $J_i>J_{i(t)}$. This implies that Eq. (3.5) can provide an explanation for the well-known phenomenon that the singularities near a crack tip shield the crack from remotely applied stresses. In addition, it should be pointed out that Eq. (3.5) represents a universal relation, valid for any singularity-crack interaction problem. Similarly, this relation also stands for macrocrack-microcrack interaction problems. Chen [21] found the relation for the case $i=1$ in Eq. (3.5) for macrocrack-microcrack interaction problems, assuming that the macrocrack propagates along the original direction. His result can be looked as a special case of Eq. (3.5).

G-Criterion. Consider now Eq. (2.7). Substitution of Eq. (3.5) into Eq. (2.7) leads to

$$\frac{d\Pi}{dl}=-(J_1\cos\alpha+J_2\sin\alpha)=-\left\{\left(J_{1(t)}+\sum_{k=1}^{N}J_{1(k)}\right)\cos\alpha+\left(J_{2(t)}+\sum_{k=1}^{N}J_{2(k)}\right)\sin\alpha\right\} \tag{3.7}$$

from which it follows that

$$J_{1(t)}\cos\alpha+J_{2(t)}\sin\alpha=\left(J_{1(\infty)}-\sum_{k=1}^{N}J_{1(k)}\right)\cos\alpha+\left(J_{2(\infty)}-\sum_{k=1}^{N}J_{2(k)}\right)\sin\alpha \tag{3.8}$$

where J_i has been substituted by $J_{i(\infty)}$. Equation (3.8) represents the crack tip driving forces projected in the α-direction, which is obtained by taking out the other singularity driving forces from the remote loads, also projected in the α-direction.

Irwin [22] introduced the symbol G to denote the rate of change of potential energy associated with crack advance. Equation (3.8) can therefore be written as

$$G=J_{1(t)}\cos\alpha+J_{2(t)}\sin\alpha=\sqrt{J_{1(t)}^2+J_{t(t)}^2}\cos(\phi_t-\alpha) \tag{3.9}$$

where $\tan\phi_t=J_{2(t)}/J_{1(t)}$. Note that $J_{i(t)}$ in Eq. (3.9) is independent of the possible crack propagation angle α (see Fig. 3). Consequently, Eq. (3.9) implies that there exists a maximum value of G, corresponding to a certain value of α. Using the *MERR*-criterion (see Refs. [10–12]), we obtain

$$G_{\max}=\sqrt{J_{1(t)}^2+J_{2(t)}^2} \quad \alpha=\phi_t=\arctan\frac{J_{2(t)}}{J_{1(t)}} \tag{3.10}$$

where $-\pi/2\leqslant\alpha\leqslant\pi/2$. It follows from Eq. (3.10) that: once $G_{\max}=\sqrt{J_{1(t)}^2+J_{2(t)}^2}\geqslant G_c$ (G_c is the material fracture toughness), crack tip will advance along angle $\alpha=\arctan(J_{2(t)}/J_{1(t)})$.

Assume that the loading applied at remote consisted of a combination of mode I and mode II loads expressed by $K_{\mathrm{I}(\infty)}$ and $K_{\mathrm{II}(\infty)}$, respectively. Let the corresponding stress intensity factors at the tip of the semi-infinite crack be denoted by $K_{\mathrm{I}(t)}$ and $K_{\mathrm{II}(t)}$. The remote loads and crack driving forces represented in terms of J_i-integral are associated with the stress intensity factors as [4,5]

$$J_{1(\infty)}=\frac{1-\nu^2}{E}[K_{\mathrm{I}(\infty)}^2+K_{\mathrm{II}(\infty)}^2]$$

$$J_{2(\infty)}=-\frac{2(1-\nu^2)}{E}K_{\mathrm{I}(\infty)}K_{\mathrm{II}(\infty)} \quad \text{(for plain strain)} \tag{3.11}$$

and

$$J_{1(t)}=\frac{1-\nu^2}{E}[K_{\mathrm{I}(t)}^2+K_{\mathrm{II}(t)}^2]$$

$$J_{2(t)}=-\frac{2(1-\nu^2)}{E}K_{\mathrm{I}(t)}K_{\mathrm{II}(t)} \quad \text{(for plain strain)} \tag{3.12}$$

where E and ν are the Young's modulus and Poisson ratio of the material, respectively.

Substitution of Eq. (3.12) into Eq. (3.10) leads to

$$G_{\max}=\frac{(1-\nu^2)}{E}\sqrt{K_{\mathrm{I}(t)}^4+K_{\mathrm{II}(t)}^4+6K_{\mathrm{I}(t)}^2K_{\mathrm{II}(t)}^2} \tag{3.13}$$

and

$$\alpha=\phi_t=-\arctan\frac{2K_{\mathrm{I}(t)}K_{\mathrm{II}(t)}}{K_{\mathrm{I}(t)}^2+K_{\mathrm{II}(t)}^2}. \tag{3.14}$$

If there are no singularities near the crack tip, the vector $J_{i(t)}$ is identical to the vector $J_{i(\infty)}$, and the corresponding MERR of the crack-tip can be written as

$$G_{\max(\infty)}=\sqrt{J_{1(\infty)}^2+J_{2(\infty)}^2}=\frac{(1-\nu^2)}{E}\sqrt{K_{\mathrm{I}(\infty)}^4+K_{\mathrm{II}(\infty)}^4+6K_{\mathrm{I}(\infty)}^2K_{\mathrm{II}(\infty)}^2} \tag{3.15}$$

with

$$\alpha_{(\infty)}=\phi_{t(\infty)}=\arctan\frac{J_{2(\infty)}}{J_{1(\infty)}}=-\arctan\frac{2K_{\mathrm{I}(\infty)}K_{\mathrm{II}(\infty)}}{K_{\mathrm{I}(\infty)}^2+K_{\mathrm{II}(\infty)}^2}. \tag{3.16}$$

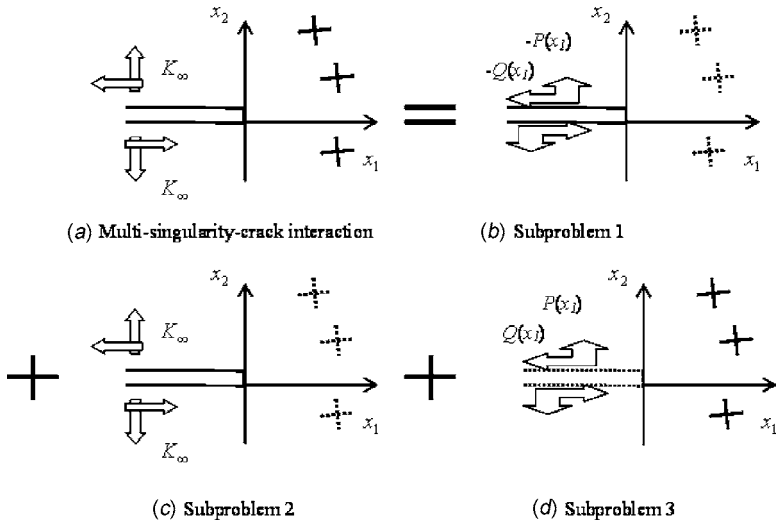
By virtue of the ratio of Eqs. (3.13) and (3.15), we now propose a new criterion to assess the shielding/amplification effect of the multisingularities on the crack tip, as

$$G_N=\frac{G_{\max}}{G_{\max(\infty)}}=\sqrt{\frac{K_{\mathrm{I}(t)}^4+K_{\mathrm{II}(t)}^4+6K_{\mathrm{I}(t)}^2K_{\mathrm{II}(t)}^2}{K_{\mathrm{I}(\infty)}^4+K_{\mathrm{II}(\infty)}^4+6K_{\mathrm{I}(\infty)}^2K_{\mathrm{II}(\infty)}^2}}. \tag{3.17}$$

