An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments

W. C. Oliver
Metals and Ceramics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6116

G. M. Pharr
Department of Materials Science, Rice University, P. O. Box 1982, Houston, Texas 77251

The indentation load-displacement behavior of six materials tested with a Berkovich indenter has been carefully documented to establish an improved method for determining hardness and elastic modulus from indentation load-displacement data. The materials included fused silica, soda-lime glass, and single crystals of aluminum, tungsten, quartz, and sapphire. It is shown that the load-displacement curves during unloading in these materials are not linear, even in the initial stages, thereby suggesting that the flat punch approximation used so often in the analysis of unloading data is not entirely adequate. An analysis technique is presented that accounts for the curvature in the unloading data and provides a physically justifiable procedure for determining the depth which should be used in conjunction with the indenter shape function to establish the contact area at peak load. The hardnesses and elastic moduli of the six materials are computed using the analysis procedure and compared with values determined by independent means to assess the accuracy of the method. The results show that with good technique, moduli can be measured to within 5%.

I. INTRODUCTION

Great strides have been made over the past few years in the development of techniques for probing the mechanical properties of materials on the submicron scale.[1-3] The advances have been made possible by the development of instruments that continuously measure force and displacement as an indentation is made.[4-7] The indentation load-displacement data thus derived can be used to determine mechanical properties even when the indentations are too small to be imaged conveniently. Since the indentation positioning capability of some of the instruments is also in the submicron regime, a means is available by which the mechanical properties of a surface can be mapped with submicron resolution. An instrument with such capabilities can be described as a mechanical properties microprobe.[8-10]

The two mechanical properties measured most frequently using load and depth sensing indentation techniques are the elastic modulus, E, and the hardness, H. In a commonly used method, data are obtained from one complete cycle of loading and unloading.[3] The unloading data are then analyzed according to a model for the deformation of an elastic half space by an elastic punch which relates the contact area at peak load to the elastic modulus. Methods for independently estimating the contact area from the indenter shape function are then used to provide separate measurements of E and H.

The elastic contact problem, which plays a key role in the analysis procedure, was originally considered in the late 19th century by Boussinesq[11] and Hertz.[12] Boussinesq developed a method based on potential theory for computing the stresses and displacements in an elastic body loaded by a rigid, axisymmetric indenter. His method has subsequently been used to derive solutions for a number of important geometries such as cylindrical and conical indenters.[13,14] Hertz analyzed the problem of the elastic contact between two spherical surfaces with different radii and elastic constants. His now-classic solutions form the basis of much experimental and theoretical work in the field of contact mechanics[15] and provide a framework by which the effects of non-rigid indenters can be included in the analysis. Another major contribution was made by Sneddon, who derived general relationships among the load, displacement, and contact area for any punch that can be described as a solid of revolution of a smooth function.[16,17] His results, which will be used extensively in this paper, show that the load-displacement relationships for many simple punch geometries can conveniently be written as

$$P = \alpha h^m \tag{1}$$

where P is the indenter load, h is the elastic displacement of the indenter, and α and m are constants. Values of the exponent m for some common punch geometries are $m = 1$ for flat cylinders, $m = 2$ for cones, $m = 1.5$ for

spheres in the limit of small displacements, and $m = 1.5$ for paraboloids of revolution.

Modeling indentation contact in a way that includes plasticity is a much more complex problem. Since the constitutive equations are nonlinear and a number of material parameters must be included to describe material behavior (e.g., yield strength and work hardening coefficient), analytical solutions are not easily obtained.[15] As a result, much of our understanding of the importance of plasticity in indenter contact problems has been derived through experimentation and finite element simulation.

The earliest experiments in which load and displacement sensing indentation methods were used to measure mechanical properties were performed by Tabor,[18] who studied the indentation of a number of metals deformed by hardened spherical indenters. A similar study was subsequently undertaken by Stillwell and Tabor to examine the behavior of conical indenters.[19] One particularly important observation resulting from these studies concerns the shape of the hardness impression after the indenter is unloaded and the material elastically recovers. The experiments revealed that, at least in metals, the impression formed by a spherical indenter is still spherical with a slightly larger radius than the indenter, and the impression formed by a conical indenter is still conical with a larger included tip angle. The importance of these experiments is that since elastic contact solutions exist for each of these geometries (i.e., a spherical indenter in a spherical hole and a conical indenter in a conical hole), the ways in which plasticity affects the interpretation of elastic unloading data can be dealt with by taking into account the shape of the perturbed surface in the elastic analysis. Tabor used these results to show that the shape of the entire unloading curve and the total amount of recovered displacement can be accurately related to the elastic modulus and the size of the contact impression for both spherical and conical indenters. Other important observations resulting from these studies include (1) the diameter of the contact impression in the surface formed by conical indenters does not recover during unloading—only the depth recovers; (2) the indentation must be loaded and unloaded a few times before the load-displacement behavior becomes perfectly reversible; i.e., a limited amount of plasticity sometimes occurs in each of the first few loadings and unloadings: and (3) effects of non-rigid indenters on the load-displacement behavior can be effectively accounted for by defining a reduced modulus, E_r, through the equation

$$\frac{1}{E_r} = \frac{(1 - \nu^2)}{E} + \frac{(1 - \nu_i^2)}{E_i} \qquad (2)$$

where E and ν are Young's modulus and Poisson's ratio for the specimen and E_i and ν_i are the same parameters for the indenter.

Interest in load and displacement sensing indentation testing as an experimental tool for measuring elastic modulus began in the early 1970's with the work of Bulychev, Alekhin, Shorshorov, and co-workers.[20–24] These investigators used instrumented microhardness testing machines to obtain indentation load-displacement data like that shown schematically in Fig. 1 which was then analyzed according to the equation

$$S = \frac{dP}{dh} = \frac{2}{\sqrt{\pi}} E_r \sqrt{A}. \qquad (3)$$

Here, $S = dP/dh$ is the experimentally measured stiffness of the upper portion of the unloading data, E_r is the reduced modulus (previously defined), and A is the projected area of the elastic contact. By measuring the initial unloading stiffness and assuming that the contact area is equal to the optically measured area of the hardness impression, the modulus can thus be derived.

Equation (3) has its origins in elastic contact theory. While originally derived for a conical indenter, Buylchev et al. showed that Eq. (3) holds equally well for spherical and cylindrical indenters and speculated that Eq. (3) may apply to other geometries as well.[21] Pharr, Oliver, and Brotzen have subsequently shown that Eq. (3) applies to any indenter that can be described as a body of revolution of a smooth function.[25] Bulychev et al. also argued that significant deviations from the behavior predicted by Eq. (3) should not occur for
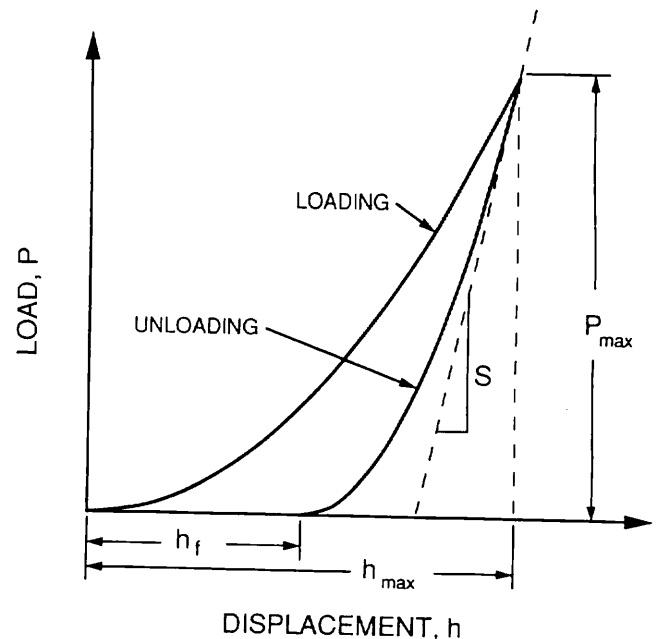
pyramidal indenters. They noted good agreement between generally accepted values of moduli and values computed from Eq. (3) using indentation data obtained with a Vickers indenter. That Eq. (3) works well for at least some indenters that cannot be described as bodies of revolution has been confirmed by the finite element calculations of King.[26] These calculations show that the deviations from Eq. (3) for flat ended punches with square and triangular cross sections are only 1.2% and 3.4%, respectively.

In the early 1980's, it was realized that load and depth sensing indentation methods could be very useful in the measurement of the mechanical properties of very thin films and surface layers, and instruments for producing submicron indentations were developed.[4–7] For practical reasons, some means other than direct observation of the hardness impressions was needed to measure contact areas since imaging very small indentations is both time-consuming and difficult. Oliver, Hutchings, and Pethica suggested a simple method based on measured indentation load-displacement curves and a knowledge of the indenter area function (or shape function), that is, the cross-sectional area of the indenter as a function of the distance from its tip.[1,2] The method is based on the notion that, at peak load, the material conforms to the shape of the indenter to some depth; if this depth can be established from the load-displacement data, the projected area of contact can be estimated directly from the shape function. Two obvious choices for the depth, though not the only ones, are the depth at peak load, h_{max} (i.e., the maximum displacement in the loading cycle), and the final depth, h_f (i.e., the residual depth of the hardness impression after final unloading), both of which are easily determined from indentation load-displacement data (see Fig. 1). Using TEM replication methods to establish the shape function, Oliver et al. found that the final depth gives a better estimate of the contact area than the depth at peak load.[2]

Doerner and Nix subsequently put together many of these ideas to produce the most comprehensive method to date for determining hardness and modulus from indentation load-displacement data.[3] Their approach is based on the observation that during the initial stages of unloading, the elastic behavior of the indentation contact is similar to that of a flat cylindrical punch; that is, the area of contact remains constant as the indenter is unloaded. This was justified by experimental observations which suggest that, for some materials, the initial portions of unloading curves are linear, as would be expected for the flat-punch geometry. The unloading stiffness dP/dh is then related to the modulus and contact area through a relationship equivalent to Eq. (3). To evaluate independently the contact area, Doerner and Nix proposed a simple empirical method

based on extrapolating the initial linear portion of the unloading curve to zero load and using the extrapolated depth with the indenter shape function to determine the contact area. Careful experiments in one material, METGLAS® 2826, confirmed that the extrapolated depth gives a better estimate of the contact area than either the depth at peak load or the final depth. This observation was later confirmed with finite element simulations of the indentation of silicon and nickel by conical indenters.[27] With the contact area so determined, the modulus can be computed from Eq. (3) and the hardness from its normal definition:

$$H = \frac{P_{max}}{A}, \tag{4}$$

where P_{max} is the peak indentation load and A is the projected area of the hardness impression. As a practical matter, Doerner and Nix suggested that the unloading stiffness can be computed from a linear fit of the upper one-third of the unloading curve.

One disturbing feature of the Doerner–Nix method concerns their assumption of linear unloading. During the past several years, we have conducted load and displacement sensing indentation tests on a large number of materials, and our observations have led us to believe that unloading curves are rarely, if ever, linear, even in the initial stages of unloading. Rather, unloading data are better described by power laws like Eq. (1) with exponents ranging from about 1.2 to 1.6. In addition, by employing a special dynamic technique by which stiffness can be measured continuously during indentation, we have found that unloading contact stiffnesses change immediately and continuously as the indenter is withdrawn, as would be expected from continuous changes in contact area. Together, these observations suggest that the flat punch approximation is not an entirely adequate description of real material behavior.

In this paper, a new method for analyzing indentation load-displacement data is outlined which addresses these problems. We begin by presenting data for a variety of materials to show that unloading data are usually nonlinear. An analysis technique is then presented which accounts for the curvature in the unloading data and provides a physically justifiable procedure for determining the depth that should be used in conjunction with the indenter shape function to establish the contact area at peak load. The hardnesses and moduli of several materials are then computed using the analysis and compared with values determined by independent means to establish the accuracy of the method. In addition, several practical issues concerning procedures for taking data and methods for determining load frame compliance and indenter shape functions are discussed.

II. EXPERIMENTAL

A. Materials

The materials used in the study were chosen to span a wide range of hardness and modulus. They included a soft fcc metal (aluminum), a hard bcc metal (tungsten), two amorphous glasses (soda-lime glass and fused silica), and two crystalline ceramics (sapphire and quartz). Details concerning purity, finish, and crystal orientation are given in Table I. One seemingly obvious choice which does not appear in the table is silicon, which was intentionally excluded because its indentation behavior is complicated by cracking and pressure-induced phase transformations.[28-30]

Aluminum and tungsten were chosen in preference to other metals because both are relatively isotropic in their elastic properties; their anisotropy ratios are 1.21 and 1.00, respectively.[31] The question of whether the modulus measured in an indentation test represents that of some specific crystallographic direction or some average value is therefore not an issue. The same is true of the amorphous glasses, which are isotropic in both their elastic properties and their hardnesses. Quartz and sapphire, on the other hand, have complex hexagonal crystal structures, and because of this, results from these materials can be used to provide some insight into the importance of elastic anisotropy.

B. Indentation procedure

All experiments were performed using a Nano-indenter® (Nanoindenter is a registered trademark of Nano Instruments, Inc., Knoxville, TN) at the Oak Ridge National Laboratory, a schematic illustration of which is shown in Fig. 2. The system has load and displacement resolutions of 0.3 μN and 0.16 nm, respectively. A Berkovich indenter, a three-sided pyramid with an area-
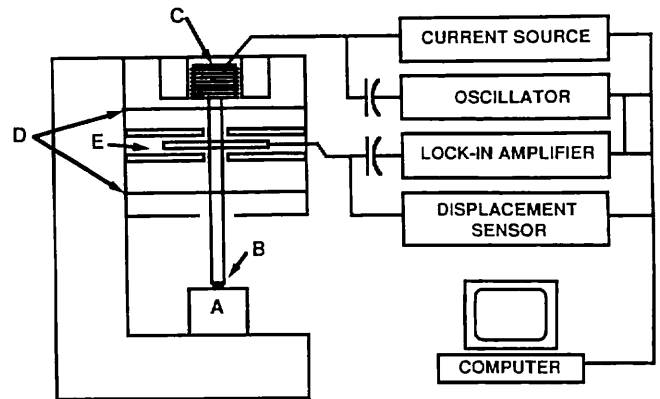

FIG. 2. A schematic representation of the experimental apparatus used to perform the indentation experiments: (A) sample; (B) indenter; (C) load application coil; (D) indentation column guide springs; (E) capacitive displacement sensor.

to-depth function which is the same as that of a Vickers indenter, was used in all experiments.

Most experiments were performed using a load-time sequence like that shown in Fig. 3. The indenter was first loaded and unloaded three times in succession at a constant rate of loading with each of the unloadings terminated at 10% of the peak load to assure that contact was maintained between the specimen and the indenter. The reason for performing multiple loadings and unloadings was to examine the reversibility of the deformation and thereby make sure that the unloading data used for analysis purposes were mostly elastic. After the third unloading, the load was held constant for a period of 100 s at 10% of the peak value while the displacement was carefully monitored to establish the rate of displacement produced by thermal expansion in the system. Even though the system is thermally buffered from its surroundings and the room in which it is housed is temperature-controlled to within ±1 °C, small thermal fluctuations cause some of the machine

TABLE I. Summary of materials used in indentation study.

| Material	Description
Aluminum	99.995% pure single crystal mechanically polished with colloidal silica
Tungsten	99.95% pure single crystal mechanically polished with colloidal silica
Soda-lime glass	Commercial microscope slide
Fused silica	Optically flat substrate material
Quartz	(001) single crystal; optically flat
Sapphire	99.995% Al_2O_3; (001) single crystal mechanically polished to optical flatness
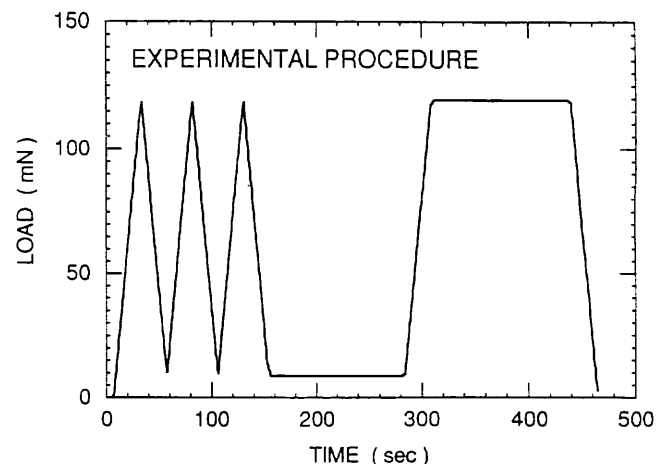
FIG. 3. A typical load-time sequence; peak load = 120 mN.

components to expand, and the expansion is manifested in the data as an apparent displacement in the specimen. Thermal drift becomes particularly important for small indentations made over long periods of time, as was the case for some of the indentations in this study. To account for thermal drift, the rate of displacement was measured during the last 80 s of the hold period, and the displacement data were corrected by assuming that this drift rate was constant throughout the test. In this procedure, it is implicitly assumed that the displacements measured at the low load during the hold period are due mostly to thermal expansion rather than to time dependent plasticity in the specimen. Following the hold period, the specimen was loaded for a fourth and final time, with another 100 s hold period inserted at peak load to allow any final time dependent plastic effects to diminish, and the specimen was fully unloaded.

For each material, six separate peak loads were investigated starting at 120 mN and successively reducing the load by a factor of 3 to produce indentations at 120, 40, 13.3, 4.4, 1.5, and 0.5 mN. The loading/unloading rate was also reduced by a factor of 3 starting at a value of 5000 μN/s for the 120 mN indentations. The reason for reducing the loading rate in proportion to the peak load was to assure that all load-displacement curves had about the same number of data points since (the system samples data at a constant rate of approximately 1.6 data points per second). Five indentations were made at each load, with most of the results presented here representing averages for the group. Wherever possible, scatter bars representing ± one standard deviation have been included in the data.

The indentation system also had the ability to continuously measure contact stiffness during indentation. This is accomplished by superimposing a small oscillation on the force signal and measuring the displacement response at the same frequency. Continuous stiffness measurement was used to provide direct evidence that the contact stiffness changes continuously during loading and unloading. Details of the technique and analysis procedure, parts of which have been published elsewhere,[32-34] are included in Appendix I.