It is assumed that the singularities have a shielding effect on the crack tip if $G_N<1$ whereas they have an amplification effect if $G_N>1$. Previously, criteria such as $K_{\mathrm{I}(t)}/K_{\mathrm{I}(\infty)}$ and $K_{\mathrm{II}(t)}/K_{\mathrm{II}(\infty)}$ have been used to scale the shielding effect (e.g., [16]), where $K_{\mathrm{I}(t)}/K_{\mathrm{I}(\infty)}<1$, (or $K_{\mathrm{II}(t)}/K_{\mathrm{II}(\infty)}<1$) means that singularities have a shield effect on crack tip, and vice versa.

4 Closed-Form Solution for a Semi-Infinite Crack Interacting with Singularities

Although we proposed a criterion to evaluate the shielding effect of multisingularity interaction on a crack tip in a general sense, the difficulty of obtaining both the analytical and numerical



(*a*) Multi-singularity-crack interaction (*b*) Subproblem 1

(*c*) Subproblem 2 (*d*) Subproblem 3

Fig. 4 Decomposition of multisingularity-crack interaction probelem

solutions for this problem is obvious. In the literature, the problem of a semi-infinite crack interacting with a discrete singularity has been partially studied (see, e.g., Refs. [14,23]). In the following, we will study the more complicated multisingularity-crack interaction problems and try to find a general closed-form solution for this problem.

Consider a semi-infinite crack, situated at $x_1 \leqslant 0$ and $x_2=0$, interacting with multisingularities as shown in Fig. 4(*a*). The singularities are situated at s_i $(i=1,\ldots,N)$. The material is assumed to be linear elastic, characterized by Young's modulus E and Poisson ratio ν. At infinity, the loading condition is $K_\infty=K_{\mathrm{I}(\infty)}+iK_{\mathrm{II}(\infty)}$, and the crack surfaces are traction-free.

The original interaction problem shown in Fig. 4(*a*) can be decomposed into three sub-problems as shown in Figs. 4(*b*)–4(*d*).

The stresses and displacements for a homogeneous media under plane deformation can be represented by two standard Muskhelishvili complex potentials $\phi(z)$ and $\psi(z)$, or $\Phi(z)$ and $\Omega(z)$ which are related to $\phi(z)$ and $\psi(z)$ as $\Phi(z)=\phi'(z)$ and $\Omega(z)=[z\phi'(z)+\psi(z)]'$, as follows

$$\sigma_{xx}+\sigma_{yy}=2[\Phi(z)+\overline{\Phi(z)}]$$

$$\sigma_{yy}+i\sigma_{xy}=\overline{\Phi(z)}+\Omega(z)+(\bar{z}-z)\Phi'(z)$$

$$-2i\mu\frac{\partial}{\partial x}(u_y+iu_x)=\kappa\overline{\Phi(z)}-\Omega(z)-(\bar{z}-z)\Phi'(z) \qquad (4.1)$$

where $z=x+iy$, $x=x_1$, $y=x_2$, $\kappa=3-4\nu$ for plane strain. By virtue of Eq. (4.1), each subproblem in Fig. 4 can be separately solved.

For subproblem 1, the semi-infinite crack is subjected to unknown pseudo-tractions $-[P(x_1)+iQ(x_1)]$, $x_1\in(-\infty,0)$. Using crack surface traction condition and the second equation of Eq. (4.1), one finds

$$\bar{\Phi}(x^-)+\Omega(x^+)=(\sigma_{yy}+i\sigma_{xy})^+=-[P(x)+iQ(x)] \quad x\in(-\infty,0) \qquad (4.2)$$

$$\bar{\Phi}(x^+)+\Omega(x^-)=(\sigma_{yy}+i\sigma_{xy})^-=-[P(x)+iQ(x)] \quad x\in(-\infty,0). \qquad (4.3)$$

It follows from Eqs. (4.2) and (4.3) that

$$\Omega(x^+)-\bar{\Phi}(x^+)=\Omega(x^-)-\bar{\Phi}(x^-) \quad x\in(-\infty,0). \qquad (4.4)$$

By the standard analytic continuation arguments, one writes

$$\Omega(z)=\bar{\Phi}(z) \quad z=x+iy \quad x,y\in(-\infty,+\infty). \qquad (4.5)$$

The combination of Eqs. (4.2) and (4.3) results in

$$[\bar{\Phi}(x^+)+\Omega(x^+)]+[\bar{\Phi}(x^-)+\Omega(x^-)]=-2[P(x)+iQ(x)] \quad x\in(-\infty,0) \qquad (4.6)$$

which is a standard Riemann-Hilbert problem.

For the problem of concern, the branch function can be chosen as

$$X(z)=\frac{1}{\sqrt{z}} \qquad (4.7)$$

Then, Eq. (4.6) can be rewritten as

$$\frac{[\bar{\Phi}(x^+)+\Omega(x^+)]}{X(x^+)}-\frac{[\bar{\Phi}(x^-)+\Omega(x^-)]}{X(x^-)}=-\frac{2[P(x)+iQ(x)]}{X(x^+)}. \qquad (4.8)$$

It follows that

$$[\bar{\Phi}(z)+\Omega(z)]=\frac{1}{\sqrt{z}}\left\{\frac{1}{2\pi i}\int_{-\infty}^{0}\frac{-2i[P(x)+iQ(x)]\sqrt{|t|}}{t-z}dt+P_0(z)\right\} \qquad (4.9)$$

where $P_0(z)$ is a polynomial. Using Eq. (4.5), one gets

$$\Omega(z)=-\frac{1}{\sqrt{z}}\left\{\frac{1}{2\pi}\int_{-\infty}^{0}\frac{[P(x)+iQ(x)]\sqrt{|t|}}{t-z}dt+P_0(z)\right\} \qquad (4.10)$$

$$\Phi(z)=-\frac{1}{\sqrt{z}}\left\{\frac{1}{2\pi}\int_{-\infty}^{0}\frac{[P(x)-iQ(x)]\sqrt{|t|}}{t-z}dt+\overline{P_0}(z)\right\}. \qquad (4.11)$$

In this subproblem, clearly, $P_0(z)=0$.

For subproblem 2 of Fig. 4, the semi-infinite crack is subjected to remote loading $K_\infty=K_{\mathrm{I}(\infty)}+iK_{\mathrm{II}(\infty)}$, whose solution can be directly derived by virtue of Eqs. (4.10) and (4.11). That is:

$$\Omega(z)=\frac{K_\infty}{\sqrt{8\pi z}} \qquad \Phi(z)=\frac{\overline{K_\infty}}{\sqrt{8\pi z}}. \tag{4.12}$$

For subproblem 3, the multisingularities are located at s_i $(i=1,\ldots,N)$ in an infinite solid which does not contain any other defect or crack. Here we denote the potentials for the multisingularities in an infinite solid as $\Phi_0(z)$ and $\Omega_0(z)$.

Superposing the potentials for the three subproblems together, one obtains the potentials for the original problem in Fig. 4(0), as

$$\Omega(z)=\frac{1}{\sqrt{z}}\left\{\frac{1}{2\pi}\int_{-\infty}^{0}\frac{-[P(t)+iQ(t)]\sqrt{|t|}}{t-z}dt+\frac{K_\infty}{\sqrt{8\pi}}\right\}+\Omega_0(z)$$

$$\Phi(z)=\frac{1}{\sqrt{z}}\left\{\frac{1}{2\pi}\int_{-\infty}^{0}\frac{-[P(t)-iQ(t)]\sqrt{|t|}}{t-z}dt+\frac{\overline{K_\infty}}{\sqrt{8\pi}}\right\}+\Phi_0(z). \tag{4.13}$$

Finally, from the traction-free condition along the crack surfaces, the Muskhelishvili complex potentials for the original problem are obtained as

$$\Omega(z)=\frac{1}{\sqrt{z}}\left\{\frac{1}{2\pi}\int_{-\infty}^{0}\frac{-[\overline{\Phi_0}(t)+\Omega_0(t)]\sqrt{|t|}}{t-z}dt+\frac{K_\infty}{\sqrt{8\pi}}\right\}+\Omega_0(z)$$

$$\Omega(z)=\frac{1}{\sqrt{z}}\left\{\frac{1}{2\pi}\int_{-\infty}^{0}\frac{-[\Phi_0(t)+\overline{\Omega_0}(t)]\sqrt{|t|}}{t-z}dt+\frac{\overline{K_\infty}}{\sqrt{8\pi}}\right\}+\Phi_0(z). \tag{4.14}$$

The singularities in a general sense can represent dislocations, inhomogeneities, concentrated forces, and concentrated moments, to name just a few. Whatever they are, once the complex potentials $\Omega_0(z)$ and $\Phi_0(z)$ are known, the multisingularity-crack interaction potentials are given by Eq. (4.14). Then the stress and displacement fields can be deduced by inserting Eq. (4.14) into Eq. (4.1), from which the stress intensity factors are straightforwardly obtained as

$$K_t=K_{\mathrm{I}(t)}+iK_{\mathrm{II}(t)}=\lim_{x\to 0^+}\sqrt{2\pi x}[\sigma_{yy}(x)+i\sigma_{xy}(x)]$$

$$=\lim_{x\to 0^+}\sqrt{2\pi x}[\bar{\Phi}(x)+\Omega(x)]. \tag{4.15}$$

Further, the appraisement on the shielding effect can be performed according to Eq. (3.17). Generally, the potentials for a discrete singularity in an infinite solid can be written as [14,23]

$$\Phi_{0d}(z)=\sum_{m=1}^{M}\frac{A_m}{(z-s)^m} \qquad \Omega_{0d}(z)=\sum_{m=1}^{M}\frac{B_m}{(z-s)^m}. \tag{4.16}$$

Using the superposition principle, the potentials of the multisingularities in an infinite solid mentioned in subproblem 3 can be obtained by directly adding the corresponding singularity's potentials together.

5 An Example: Dislocation-Crack Interaction Problem

5.1 Analytical Solution for Dislocation-Crack Interaction Problem. Consider a single edge dislocation of Burgers vector $\mathbf{b}(b\cos\psi, b\sin\psi)$, located at $s=r_0e^{i\theta_0}$ relative to a semi-infinite crack as shown in Fig. 5.

The Muskhelishvili complex potentials for a dislocation in an isotropic infinite plane solid are given by [23]

$$\Phi_0(z)=\frac{B}{(z-s)} \qquad \Omega_0(z)=\frac{B(\bar{s}-s)}{(z-s)^2}+\frac{\bar{B}}{(z-s)} \tag{5.1}$$

where

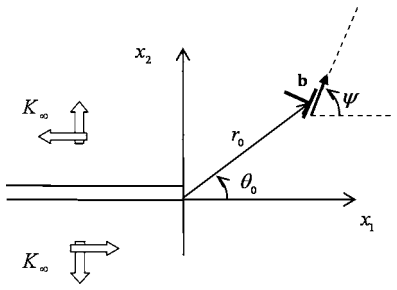


Fig. 5 Coordinates of the dislocation-crack system

$$B=\frac{\mu}{\pi i(\kappa+1)}(b_x+ib_y).$$

Substitution of Eq. (5.1) into Eq. (4.14) results in

$$\Omega(z)=\frac{1}{\sqrt{z}}\left\{\frac{K_\infty}{\sqrt{8\pi}}-\frac{B}{4}\frac{[(s-\bar{s})\sqrt{s}+z\sqrt{s}-2\sqrt{z}(\bar{s}-s)-z\bar{s}\sqrt{1/s}]}{(s-z)^2}\right.$$
$$\left.-\frac{\bar{B}}{2}\left[\frac{\sqrt{z}}{z-\bar{s}}-\frac{\sqrt{z}}{z-s}+\frac{\sqrt{s}}{s-z}+\frac{\sqrt{\bar{s}}}{\bar{s}-z}\right]\right\}$$

$$\Phi(z)=\frac{1}{\sqrt{z}}\left\{\frac{\bar{K}_\infty}{\sqrt{8\pi}}-\frac{\bar{B}}{4}\frac{[(\bar{s}-s)\sqrt{\bar{s}}+z\sqrt{\bar{s}}+2\sqrt{z}(s-\bar{s})-zs\sqrt{1/\bar{s}}]}{(\bar{s}-z)^2}\right.$$
$$\left.-\frac{B}{2}\left[\frac{\sqrt{z}}{z-\bar{s}}-\frac{\sqrt{z}}{z-s}+\frac{\sqrt{s}}{s-z}+\frac{\sqrt{\bar{s}}}{\bar{s}-z}\right]\right\}. \tag{5.2}$$

From Eq. (5.2), the stress field in front of the crack tip is obtained as

$$\sigma_{yy}(x)+i\sigma_{xy}(x)=\frac{1}{\sqrt{x}}\left\{\frac{K_\infty}{\sqrt{2\pi}}-\frac{B}{2}\frac{\left[(s-\bar{s})\sqrt{s}+x\sqrt{s}-x\bar{s}\sqrt{\frac{1}{s}}\right]}{(s-x)^2}\right.$$
$$\left.-\bar{B}\left[\frac{\sqrt{s}}{(s-x)}+\frac{\sqrt{\bar{s}}}{(\bar{s}-x)}\right]\right\}. \quad x\in(0,\infty). \tag{5.3}$$