III. EXPERIMENTAL RESULTS

A. General characteristics of load-displacement curves

We begin the discussion of experimental results with an overview of the characteristics of the load-displacement curves for the six materials studied. Figures 4–9 present experimental data for each of the materials for indentations made to peak loads of 120 mN. The differences in hardness of the materials are apparent from the large differences in the depth attained a maximum load. The softest material is aluminum, with a peak depth
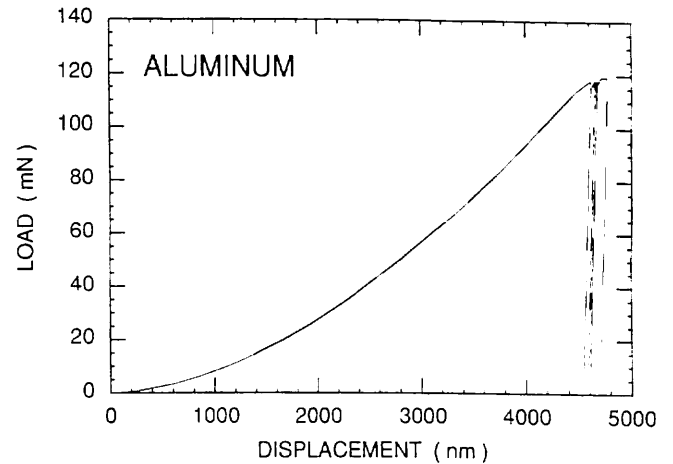

FIG. 4. Load versus indenter displacement for the highest peak load experiment performed on a mechanically polished aluminum single crystal.

of almost 5000 nm, while the hardest is sapphire, which was penetrated to a depth of only about 500 nm. Both the aluminum and tungsten data are typical of materials in which the hardness is relatively small compared to the modulus, as is observed in most metals; most of the indenter displacement in these two materials is accommodated plastically, and only a small portion is recovered on unloading. The other materials show varying degrees of elastic recovery during unloading, the largest being that for fused silica.

The unloading/reloading behavior of the materials exhibits some variability. Figure 10 shows an expanded view of the sapphire results, demonstrating that the unloading and reloading curves for this material are nearly the same. This behavior is also observed in fused silica and quartz, with the near perfect reversibility suggesting that deformation in these materials after the initial loading is almost entirely elastic.
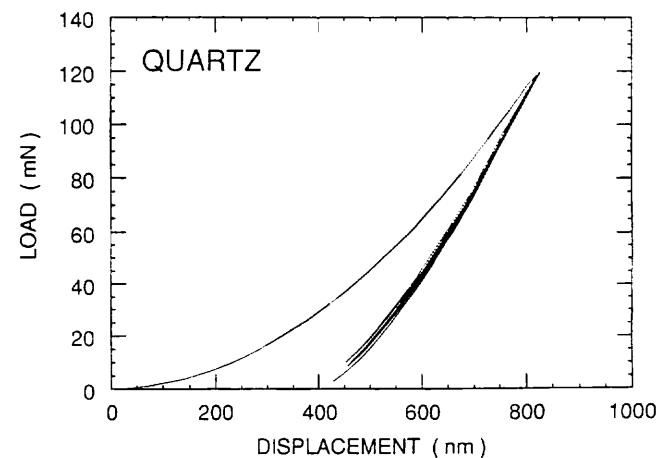

FIG. 5. Load versus indenter displacement for the highest peak load experiment performed on a (001) quartz single crystal.
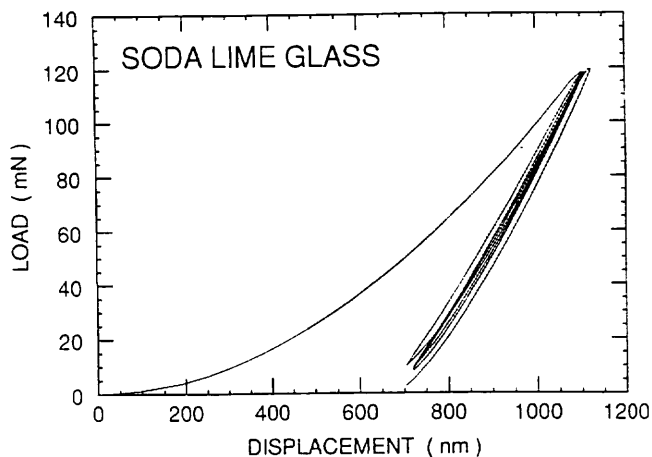
FIG. 6. Load versus indenter displacement for the highest peak load experiment performed on soda-lime glass.
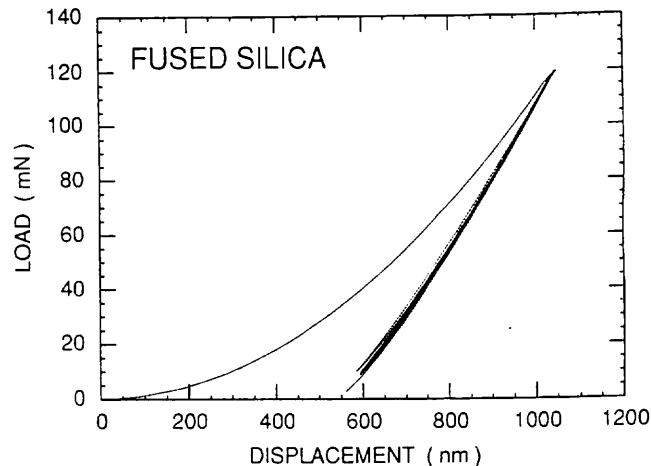

FIG. 8. Load versus indenter displacement for the highest peak load experiment performed on fused silica.

For aluminum. on the other hand, the expanded view in Fig. 11 shows that the peak load displacements shift to higher values in successive cycles. In addition, the relatively large displacement just prior to final unloading is due to creep during the 100 s hold period at peak load. Thus, it is apparent that at least some plastic deformation occurs in this material after the initial loading, and some of it is time dependent. The behavior of tungsten is similar to that of aluminum.

A third type of behavior is that exhibited by soda-lime glass. As shown in Fig. 12, this material exhibits distinct hysteresis loops, as might be expected if there were a small amount of reverse plasticity upon unloading. The looping degenerates with cycling, however, and after 3 or 4 cycles, the load-displacement behavior is largely elastic.

The main reason for discussing the load-displacement behavior during unloading/reloading is to point out

that in some materials, displacements recovered during first unloading may not be entirely elastic, and because of this, the use of first unloading curves in the analysis of elastic properties can sometimes lead to inaccuracies. One way to minimize nonelastic effects is to include peak load hold periods in the loading sequence to allow time dependent plastic effects to diminish. and to use unloading curves obtained after several cycles of loading to minimize effects of reverse plasticity. The procedure we have adopted here to avoid these problems is to load and unload the indentation four times and include one long period at peak load. Our analyses are then performed using the final unloading data only. Exactly how nonelastic effects can influence the measurement of mechanical properties is documented later in this paper.
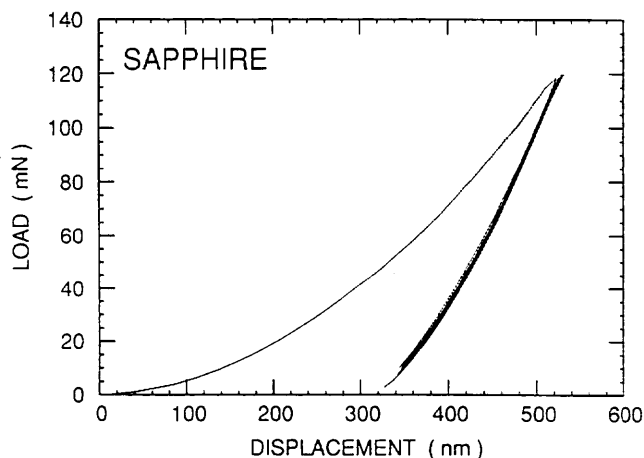

FIG. 7. Load versus indenter displacement for the highest peak load experiment performed on a (001) sapphire single crystal.
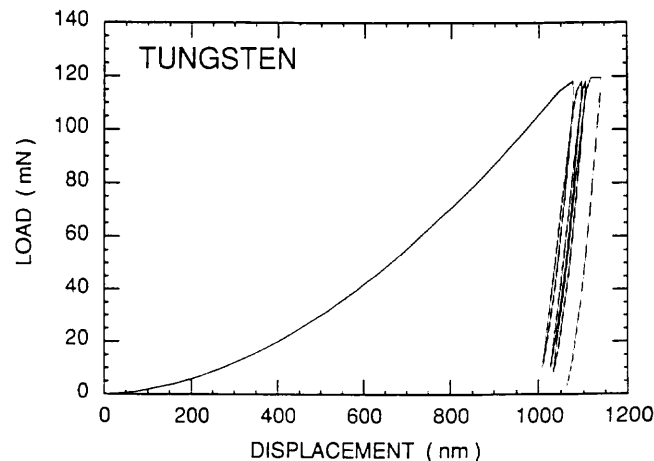