Then, we can get the SIF of the crack tip as

$$K_t=\lim_{x\to 0}\sqrt{2\pi x}[\sigma_{yy}(x)+i\sigma_{xy}(x)]=K_\infty-\sqrt{\frac{\pi}{2}}\frac{B(s-\bar{s})}{s\sqrt{s}}$$
$$-\bar{B}\sqrt{2\pi}\left[\frac{1}{\sqrt{s}}+\frac{1}{\sqrt{\bar{s}}}\right]. \tag{5.4}$$

It immediately follows that the change in SIF, ΔK, induced purely by the presence of the dislocation is given by

$$\Delta K=\Delta K_{\mathrm{I}}+i\Delta K_{\mathrm{II}}=(K_t-K_\infty)=-\sqrt{\frac{\pi}{2}}\frac{B(s-\bar{s})}{s\sqrt{s}}$$
$$-\bar{B}\sqrt{2\pi}\left[\frac{1}{\sqrt{s}}+\frac{1}{\sqrt{\bar{s}}}\right]. \tag{5.5}$$

Using

$$B=\frac{\mu}{\pi i(\kappa+1)}(b_x+ib_y)=\frac{\mu}{\pi(\kappa+1)}be^{i(\psi-\pi/2)}$$

and $s=r_0e^{i\theta_0}$, Eq. (5.5) can be expressed in the polar coordinate system as

$$\Delta K_{\mathrm{I}}+i\Delta K_{\mathrm{II}}=-\frac{\mu b\sqrt{2\pi}}{\pi(\kappa+1)\sqrt{r_0}}\left\{\begin{array}{l}\sin\left(\psi+\dfrac{\theta_0}{2}\right)+\dfrac{3}{2}\sin\left(\psi-\dfrac{\theta_0}{2}\right)-\dfrac{1}{2}\sin\left(\psi-\dfrac{5\theta_0}{2}\right)\\ +i\left\{\cos\left(\psi+\dfrac{\theta_0}{2}\right)+\dfrac{1}{2}\cos\left(\psi-\dfrac{\theta_0}{2}\right)+\dfrac{1}{2}\cos\left(\psi-\dfrac{5\theta_0}{2}\right)\right\}\end{array}\right\}. \tag{5.6}$$

Equation (5.6) is completely consistent with the solution obtained by Lakshamnan and Li [16].

Denoting $K_\infty=K_{\mathrm{I}(\infty)}+iK_{\mathrm{II}(\infty)}=|K_\infty|e^{i\omega_0}$, using Eqs. (5.5) and (5.6), and considering Eq. (3.17), one arrives at

$$G_N=\sqrt{\frac{K_{\mathrm{I}(t)}^4+K_{\mathrm{II}(t)}^4+6K_{\mathrm{I}(t)}^2K_{\mathrm{II}(t)}^2}{K_{\mathrm{I}(\infty)}^4+K_{\mathrm{II}(\infty)}^4+6K_{\mathrm{I}(\infty)}^2K_{\mathrm{II}(\infty)}^2}}=\frac{\sqrt{(\cos\omega_0+\Delta k_1)^4+(\sin\omega_0+\Delta k_2)^4+6(\cos\omega_0+\Delta k_1)^2(\sin\omega_0+\Delta k_2)^2}}{\sqrt{\dfrac{1}{2}(3-\cos 4\omega_0)}} \tag{5.7}$$

where

$$\Delta k_1=\frac{\Delta K_{\mathrm{I}}}{|K_\infty|}=-\frac{\mu b\sqrt{2\pi}}{\pi(\kappa+1)\sqrt{r_0}|K_\infty|}\left\{\sin\left(\psi+\frac{\theta_0}{2}\right)+\frac{3}{2}\sin\left(\psi-\frac{\theta_0}{2}\right)-\frac{1}{2}\sin\left(\psi-\frac{5\theta_0}{2}\right)\right\}$$

$$\Delta k_2=\frac{\Delta K_{\mathrm{II}}}{|K_\infty|}=-\frac{\mu b\sqrt{2\pi}}{\pi(\kappa+1)\sqrt{r_0}|K_\infty|}\left\{\cos\left(\psi+\frac{\theta_0}{2}\right)+\frac{1}{2}\cos\left(\psi-\frac{\theta_0}{2}\right)+\frac{1}{2}\cos\left(\psi-\frac{5\theta_0}{2}\right)\right\}. \tag{5.8}$$

To compare with the new criterion, the traditional criteria in the literature are also presented in normalized forms, respectively, as follows:

(i) $$K_{IN}=\frac{K_{\mathrm{I}(t)}}{|K_\infty|}=\frac{K_{\mathrm{I}(\infty)}+\Delta K_{\mathrm{I}}}{|K_\infty|}=\cos\omega_0+\Delta k_1 \quad \text{(for mode I)}. \tag{5.9}$$

If $K_{IN}<\cos\omega_0$, the dislocation has a shielding effect on the crack tip, and vice versa.

(ii) $$K_{\mathrm{II}N}=\frac{K_{\mathrm{II}(t)}}{|K_\infty|}=\frac{K_{\mathrm{II}(\infty)}+\Delta K_{II}}{|K_\infty|}=\sin\omega_0+\Delta k_2 \quad \text{(for mode II)} \tag{5.10}$$

If $K_{\mathrm{II}N}<\sin\omega_0$, the dislocation has a shielding effect on the crack tip, and vice versa.

The above criteria are usually used to assess the shield effect on the crack tip due to the singularities under pure mode I and mode II remote loads, respectively. Under this condition, the normalized SIF can be simplified as $K_{IN}=1+\Delta k_1$ and $K_{\mathrm{II}N}=1+\Delta k_2$, respectively.

5.2 Numerical Example. Our aim here is to show the significant advantage of G-criterion relative to the SIF-criteria. For numerical convenience, we take: $\mu\sqrt{b}/|K_\infty|=10$, $\nu=1/3$, $b/r_0=1/100$, and $\psi=\theta_0$ in the following calculations. Three cases are evaluated, corresponding to three different kinds of remote loading (see Table 1).

The variation of the normalized G (i.e., G_N), the normalized $K_{\mathrm{I}(t)}$ (i.e., K_{IN}), and the normalized $K_{\mathrm{II}(t)}$ (i.e., $K_{\mathrm{II}N}$) with dislocation location angle θ_0, is plotted in Figs. 6(*a*)–6(*c*) each of the three loading cases.

Table 1 Three cases of remote loading

Case 1	Case 2	Case 3
$K_{\mathrm{I}(\infty)}\neq 0$	$K_{\mathrm{I}(\infty)}=0$	$K_{\mathrm{I}(\infty)}=K_{\mathrm{II}(\infty)}=\sqrt{2}/2\,\lvert K_\infty\rvert$
$K_{\mathrm{II}(\infty)}=0$	$K_{\mathrm{II}(\infty)}\neq 0$	$\lvert K_\infty\rvert\neq 0$

Figure 6(*a*) shows the normalized G and the normalized $K_{\mathrm{I}(t)}$ varying with the dislocation location angle θ_0 under mode I remote loading $K_{\mathrm{I}(\infty)}$ (case 1). It is seen that the maximum shielding angle and amplification angle predicted by the SIF-criterion are about 70 deg and −70 deg, respectively. Although using the G-criterion leads to the same results, the SIF-criterion predicts that the shielding angle is within [0 deg, 180 deg] and the amplification angle is within [−180 deg, 0 deg], whereas the G-criterion predicts that the shielding angle is approximately within [25 deg, 180 deg] and the amplification angle is within [−180 deg, 25 deg]. Another evident difference is that the maximum value of the normalized G is much higher than the value of the normalized $K_{\mathrm{I}(t)}$. Thus the SIF-criterion may underestimate the influence of dislocations.

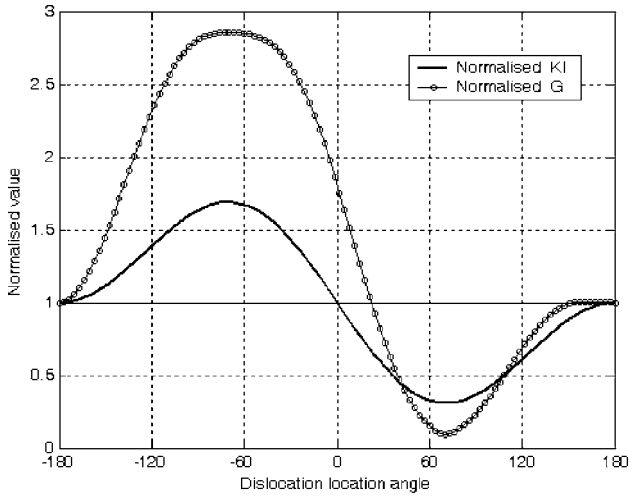
Figure 6(*b*) presents the normalized G and normalized $K_{\mathrm{I}(t)}$ varying with the dislocation location angle θ_0 under mode II remote loading $K_{\mathrm{II}(\infty)}$ (case 2). Apparently, the shielding/amplification effects and their amplitudes estimated by the SIF- and G-criteria are different, although the dislocation location angles predicted by the two criteria for the maximum shield effect are identical.

In summary, for pure mode I or mode II remote loading, it can be concluded that using the two different criteria leads to: (1) nearly identical dislocation location angle for either maximum shielding or maximum amplification effect; (2) different ranges of dislocation location angle for shielding/amplification; (3) the SIF-criterion gives a much lower amplitude of the shielding/amplification effect than that predicted by the G-criterion.

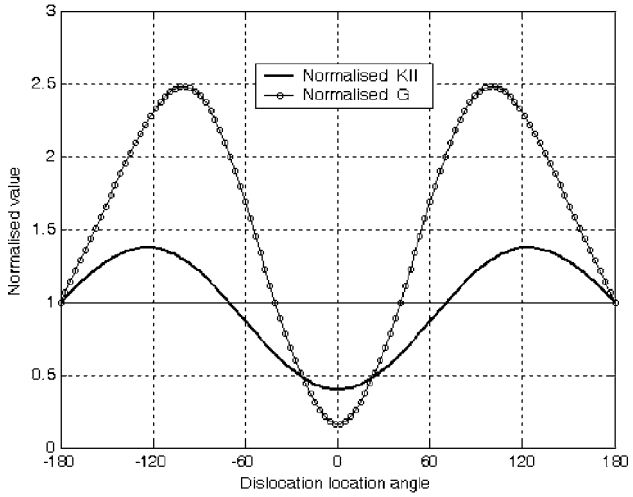
Figure 6(*c*), plotted for case 3, can perhaps fully reveal the advantage of the G-criterion over the SIF-criterion. Under mixed mode remote loading $K_{\mathrm{I}(\infty)}$ and $K_{\mathrm{II}(\infty)}$, Fig. 6(*c*) clearly shows that with the SIF-criterion it is very difficult to precisely make out the shielding/amplification effect dislocation location angle as well as the amplitudes of shielding/amplification. In contrast, the G-criterion employed here can reveal the shielding/amplification effect very well. The main reason, as mentioned in Sec. 1, is that the SIF-criterion is established based on a single stress component out of the six independent stress components, whereas the G-criterion is proposed rationally based on energy principles.

6 Conclusions

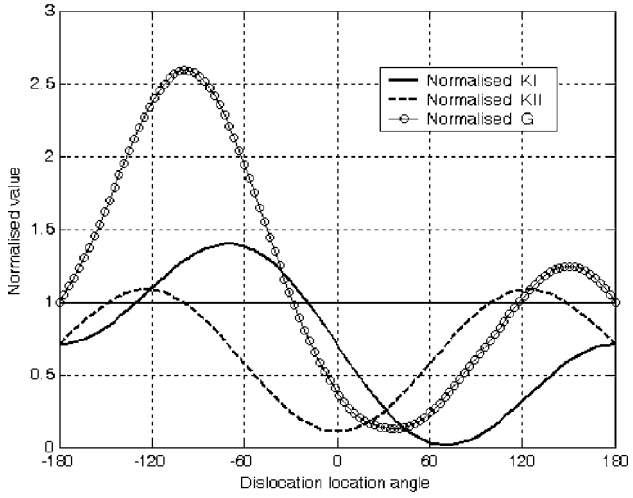
In the present paper we use the analysis of energy to establish the relationship between the potential energy release rate and the conservative vector J_i-integral for multisingularity-crack interaction problems. The individual components of the vector J_i-integral have been re-examined. It is shown that the conservative integrals J_1 and J_2 are equally important in crack problems.



(a)



(b)



(c)

Fig. 6 (***a***) **Normalized** G **and** $K_{I(t)}$ **versus dislocation location angle** θ_0 **under pure** $K_{I(\infty)}$ **remote loading (case 1).** (***b***) **Normalized** G **and** $K_{II(t)}$ **versus dislocation location angle** θ_0 **under pure** $K_{II(\infty)}$ **remote loading (case 2).** (***c***) **Normalized** G, $K_{I(t)}$ **and** $K_{II(t)}$ **versus dislocation location angle** θ_0 **under mixed mode remote loading** $K_{I(\infty)}=K_{II(\infty)}\neq 0$ **(case 3).**

It is found that the remote loading $J_{i(\infty)}$ is redistributed and divided into two main parts due to the presence of multisingularities. One part is acting on the tip of the main crack, and the other is acting on the multisingularities. On the basis of this analysis, a novel G-criterion is proposed to assess the shielding/amplification effect of the singularities on the crack tip.

A general multisingularity-crack interaction problem is solved. The closed-form solution provides a convenient way to perform calculation to evaluate the shield effect due to the presence of multisingularities. Based on it, numerical analysis for discrete-dislocation-crack interaction problem is performed. The results clearly demonstrate that, in contrast to the SIF-criterion, the G-criterion not only can predict the shielding effect on the crack-tip under pure $K_{I(\infty)}$ or $K_{II(\infty)}$ remote loading, but also can deal with the case of a combination of remote loads $K_{I(\infty)}$ and $K_{II(\infty)}$.

Acknowledgment

L.M. would like to thank the National Natural Science Foundation of China (Grant No. 10502040) for partial financial support of this work. LTJ wishes to thank the National Basic Research Program of China through Grant No. 2006CB601202, and the National Natural Science Foundation of China through Grant Nos. 10328203 and 10572111 for partial financial support.