FIG. 9. Load versus indenter displacement for the highest peak load experiment performed on a mechanically polished tungsten single crystal.
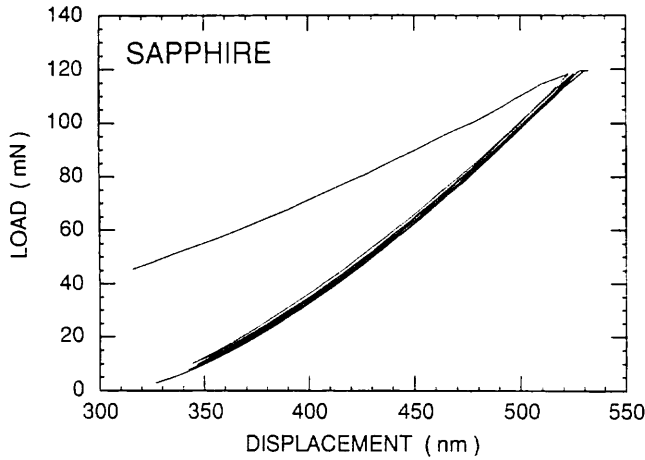
FIG. 10. An expanded view of the unloading/reloading portion of the load versus indenter displacement data for the highest peak load experiment performed on sapphire.
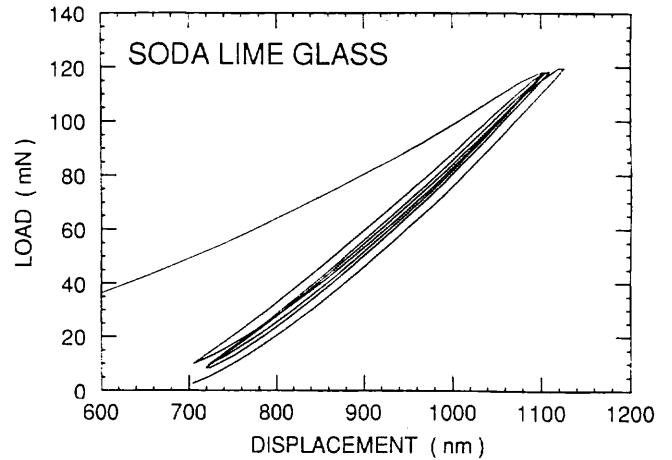

FIG. 12. An expanded view of the unloading/reloading portion of the load versus indenter displacement data for the highest peak load experiment performed on soda-lime glass.

B. Shapes of unloading curves

To examine carefully the shapes of unloading curves, final unloading data from all six materials are plotted on a common set of axes in Fig. 13. For the sake of comparison, the displacements have been shifted laterally by subtracting the final depth, h_f, from the total depth, h, to force all the curves to pass through a common origin. Careful examination of the figure shows that none of the data is linear; rather, each unloading curve is slightly concave up over its entire span. This can be better appreciated by plotting the data logarithmically, as has been done in Fig. 14. The fact that the data are linear on logarithmic axes implies that the unloading curves are well described by a power law relation like that of Eq. (1). The exact values of the experimentally

observed power law exponents, as well as the correlation coefficients for the power-law fits, are given in Table II. The power law exponents for the six materials studied here vary from 1.25 to 1.51, and the fact that they are distinctly greater than one implies that none of the data is consistent with flat punch behavior.

C. Continuous stiffness measurement

Additional evidence that the flat punch approximation is inappropriate is provided by the continuous measurement of stiffness. Here, we discuss results for a material not included in Table I, electropolished tungsten. This material is particularly interesting because well-annealed single crystals indented by a Berkovich indenter at sufficiently low loads can be deformed purely elastically—that is, the contact impression fully recovers
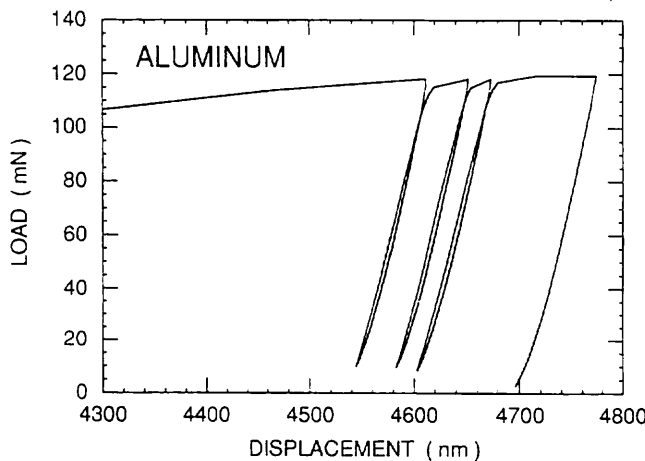

FIG. 11. An expanded view of the unloading/reloading portion of the load versus indenter displacement data for the highest peak load experiment performed on a mechanically polished aluminum single crystal.
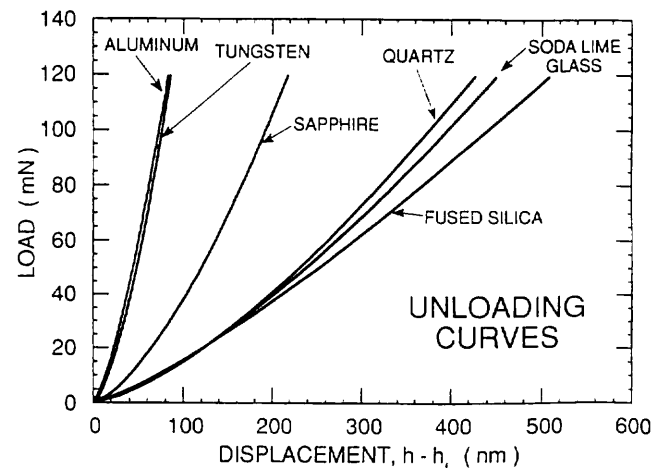

FIG. 13. The final unloading segment of the load versus indenter displacement data from the highest peak load experiments for all six materials plotted so that the curves pass through a common origin.
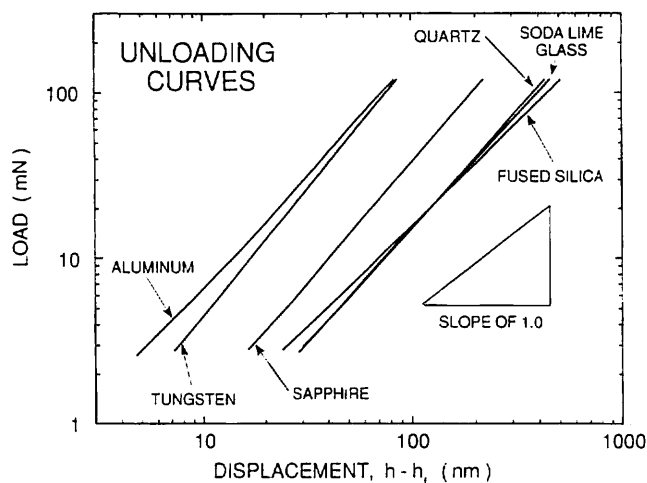
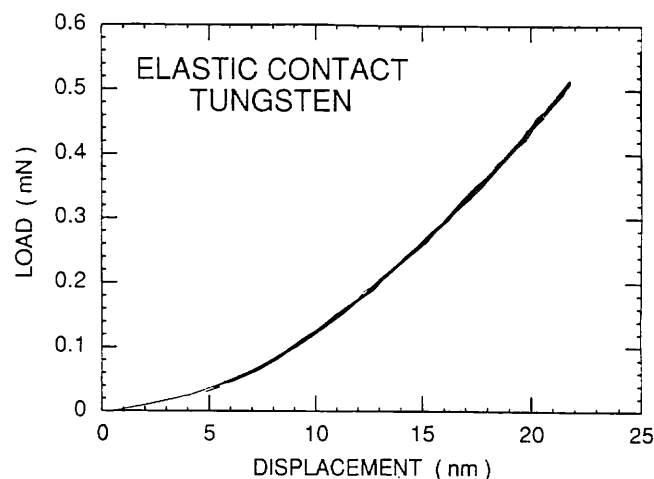FIG. 14. The data from Fig. 13 replotted on logarithmic axes.

FIG. 15. Load versus indenter displacement for a fully elastic contact on an electropolished single crystal of tungsten. The data consist of four sequential loading/unloading cycles.

upon unloading, no hardness impression remains after indentation, and the loading and unloading curves trace one another perfectly.[32,33] As shown in Fig. 15, this behavior was observed in electropolished tungsten for peak loads of 0.5 mN and less. For purely elastic contact, both the contact area and the contact stiffness are unique, reversible functions of the load, and they increase and decrease with load according to simple functional relationships that depend on the geometry of the indenter. Because of this, contact stiffness correlates directly with contact area, and hence, changes in contact stiffness in an indentation experiment can be correlated with changes in contact area. Continuous stiffness measurement results for the indentation data in Fig. 15 are presented in Fig. 16, plotted as stiffness versus time. Comparison of these data with the load-time history shown in Fig. 3 shows that the measured contact stiffness, and thus the contact area, does indeed increase and decrease in the way that would be expected based on the loading history (note that the hold periods in Fig. 3 were not included in acquiring the data of Figs. 15 and 16).

At higher peak loads, the indentation contact in electropolished tungsten is not just elastic. Rather, as shown in Fig. 17, when a threshold load of about 1.0 mN

is reached, a sudden jump in displacement corresponding to the onset of plasticity is observed, and a permanent hardness impression is formed. The continuous stiffness measurement for this indentation is shown in Fig. 18. It is seen that for each of the four unloadings, the contact stiffness changes *immediately* and *continuously* as the specimen is unloaded. Thus, the contact area, which varies in the same way as the contact stiffness, is not constant during the unloading of the plastic hardness impression, even during the initial stages of unloading, and the flat punch approximation is therefore questionable. Continuous and immediate changes in contact stiffness during unloading were also observed in each of the other materials examined in this study. Tungsten was chosen to illustrate the behavior because of its unique elastic behavior at low loads.

TABLE II. Parameters describing power law fits of unloading curves.

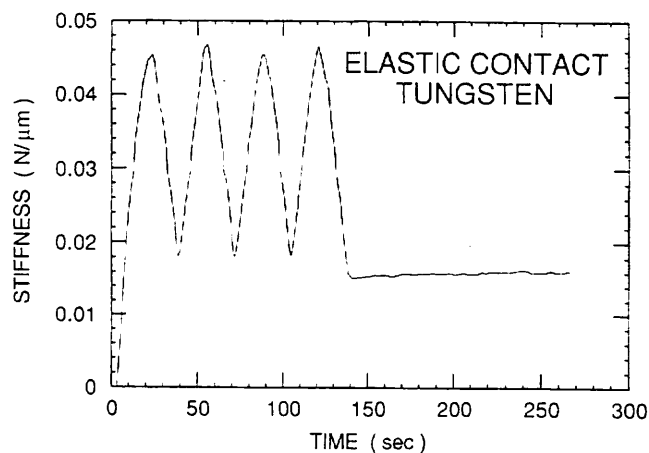| Material	A (mN/nm^m)	m	Correlation coefficient
Aluminum	0.265	1.38	0.999938
Quartz	0.0215	1.43	0.999985
Soda-lime glass	0.0279	1.37	0.999997
Sapphire	0.0435	1.47	0.999998
Fused silica	0.0500	1.25	0.999997
Tungsten	0.141	1.51	0.999986

FIG. 16. Contact stiffness versus time for a fully elastic contact on an electropolished single crystal of tungsten measured with the continuous stiffness measurement technique.
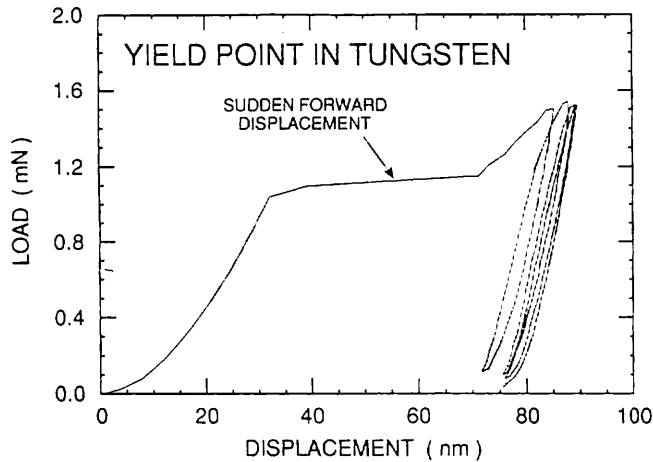
FIG. 17. Load versus indenter displacement for a plastic contact on an electropolished single crystal of tungsten. The data consist of four sequential loading/unloading cycles.

D. Comments on linear unloading

Before leaving this section, a comment is warranted as to why the observations here are different from other reports in the literature which suggest that unloading curves in some materials are linear, at least in the initial stages of unloading. The primary observations of this sort are those of Doerner and Nix, who report linear unloading in metals over most of the unloading range and in silicon for at least the first one-third of the unloading curve.[3] The data they present as representative of metallic behavior are that of aluminum. The data consist of one loading and unloading plotted with the unloading and loading curves on a single set of axes. On such a plot, the unloading curves are so steep that they give the appearance of being linear even though they may not be. In Fig. 4, for example, the unloading curves
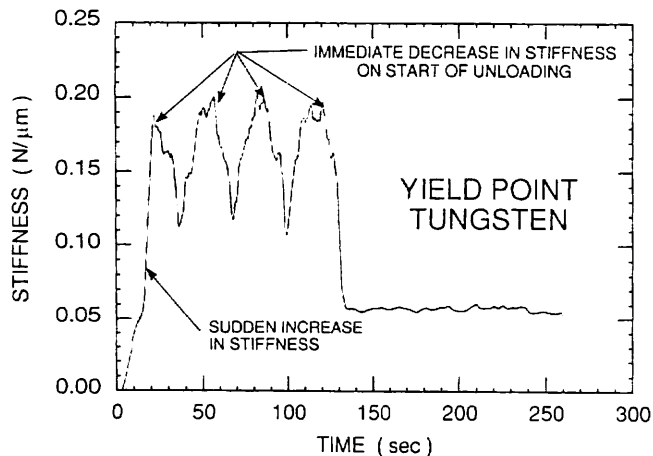

FIG. 18. Contact stiffness versus time for a plastic contact on an electropolished single crystal of tungsten measured using the continuous stiffness measurem⁻⁻ ⁻⁻⁻⁻ue.

for aluminum appear to be linear, but when plotted on expanded or logarithmic axes such as those in Figs. 11, 13, and 14, the nonlinearity becomes apparent. We also wish to note that improper values for the load frame stiffness can influence unloading curves in a way that makes them appear to be more linear than they really are.

With respect to the observations of linear unloading in silicon, we note that the indentation behavior of this material stands in a class of its own. Studies have shown that a reversible, pressure induced transformation to a denser phase occurs beneath the indenter.[28-30] When the indenter is withdrawn, the transformation reverts, and a portion of the unloading displacement is recovered by a nonelastic process. Unloading curves for silicon are thus unlike those for any other material, and they sometimes may appear to have a linear region near the top. In our experience, however, careful examination of such data has shown that it is curved with power law exponents in the range reported here for the other materials.

IV. A NEW METHOD OF ANALYSIS

The observations presented in the previous section suggest that the flat punch approximation is not an entirely adequate description of the behavior of materials when indented by a Berkovich indenter. Here, we propose a new method of analysis based on analytical solutions for other indenter geometries. In addition to accounting for curvature in the unloading data, the method also provides a physically justifiable procedure for determining the depth that should be used in conjunction with the indenter shape function to establish the contact area at peak load.

As mentioned in the introduction, Sneddon has derived closed form analytical solutions for punches of several geometries. In addition to the flat punch, he has also considered punches with conical geometries and paraboloids of revolution.[16] We proceed here on the assumption that the behavior of one of these latter geometries gives a better description of the elastic unloading of an indentation made with a Berkovich indenter.

The conical indenter is a natural choice since, like the Berkovich indenter, its cross-sectional area varies as the square of the depth of contact and its geometry is singular at the tip. The paraboloid of revolution, whose behavior is the same as that of a spherical indenter in the limit of small displacements, is also a potentially useful geometry in that no real indenter is ever perfectly sharp; i.e., at some scale the tip exhibits some rounding. In addition, because of plasticity, an elastic singularity cannot really exist at the tip of the indenter, and the paraboloid geometry potentially accounts for this. For both geometries, the load-displacement relationships are nonlinear and the contact area changes continuously

1572

J. Mater. Res., Vol. 7, No. 6, Jun 1992

during unloading. An analysis of the geometry that works best is presented in the next section by comparing predictions of the method with actual experimental results.

An important fundamental assumption in the development is that the equations describing the elastic unloading of a flat, semi-infinite half space are the same as those for an indented surface; that is, Sneddon's solutions apply equally well to a flat surface or a surface with a hardness impression. The justification for this assumption is based largely on the empirical observations discussed in the introduction that hardness impressions formed in metals by conical and spherical indenters are also conical and spherical with slightly different included tip angles and radii. As discussed in greater detail elsewhere,[25] the mathematical description of the elastic loading and unloading of these perturbed surfaces is exactly the same as that for the flat surface if adjustments are made to the geometrical parameters describing the tip angle of the cone and the effective radius of the sphere. Since spherical and conical indenters represent two very different geometries, i.e., one very sharp and one very blunt, it seems reasonable that this behavior may hold for other axisymmetric indenters as well.

Figure 19 shows a cross section of an indentation and identifies the parameters used in the analysis. At any time during loading, the total displacement h is written as

$$h = h_c + h_s \qquad (5)$$

where h_c is the vertical distance along which contact is made (hereafter called the contact depth) and h_s is the displacement of the surface at the perimeter of the contact. At peak load, the load and displacement are P_{max} and h_{max}, respectively, and the radius of the contact circle is a. Upon unloading, the elastic displacements are recovered, and when the indenter is fully withdrawn, the final depth of the residual hardness impression is h_f.
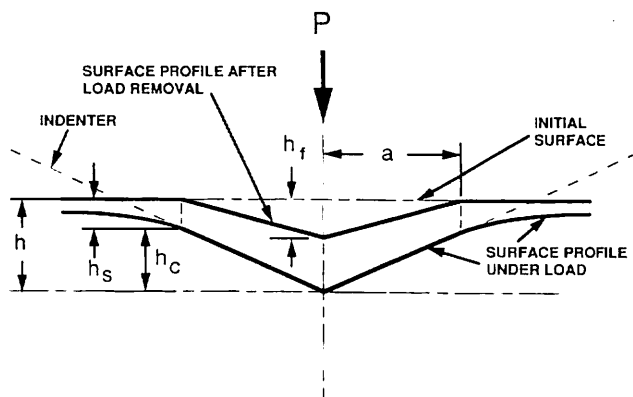
FIG. 19. A schematic representation of a section through an indentation showing various quantities used in the analysis.

The experimental parameters needed to determine hardness and modulus are shown in the schematic load-displacement data shown in Fig. 20. The three key parameters are the peak load (P_{max}), the depth at peak load (h_{max}), and the initial unloading contact stiffness (S_{max}). It should be noted that the contact stiffness is measured only at peak load, and no restrictions are placed on the unloading data being linear during any portion of the unloading.