References

[1] Knowles, J. K., and Stenberg, E., 1972, "On a Class of Conservation Laws in Linearized and Finite Elastostatics," Arch. Ration. Mech. Anal., **44**, pp. 187–211.
[2] Budiansky, B., and Rice, J. R., 1973, "Conservation Laws and Energy-Release Rates," ASME J. Appl. Mech., **40**, pp. 201–203.
[3] Eshelby, J. D., 1951, "The Force on an Elastic Singularity," Proc. R. Soc. London, Ser. A, **244**, pp. 87–112.
[4] Rice, J. R., 1968, "A Path Independent Integral and the Approximate Analysis of Strain Concentration by Notches and Cracks," ASME J. Appl. Mech., **35**, pp. 379–386.
[5] Rice, J. R., 1968, "Mathematical Analysis in the Mechanics of Fracture," in *Mathematical Fundamentals*, Fracture An Advanced Treatise Vol. II, H. Liebowitz, ed., Pergamon, New York, pp. 191–311.
[6] Cherepanov, G. P., 1979, *Mechanics of Brittle Fracture*, Nauka, Moscow, English translation published by McGraw-Hill International, New York.
[7] Erdogan, F., and Sih, G. C., 1963, "On the Crack Extension in Plates Under Plane Loading and Transverse Shear," ASME J. Basic Eng., **91**, pp. 764–769.
[8] Otsuka, A., Mori, K., and Miyata, T., 1975, "The Condition of Fatigue Crack Growth in Mixed Mode Condition," Eng. Fract. Mech., **7**, pp. 429–432.
[9] Sih, G. C., 1973, "A Special Theory of Crack Propagation: Methods of Analysis and Solutions of Crack Problems," in *Mechanics of Fracture I*, G. C. Sih, ed., Noordhoff, Leyden, pp. 21–45.
[10] Wu, C.-H., 1978, "Fracture Under Combined Loads by Maximum-Energy-Release-Rate Criterion," ASME J. Appl. Mech., **45**, pp. 553–558.
[11] Shen, M. S., and Shen, M.-H. H., 1995, "Direction of Crack Extension Under General Plane Loading," Int. J. Fract., **70**, pp. 51–58.
[12] Ma, L. F., and Korsunsky, A. M., 2005, "On the Use of Vector *J*-Integral in Crack Growth Criteria for Brittle Solids," Int. J. Fracture, in press.
[13] Rice, J. R., and Thomson, R., 1974, "Ductile Versus Brittle Behavior of Crystals," Philos. Mag., **29**, pp. 73–97.
[14] Hirth, J. P., and Wagoner, R. H., 1976, "Elastic Fields of Line Defects in a Cracked Body," Int. J. Solids Struct., **12**, pp. 117–123.
[15] Lin, I.-H., and Thomson, R., 1986, "Cleavage, Dislocation Emission, and Shielding for Cracks Under General Loading," Acta Metall., **34**(2), pp. 187–206.
[16] Lakshmanan, V., and Li, J. C. M., 1988, "Edge Dislocation Emitted Along Inclined Planes Form a Mode I Crack," Mater. Sci. Eng., A, **104**, pp. 95–104.
[17] Zhang, T.-Y., Tong, P., Ouyang, H., and Lee, S., 1995, "Interaction of an Edge Dislocation With a Wedge Crack," J. Appl. Phys., **78**, pp. 4873–4880.
[18] Cleveringa, H. H. M., Giessen, E. V. D., and Needleman, A., 2000, "A Dislocation Analysis of Mode I Crack Growth," J. Mech. Phys. Solids, **48**, pp. 1133–1157.
[19] Cai, H., and Faber, K. T., 1992, "On the Use of Approximation Methods for Micro-Crack Shielding Problems," ASME J. Appl. Mech., **59**, pp. 497–501.
[20] Andersson, P., and Ståhle, P., 1997, "Shielding Effects and Residual Stresses at Cleavage due to Pre-Existing Dislocations," Int. J. Fract., **85**, pp. 365–380.
[21] Chen, Y. H., 1996, "On the Contribution of Discontinuities in a Near-Tip Stress Field to the *J* Integral," Int. J. Eng. Sci., **34**, pp. 819–829.
[22] Irwin, G. R., 1957, "Analysis of Stresses and Strains Near the End of a Crack Traversing a Plate," ASME J. Appl. Mech., **24**, pp. 361–364.
[23] Suo, Z., 1989, "Singularities Interacting with Interfaces and Cracks," Int. J. Solids Struct., **25**, pp. 1133–1142.

Journal of Applied Mechanics

Discussion

Discussion: "A Paradox in Sliding Contact Problems With Friction" (Adams, G. G., Barber, J. R., Ciavarella, M., and Rice, J. R., 2005, ASME J. Appl. Mech., 72, pp. 450–452)

D. A. Hills[1]
e-mail: david.hills@eng.ox.ac.uk

A. Sackfield

C. M. Churchman

Department of Engineering Science,
University of Oxford,
Parks Road,
Oxford OX1 3PJ, UK

In this interesting paper, the authors address an anomaly which arises when a rigid, square-ended block is pressed against a linear elastic half plane and slid along. The authors note that, within the framework of linear elasticity, the singularity in the contact pressure, and hence shearing traction, produces, adjacent to the edges, regimes in which the implied local relative slip direction dominates the rigid-body sliding velocity, and hence produces a violation of the Coulomb friction law. They seek to resolve the paradox by appealing to a more sophisticated strain definition. All of this is within the context of a quasistatic formulation. The authors recognize, of course, that in any real problem the paradox is unlikely to arise because of (a) the finite strength of the contact giving rise to a yield zone, and (b) the absence of an atomically sharp corner at the contact edge where there is, in all probability, a finite edge radius. Here, we wish to address these issues quantitatively, and so demonstrate that it is unlikely that the paradox described, though interesting, will have any bearing in a real contact. [DOI: 10.1115/1.2201886]

1 Basic Formulation

The punch is moving at velocity U_0 in the positive x direction, relative to the half plane (with the coordinate system (x,y) moving with it), and the surface normal displacement, $v(x)$, and tangential displacement, $u(x)$, are given by

$$\frac{1}{A}\frac{du}{dx} = -\frac{1}{\pi}\int_{\text{contact}} \frac{q(\xi)d\xi}{x-\xi} - \beta p(x) \tag{1}$$

$$\frac{1}{A}\frac{dv}{dx} = \frac{1}{\pi}\int_{\text{contact}} \frac{p(\xi)d\xi}{x-\xi} - \beta q(x) \tag{2}$$

where

$$A = \frac{1-\nu}{\mu}, \quad \beta = \frac{1-2\nu}{2(1-\nu)} \tag{3}$$

ν being the Poisson's ratio and μ the modulus of rigidity of the half plane. The slip velocity of particles on the punch surface relative to particles on the half-plane surface, $U(x)$, is given in the original paper by

$$U \equiv \frac{du}{dt} = \frac{du}{dx}\frac{dx}{dt} = -U_0\frac{du}{dx} \tag{4}$$

As the contact is sliding the tractions are related everywhere by $q(x)=fp(x)$, where f is the coefficient of friction, so that Eq. (1) becomes

$$\frac{du}{dx} = -\frac{fA}{\pi}\int_{\text{contact}} \frac{p(\xi)d\xi}{x-\xi} - A\beta p(x) \tag{5}$$

and from Eq. (2) we have

$$\frac{1}{\pi}\int_{\text{contact}} \frac{p(\xi)d\xi}{x-\xi} = \frac{1}{A}\frac{dv}{dx} + \beta f p(x) \tag{6}$$

Therefore, the surface normal displacements and surface tangential displacements are given by

$$\frac{du}{dx} = -f\frac{dv}{dx} - A\beta(1+f^2)p(x) \tag{7}$$

and hence the relative slip velocity is

$$U = U_0\left[f\frac{dv}{dx} + A\beta(1+f^2)p(x)\right] \tag{8}$$

The slip direction is reversed when $U/U_o > 1$.