The analysis begins by rewriting Eq. (3) as

$$E_r = \frac{\sqrt{\pi}}{2} \frac{S}{\sqrt{A}} \qquad (6)$$

which relates the reduced modulus, E_r, to the contact area, A, and the measured stiffness, S. As discussed previously, this relationship holds for any indenter that can be described as a body of revolution of a smooth function and is thus not limited to a specific geometry.[25] Measurement of the initial unloading slope can thus be used to determine the reduced modulus if the contact area at peak load can be measured independently.

The area of contact at peak load is determined by the geometry of the indenter and the depth of contact, h_c. Following Oliver et al.,[1,2] we assume that the indenter geometry can be described by an area function $F(h)$ which relates the cross-sectional area of the indenter to the distance from its tip, h. Given that the indenter does not itself deform significantly, the projected contact area
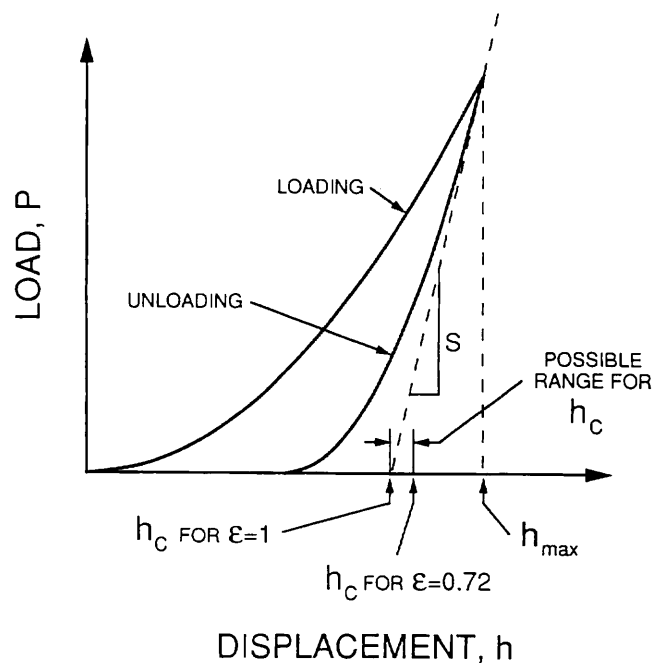
FIG. 20. A schematic representation of load versus indenter displacement showing quantities used in the analysis as well as a graphical interpretation of the contact depth.

at peak load can then be computed from the relation

$$A = F(h_c) \tag{7}$$

The functional form of F must be established experimentally prior to analysis.

To determine the contact depth from the experimental data, it is noted that

$$h_c = h_{max} - h_s \tag{8}$$

which follows directly from Eq. (5). Since h_{max} can be experimentally measured, the key to the analysis then becomes how the displacement of the surface at the contact perimeter, h_s, can be ascertained from the load-displacement data.

The deflection of the surface at the contact perimeter depends on the indenter geometry. For a conical indenter, Sneddon's expression for the shape of the surface outside the area of contact[16] can be used to give

$$h_s = \frac{(\pi - 2)}{\pi}(h - h_f) \tag{9}$$

The quantity $(h - h_f)$ appears in this expression rather than h by itself since Sneddon's solution applies only to the elastic component of the displacement. In addition, Sneddon's force-displacement relationship for the conical indenter yields

$$(h - h_f) = 2\frac{P}{S} \tag{10}$$

where S is the stiffness. Substituting Eq. (10) into Eq. (9) and noting that the contact area of interest is that at peak load, one obtains

$$h_s = \epsilon\frac{P_{max}}{S} \tag{11}$$

where the geometric constant ϵ for the conical indenter is given by

$$\epsilon = \frac{2}{\pi}(\pi - 2) \tag{12}$$

or $\epsilon = 0.72$. When similar arguments are made for the flat punch and the paraboloid of revolution, Eq. (11) is again obtained with different geometric constants. For the flat punch, $\epsilon = 1$, and for the paraboloid of revolution, $\epsilon = 0.75$.

The graphical interpretation of Eq. (11) is shown in Fig. 20. For $\epsilon = 1$, the value for the flat punch, $h_s = P_{max}/S$, and the contact depth h_c is given by the intercept of the initial unloading slope with the displacement axis. Interestingly, this is precisely the depth used by Doerner and Nix in their analysis based on the flat punch approximation. Thus, the current method is consistent with the Doerner and Nix approach when the flat punch geometry is assumed. For the conical and paraboloid indenters, however, the contact depths are

greater than those for the flat punch, and this must be accounted for in analyses using these indenter geometries if accurate measurements are to be obtained. The range of h_c for the indenters considered here is shown in Fig. 20.

In addition to the modulus, the data obtained using the current method can be used to determine the hardness, H. We define the hardness as the mean pressure the material will support under load. With this definition, the hardness is computed from

$$H = \frac{P_{max}}{A} \tag{13}$$

where A is the projected area of contact at peak load evaluated from Eq. (7). It should be noted that hardness measured using this definition may be different from that obtained from the more conventional definition in which the area is determined by direct measurement of the size of the residual hardness impression. The reason for the difference is that, in some materials, a portion of the contact area under load may not be plastically deformed, and as a result, the contact area measured by observation of the residual hardness impression may be less than that at peak load. However, for most of the materials considered in this study, the contact areas computed using the procedure outlined here compare favorably with residual contact areas measured in the SEM, and the two definitions of hardness give similar results. This will be discussed further in the next section.

V. IMPLEMENTATION OF THE METHOD

A. Choice of indenter geometries

The first step in implementing the new analysis procedure is to identify which of the indenter geometries best describes the experimental data and therefore which value of ϵ should be used. While it is conceivable that the best description varies from one material to another, the data presented in Fig. 21 suggest otherwise. The figure shows the unloading curves of Fig. 13 replotted with the loads and displacements normalized with respect to their maximum values. When plotted this way, it is apparent that the unloading data are all remarkably similar in form. In fact, with the exception of fused silica data, which are slightly higher than the rest, the curves for the materials are so similar that they can hardly be distinguished. This suggests that unloading behavior of a wide variety of materials can be described by a single indenter geometry.

As for which indenter geometry works the best, the primary clue is found in the power law exponents, m, describing the unloading behavior. Table III summarizes the power law exponents predicted by Sneddon's analyses along with the associated values of ϵ. The experimentally determined values of m listed in Table II
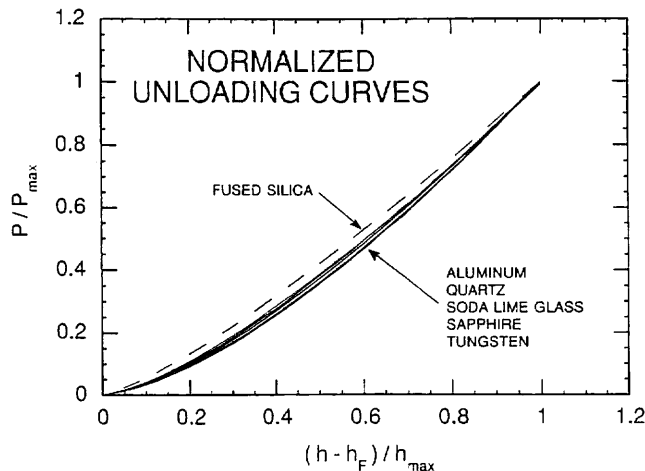
FIG. 21. Normalized final unloading curves for all six materials showing similarities in shape.
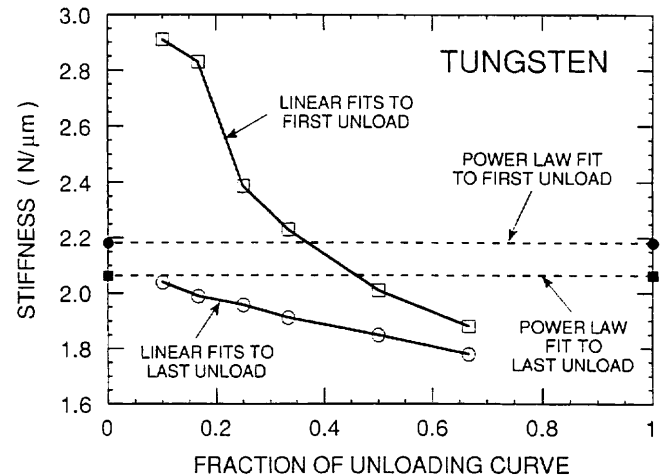

FIG. 22. The peak load stiffness of a 120 mN indentation in tungsten computed using linear fits of the unloading data and the power law fitting technique. The dashed lines are power law fits applied to the first and last unloading segments.

range from 1.25 to 1.51 with a mean value of 1.40. Comparison of the two tables thus shows that based on the power law exponents, the unloading behavior is best described by the paraboloid geometry, and the value of ϵ we use in the analysis procedure is 0.75. Why the paraboloid is a better description of the unloading behavior is probably due to the fact that the elastic singularity characteristic of the conical geometry is not physically realizable when plasticity occurs, and the pressure distribution which actually forms around the tip of the indenter is more like that predicted by the parabola of revolution. It is notable, however, that even had the conical geometry been chosen, the only difference in the analysis would have been to use a slightly smaller value of ϵ, i.e., $\epsilon = 0.72$.

B. Measurement of the initial unloading stiffness

One important practical question that arises in the analysis procedure is how the initial unloading stiffness, S, should be measured from the unloading data. One simple way to accomplish the measurement is to fit a straight line to a fraction of the upper portion of the unloading curve and use its slope as a measure of the stiffness. The problem with this is that for nonlinear unloading data, the measured stiffness depends on how much of the data is used in the fit.