2 Asymptotic Representation

Because the region of apparent reverse slip is so small the problem can conveniently be readdressed using an asymptote which gives the problem additional simplicity and applicability. Suppose that a rigid quarter plane is pressed onto the elastic half plane and slid in the positive x direction, with friction. The local contact pressure, $p(s)$, and shearing traction, $q(s)$, may be written in the form [1]

$$\frac{q(s)}{f} = p(s) = K_N s^{\lambda-1} \tag{9}$$

where K_N is a generalized stress intensity factor and s is a coordinate measured from the contact edge. The exponent, λ, is given by the characteristic equation

[1]Author to whom correspondence should be addressed.

$$\tan \pi\lambda = \frac{1}{f\beta}, \quad 0 < \lambda < 1 \tag{10}$$

We now turn our attention to the finite, flat-ended rigid punch, of half-width a, to which a normal load, P, is applied, together with a force sufficient to cause sliding in the positive x direction. This develops a contact pressure, $p(x)$, [2] given by

$$p(x) = \frac{P}{a\pi}\sin \pi\lambda \left(1 + \frac{x}{a}\right)^{\lambda-1}\left(1 - \frac{x}{a}\right)^{-\lambda} \tag{11}$$

such that $p(x)$ is positive in compression. If we apply the change of coordinate $s/a = 1 + x/a$ we get

$$p(s) = \frac{P}{a\pi}\sin \pi\lambda (s/a)^{\lambda-1}[2 - (s/a)]^{-\lambda} \tag{12}$$

$$\simeq \frac{p_o}{\pi}\sin \pi\lambda \left(\frac{s}{2a}\right)^{\lambda-1}[1 + \lambda(s/a)/2 + \cdots] \tag{13}$$

where we have introduced an average pressure $p_o = P/2a$, and hence

$$K_N = \frac{(2a)^{1-\lambda}p_o}{\pi}\sin \pi\lambda \tag{14}$$

We now apply the relationship for the slip velocity $U(x)$ to this solution where, of course, $dv/dx = 0$, and the region where the implied slip direction is reversed is when

$$A\beta(1 + f^2)K_N s^{\lambda-1} > 1 \tag{15}$$

i.e., over a region $s < e$ where

$$e = [A\beta K_N(1 + f^2)]^{1/(1-\lambda)} \tag{16}$$

Therefore, for this specific geometry, when we employ the calibration for K_N, the reversal in slip direction occurs over a region $s < e$ where

$$\frac{e}{a} = 2\left[\frac{1 - 2\nu}{2\pi}\frac{p_o}{\mu}(1 + f^2)\sin \pi\lambda\right]^{1/(1-\lambda)} \tag{17}$$

Thus, the phenomenon presented in the original paper occurs whenever the half plane has a finite compressiblity ($\nu < 1/2$), even if no shearing tractions arise ($f = 0$), but clearly the region of violation increases in size with (a) friction, (b) reduced Poisson effect, (c) dimensionless contact pressure (p_o/μ). The most extreme values one might expect to encounter in practice might be a Poisson's ratio of 0.2, a coefficient of friction of 0.8, and a mean contact pressure of $p_o/\mu = 0.001$, giving $\lambda = 0.41$, so that

$$\frac{e}{a} = 7.1 \times 10^{-7} \tag{18}$$

which is itself tiny, and readily swamped by local plasticity or the effects of rounding.

3 Local Plasticity

We turn now to the question of envelopment of the region of reverse slip by plasticity. As the complete stress field associated with the asymptote is known, through the Muskhelishvili potential, it is straightforward to establish an estimate of the size of the edge plastic zone, simply by seeing where the yield condition is exceeded, in the spirit of the usual fracture mechanics crack-tip plasticity correction.

The Muskhelishvili potential for a rigid punch on a half plane is given by

$$\Phi(z) = \frac{(1 - if)}{2}K_N z^{\lambda-1} \tag{19}$$

where $z = s + iy$ is a complex coordinate in the half plane (and $i = \sqrt{-1}$). The second invariant of the stress tensor as described by von Mises' equivalent stress is (along the interface, $y = 0$)

$$\sigma_e = |K_N| s^{\lambda-1}\sqrt{3f^2 + (1 - \nu)^2} \tag{20}$$

and yield is expected to occur when $\sigma_e = \sigma_Y = \sqrt{3}\tau_Y$. This condition is satisfied over a region $s < r_p$, where r_p is

$$r_p = \left(\frac{|K_N|}{\sigma_Y}\right)^{1/(1-\lambda)}[3f^2 + (1 - \nu)^2]^{1/2(1-\lambda)} \tag{21}$$

and again using the calibration for K_N from the finite geometry

$$\frac{r_p}{a} = 2\left(\frac{p_o}{\sigma_Y}\frac{\sin \pi\lambda}{\pi}\right)^{1/(1-\lambda)}[3f^2 + (1 - \nu)^2]^{1/2(1-\lambda)} \tag{22}$$

which for $\nu = 0.2, f = 0.8$, and $\sigma_Y/\mu = 3 \times 10^{-3}$ (hence $p_o/\sigma_Y = \frac{1}{3}$) gives

$$\frac{r_p}{a} = 0.0934 \tag{23}$$

Thus, for this contact pressure, there will be plasticity over a region roughly 10^5 times larger than the zone over which the paradox occurs.

4 Effect of Rounding

To probe the effect of edge rounding we consider sliding contact between another rigid semi-infinite punch pressed onto the half plane and sliding. This time, contact is assumed to extend from $-d < x < \infty$, and the surface displacement gradient is defined by

$$\frac{dv}{dx} = -\frac{1}{R}x, \quad -d \le x \le 0 \tag{24}$$

$$= 0, \quad x > 0 \tag{25}$$

i.e., the indenter has the form of a parabolic arc of equivalent radius R to the left of the origin, and is flat to the right of the origin. This profile is substituted into integral Eqs. (1) and (2) and solved, giving a pressure along the interface of

$$p(x) = \frac{d\sin^2 \lambda\pi}{AR\pi}(s/d)^{\lambda}\left[\frac{1}{1 - \lambda} + \frac{1 - (s/d)}{\lambda}{}_2F_1(1, \lambda; 1 + \lambda; (s/d))\right] \quad (s/d) < 1 \tag{26}$$

$$= \frac{d\sin^2 \lambda\pi}{AR\pi(1 - \lambda)(2 - \lambda)}(s/d)^{\lambda-1}{}_2F_1\left(1, 1 - \lambda; 3 - \lambda; \frac{1}{(s/d)}\right) \quad (s/d) > 1 \tag{27}$$

where $s = x + d$ and ${}_2F_1(.)$ is a standard hypergeometric function. We note that, when $(s/d) \gg 1$, $p(x) \sim s^{\lambda-1}$, so that the asymptotic form given by Eq. (27) applies, and calibration shows that the generalized stress intensity factor is given by

$$K_N = \frac{\sin^2 \lambda\pi d^{2-\lambda}}{AR\pi(1 - \lambda)(2 - \lambda)} \tag{28}$$