To illustrate how important this can be, plotted in Fig. 22 are stiffnesses measured for tungsten as a

TABLE III. Punch parameters used in data analyses.

Punch geometry	ϵ	m
Flat	1	1
Paraboloid	0.75	1.5
Conical	0.72	2

function of the fraction of the unloading data used in the fit. Two features are worthy of note. First, the stiffnesses obtained from the first and last unloadings are very different. The reason for this is that there is a significant amount of creep during the first unloading, causing the slope of the upper portion of the unloading curve to be abnormally high. Because of this, there is no way that the proper unloading stiffness can be identified from the first unloading data using this technique. These effects can be minimized by the inclusion of peak load hold periods in the loading sequence to diminish time dependent plastic effects, as was done just prior to the final unloading in this study. Second, it is seen that even in the final unloading data, there is a significant variation in the measured stiffness, depending on how much of the unloading curve is used in the fit. For this reason, an alternative procedure is desirable.

The procedure we have adopted is based on the observation that unloading data are well described by a simple power law relation, as directly evidenced by the excellent correlation coefficients in Table II. The actual relationship we use to describe the unloading data for stiffness measurement is

$$P = A(h - h_f)^m \tag{14}$$

where the constants A, m, and h_f are all determined by a least squares fitting procedure. The initial unloading slope is then found by analytically differentiating this expression and evaluating the derivative at the peak load and displacement.

Results for tungsten are included in Fig. 22 as horizontal dashed lines. It is seen that the stiffness computed from the final unloading curve is very close to that of the linear fits obtained using small fractions of the unloading data. Interestingly, when applied to the

first unloading data, the power law method produces a stiffness that is only slightly greater than that derived from the last unloading curve, thereby implying that the power law method is less sensitive to creep. For these reasons, the power law method is preferred and is used in all subsequent analyses.

C. Determination of load frame compliance and diamond area function

Another important practical consideration is how well the load frame compliance and area function of the diamond are known. The load frame compliance is important because the measured displacements are the sum of the displacements in the specimen and the load frame, and, thus, to accurately determine specimen displacements, the load frame compliance must be known with some precision. This is especially important for large indentations made in materials with high modulus for which the load frame displacement can be a significant fraction of the total displacement. The area function is equally important since computation of both the modulus and the hardness depend on the contact areas through Eqs. (6) and (13).

The standard procedure used in the past for determining area functions has been to make a series of indentations at various depths in materials in which the indenter displacement is predominantly plastic and measure the size of the indentations by direct imaging. Using the contact depths computed from the indentation load-displacement data, an area function is then derived by empirically fitting a function to a plot of the imaged areas versus the contact depths. Clearly, the imaging technique must have sufficient contrast, resolution, and precision to accurately determine the areas. Optical imaging works well for larger indentations, but cannot be used exclusively since it is usually necessary to characterize the area function well into the submicron range. Scanning electron microscopy is also of limited value for small indentations because the shallowness of the hardness impressions results in very poor contrast. TEM replication methods have proved useful, but they are tedious, time-consuming, and require equipment and expertise which may not be available in all laboratories.

Here, we propose a method for determining area functions that requires no imaging at all. The method is based on the assumption that the elastic modulus is independent of indentation depth, justification for which will be given shortly when results of contact area measurements for several different materials are presented. It is important to note that specific values for the modulus are not assumed.

The method follows by modeling the load frame and the specimen as two springs in series, in which case

$$C = C_s + C_f \qquad (15)$$

where C is the total measured compliance, C_s is the compliance of the specimen, and C_f is the compliance of the load frame. Since the specimen compliance during elastic contact is given by the inverse of the contact stiffness, S, Eqs. (3) and (15) combine to yield

$$C = C_f + \frac{\sqrt{\pi}}{2E_r} \frac{1}{\sqrt{A}} \qquad (16)$$

It is thus seen that if the modulus is constant, a plot of C vs $A^{-1/2}$ is linear for a given material, and the intercept of the plot is a direct measure of the load frame compliance. The best values of C_f are obtained when the second term on the right-hand side of Eq. (16) is small, i.e., for large indentations.

To find the area function and the load frame compliance, we take advantage of the fact that relatively large indentations can be made in aluminum because of its low hardness. In addition, for the largest aluminum indentations, the area function for a perfect Berkovich indenter

$$A(h_c) = 24.5h_c^2 \qquad (17)$$

can be used to provide a first estimate of the contact area. Initial estimates of C_f and E_r were thus obtained by plotting C vs $A^{-1/2}$ for the two largest indentations in aluminum. Using these values, contact areas were then computed for all six indentation sizes by rewriting Eq. (16) as

$$A = \frac{\pi}{4} \frac{1}{E_r^2} \frac{1}{(C - C_f)^2} \qquad (18)$$

from which an initial guess at the area function was made by fitting the A vs h_c data to the relationship

$$A(h_c) = 24.5h_c^2 + C_1h_c^1 + C_2h_c^{1/2} + C_3h_c^{1/4} \\ + \dots\dots + C_8h_c^{1/128} \qquad (19)$$

where C_1 through C_8 are constants. The lead term describes a perfect Berkovich indenter; the others describe deviations from the Berkovich geometry due to blunting at the tip.

The procedure is not complete at this stage because the exact form of the area function influences the values of C_f and E_r, so using the new area function the procedure was applied again and iterated several times until convergence was achieved. Figure 23 shows a plot of the final values of $(C - C_f)$ vs $A^{-1/2}$. Note that the data are linear and extrapolate to $(C - C_f) = 0$, as they should once the proper load frame compliance and area function have been found.

To extend the area function to smaller depths and to check the validity of the constant modulus assumption, the procedure was subsequently applied to all the other materials. In doing so, the load frame compliance was held constant at the value determined from the aluminum
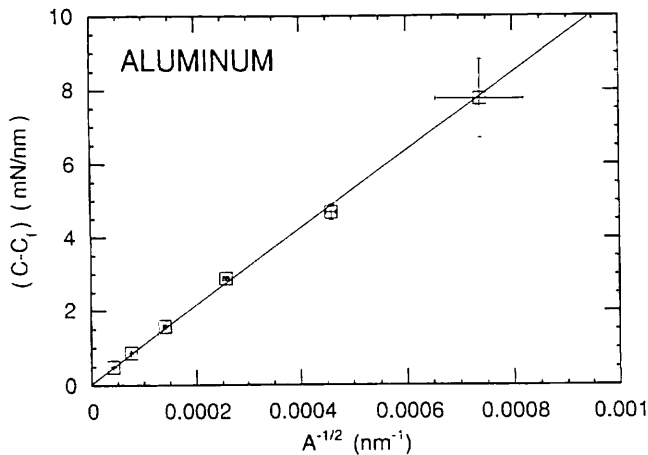
FIG. 23. A plot of $(C - C_f)$ vs $A^{-1/2}$ for aluminum. The error bars are two standard deviations in length.

FIG. 24. The computed contact areas versus contact depths for all six materials. The error bars are two standard deviations in length.

data since the largest contacts were made in this material. The area of the largest indentation in each of the other materials was computed using the aluminum area function, and the modulus for each material was estimated using Eq. (6). The areas needed to produce a constant modulus as a function of depth for each material were then computed using Eq. (18). The resulting areas and the corresponding contact depths for all materials were fit as a group using Eq. (19) to define a new area function. The process of determining E_r at the largest depth for each material (using the latest area function), recalculating areas assuming contant modulus, and redetermining the area function was then repeated iteratively until the area function accurately described the data for all of the materials at all depths. Note that nowhere in the procedure was a specific value for a modulus assumed; the only assumption is that the modulus is independent of depth.

The results for all the materials are plotted in Fig. 24 as contact area versus contact depth. The solid line through the data is the final composite area function derived by fitting Eq. (19) to all the data. The fact that one single area function adequately describes the data suggests that the constant modulus assumption and the use of a single area function are valid. The composite area function derived from these data is used in all further analyses.

Subsequent to this work, a more specific procedure was developed for determining area functions and load frame compliances using aluminum and fused quartz as calibration materials. Details are described in Appendix II.

D. Assessment of predictive capability

To assess the predictive capability of the method, load-displacement data for each of the six experimental materials were analy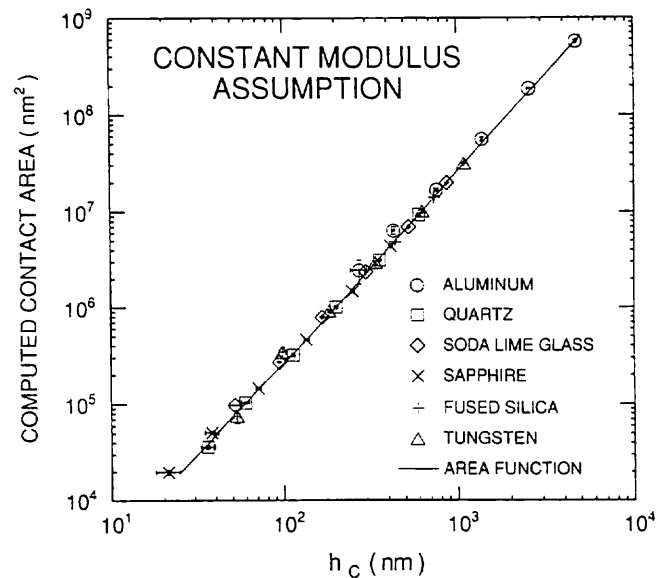zed according to the procedures described in Sec. IV to determine contact areas, moduli, and hardnesses. The elastic constants of the diamond indenter used in Eq. (2) to compute the specimen modulus from the reduced modulus were $E_i = 1141$ GPa and $\nu_i = 0.07$. Here, we compare predictions of the method with values assessed using independent means.