This scaling factor serves both to provide a connection between the applied load (K_N) and the extent of contact in the radiused portion, d, and it enables us to write the contact pressure as

$$\frac{d^{1-\lambda}p(s)}{K_N} = (2 - \lambda)(s/d)^{\lambda}\left[1 + \frac{(1 - \lambda)}{\lambda}(1 - (s/d)){}_2F_1(1, \lambda; 1 + \lambda; (s/d))\right] \quad (s/d) < 1 \tag{29}$$

$$=(s/d)^{\lambda-1}{}_2F_1\left(1,1-\lambda;3-\lambda;\frac{1}{(s/d)}\right) \quad (s/d)>1 \quad (30)$$

whereas the relative slip displacement is given by

$$\frac{U}{U_0}=\frac{d}{R}\Psi\left(\frac{s}{d},f,\beta\right) \quad (31)$$

$$\Psi\left(\frac{s}{d},f,\beta\right)=f\left[1-\left(\frac{s}{d}\right)\right]+\frac{\beta(1+f^2)}{\pi(1+f^2\beta^2)}\left(\frac{s}{d}\right)^{\lambda}\left[\frac{1}{(1-\lambda)}+\frac{1}{\lambda}\left[1-\left(\frac{s}{d}\right)\right]{}_2F_1\left(1,\lambda;1+\lambda;\left(\frac{s}{d}\right)\right)\right] \quad (32)$$

and, employing the contact law for this problem,

$$d=\left(\frac{K_N AR\pi(1-\lambda)(2-\lambda)}{\sin^2\lambda\pi}\right)^{1/(2-\lambda)} \quad (33)$$

gives

$$\frac{U}{U_0}=(K_N AR^{\lambda-1})^{1/2-\lambda}\left(\frac{\pi(1-\lambda)(2-\lambda)}{\sin^2\lambda\pi}\right)^{1/(2-\lambda)}\Psi\left(\frac{s}{d},f,\beta\right) \quad (34)$$

$$=\left(\frac{2a}{R}\right)^{(1-\lambda)/(2-\lambda)}\left(\frac{p_o}{\mu}(1-\nu)\right)^{1/(2-\lambda)}\left(\frac{(1-\lambda)(2-\lambda)}{\sin\lambda\pi}\right)^{1/(2-\lambda)}\times\Psi\left(\frac{s}{d},f,\beta\right) \quad (35)$$

Therefore there is a violation in slip direction when $U/U_0>1$, i.e., when

$$\left(\frac{R}{2a}\right)^{\lambda-1}\left(\frac{p_o}{\mu}(1-\nu)\right)>\left(\frac{\sin\lambda\pi}{(1-\lambda)(2-\lambda)}\right)\left[\Psi\left(\frac{s}{d},f,\beta\right)\right]^{\lambda-2} \quad (36)$$

The violation is more likely to occur for cases when $R/2a$ is very small and the load p_o/μ is relatively high. However, unlike the perfectly sharp solution, the function $\Psi(s/d,f,\beta)$ does not exhibit a monotonically increasing value as $s\to 0$ but shows a maximum. This cannot be found analytically, but has been found numerically for the extreme example case used earlier ($f=0.8$, $\nu=0.2$) so that the maximum of $\Psi(s/d,f,\beta)$ occurs at $s/d=0.133$ and is of magnitude 1.002. Therefore, using the same load as before, $p_o/\mu=0.001$, we find that the paradox will occur if

$$\frac{R}{2a}<5.9\times 10^{-6} \quad (37)$$

If we have a contact of total width 20 mm then in order to *avoid* the paradox occurring we require $R>120$ nm. This is such a tiny radius that for all contacts of practical importance the paradox will not exist.

5 Concluding Remarks

The region of reverse slip violating the friction law has been quantified and shown to be extremely small. Physical boundaries for its envelopment by plasticity or absence through local rounding have also been explicitly found. It is very probable that plasticity will produce the greater effect.

References

[1] Comninou, M., 1976, "Stress Singularity at a Sharp Edge in Contact Problems With Friction," ZAMP, **27**, pp. 493–499.

[2] Hills, D. A., Nowell, D., and Sackfield, A., 1993, *Mechanics of Elastic Contacts*, Butterworth-Heinmann, Oxford, UK.

Closure to "Discussion of 'A Paradox in Sliding Contact Problems With Friction' " (2006, ASME J. Appl. Mech., 73, pp. 884–886)

G. G. Adams
Department of Mechanical Engineering,
Northeastern University,
Boston, MA 02115
e-mail: adams@coe.neu.edu

J. R. Barber
Department of Mechanical Engineering,
University of Michigan,
Ann Arbor, MI 48109
e-mail: jbarber@umich.edu

M. Ciavarella
Centre of Excellence in Computational Mechanics,
Politecnico di Bari,
70125 Bari, Italy
e-mail: mciava@poliba.it

J. R. Rice
Division of Engineering and Applied Sciences,
Harvard University,
Cambridge, MA 02138
e-mail: rice@esag.deas.harvard.edu

The purpose of our technical brief was to first demonstrate that a paradox involving slip reversal exists in a class of sliding contact problems with friction. We were then motivated by the desire to determine a mechanics formulation of these problems which is self-consistent, in which case we do not have to resort to making arguments that depend on the actual dimensions and material properties of a particular situation. We also stated "The finite strain kinematics analysis shows that the paradox disappears when the correct kinematics is used. In this case it is a reasonable engineering solution to use the infinitesimal theory with the assumption that slip is always in the original direction of sliding, because the paradox occurs only in very small regions in which the infinitesimal theory is unrealistic."

We thank the authors of the Discussion for detailing some of the issues with plasticity and rounding as related to *traditional* engineering materials. The statement is made in the Discussion that a radius of greater than 120 nm is necessary to avoid the paradox. While such a tiny radius may be unrealistic in many applications, it is not at all unusual in, for example, silicon MEMS devices where radii can be nearly atomically sharp as a consequence of an anisotropic etch.

The plasticity analysis used in the Discussion assumed a ratio of yield stress to shear modulus of 0.003. However, in rubber and biological materials the "yield" stress can be a significantly higher fraction of the modulus. Even for traditional engineering materials, the elastodynamic problem behaves as the corresponding quasi-static problem with a reduced modulus. As pointed out in the original technical brief, this reduced modulus approaches zero as the speed of the punch approaches the Rayleigh wave speed.

For the reasons outlined above we disagree with the statement that "…it is unlikely that the paradox described, though interesting, will have any bearing in a real contact." We also reiterate that the purpose of our technical brief was to develop a mechanics formulation which is self-consistent.