The first comparison is between the contact areas calculated from the load displacement data and contact areas measured by direct imaging in the SEM. Figure 25 presents results for all six materials plotted as calculated versus imaged areas. The good agreement between the two measurements suggests that the method works well, with the only possible exception being aluminum at high loads, for which the calculated areas are slightly greater than imaged areas. That aluminum behaves in this way may be related to the fact that it was the only material studied here for which there was significant pile-up around the indentation.

The good agreement in Fig. 25 also suggests that the peak load contact areas computed using the new method and the imaged residual contact areas are much the same in most materials, and thus the definition of hardness given in Eq. (13) is consistent with the conventional definition. This finding was at first somewhat surprising to us, since there is significant elastic recovery upon unloading in the ceramics and glasses we examined (see Figs. 5–8), and we expected to find the residual contact areas in these materials to be measurably smaller than the peak load values. The reason for the agreement can be at least partially understood by examination of the SEM micrograph in Fig. 26. The unusual shape of this fused silica indentation suggests that during unloading, the sides of the indentation elastically recovered while
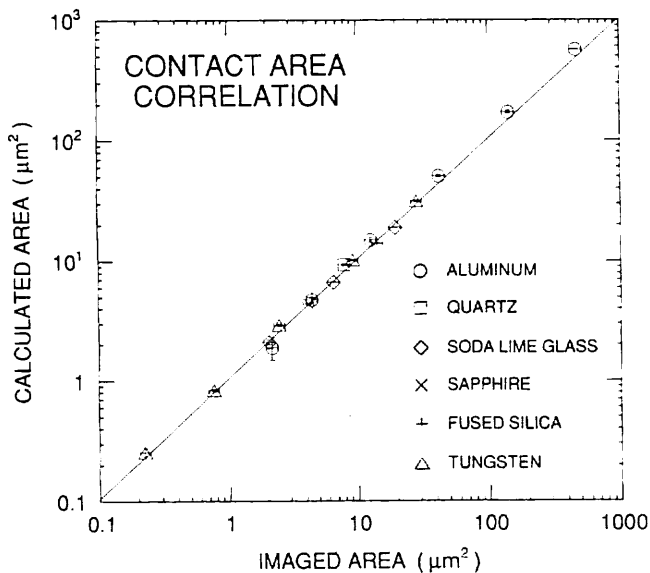
FIG. 25. Contact areas computed using the new analysis procedure versus imaged contact areas for all six materials. The error bars are two standard deviations in length.

the corners did not, thus producing the odd appearance of the edges of the impression. Apparently, the additional
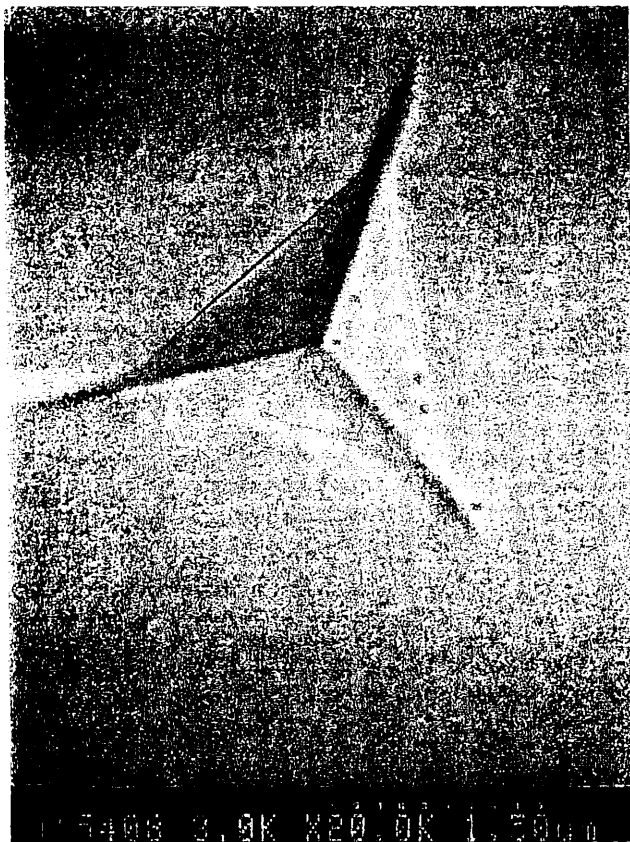

FIG. 26. A scanning electron micrograph of a 40 mN indentation in fused silica.

plasticity caused by the stress concentration at the edges the indenter serves to mark permanently the position of the corners of the indentation at peak load. Thus, if the area of the hardness impression is measured by computing the area of the triangle defined by the impression corners, as is most often done in conventional hardness testing and as was done here, then the size of the imaged indentation is a good measure of the peak load contact area. In this regard, it is notable that the same agreement would not be expected for indentations made with smooth, axially symmetric indenters (i.e., cones, paraboloids or spheres), for which elastic recovery would occur uniformly around the periphery of the impression.

The hardnesses computed from the calculated areas are shown in Fig. 27. The data show that there is very little indentation size effect (ISE) in any of the materials with the exception of aluminum and tungsten in which there is a modest increase in hardness at low loads. This could be due to surface-localized cold-work resulting from polishing.

The moduli predicted from the indentation load-displacement data are shown in Fig. 28. Note that to compute the modulus from the measured values of E_r requires that Poisson's ratio be known. The assumed values for Poisson's ratio are shown in Table IV. The data in Fig. 28 show that once again, there is very little evidence for an indentation size effect; i.e., the moduli remain more or less constant over the entire range of load.

To assess the predictive capabilities of the modulus determination, moduli obtained by averaging the experimental results at the two highest loads are compared with commonly accepted values from the literature in Table IV. For the crystalline materials, the literature values were obtained by averaging the Voigt and Reuss average elastic constants (or the Hashin and Shtrikman
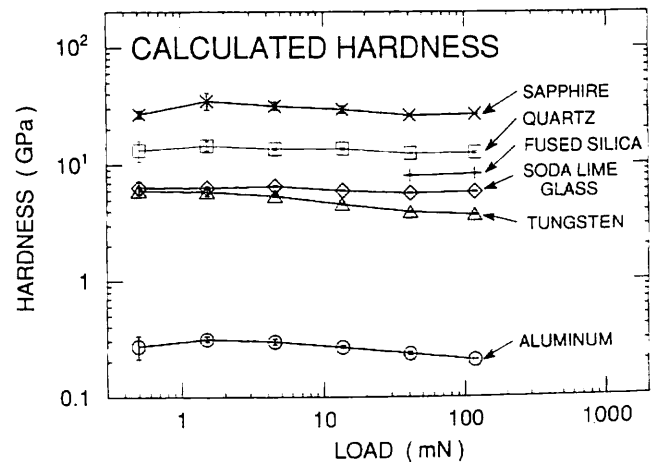

FIG. 27. The load dependence of the hardness calculated using the new analysis procedure for all six materials. The error bars are two standard deviations in length.
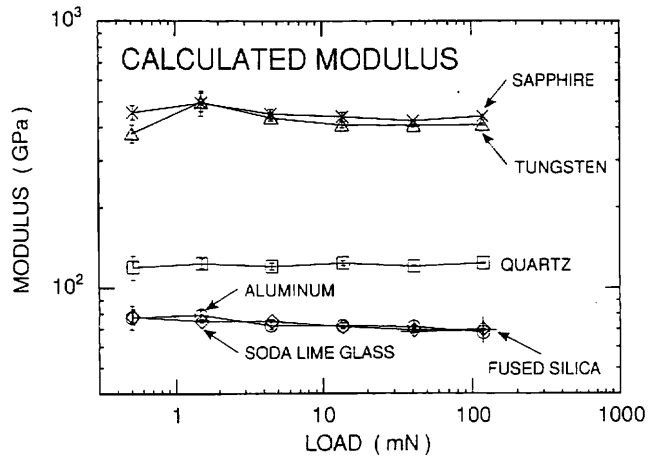
FIG. 28. The load dependence of the modulus calculated using the new analysis procedure for all six materials. The error bars are two standard deviations in length.

average elastic constants for the cubic materials) using the data tabulated in Simmons and Wang.[35]

The table shows that for aluminum, tungsten, soda lime glass, and fused quartz, the moduli computed from the indentation load-displacement data are all within 4% of the literature values. Since these materials all exhibit very isotropic elastic properties, there is no question as to what the measured modulus should be, and the good agreement between the experiment and literature moduli demonstrates that the method proposed here works well. For sapphire and quartz, on the other hand, the agreement in the table is not quite as good. The measured modulus is higher than the Voigt/Reuss average by about 9% for the sapphire and about 30% for the quartz. However, both these materials are highly anisotropic, and because of this it is not clear to what value the measured modulus should be compared. On the one hand, it can be argued that the most appropriate modulus is that in the direction of testing, since the elastic displacements are primarily in this direction. On the other hand, since the formation of the contact impression involves deformation in many directions, it can be argued that the measured modulus should be some average quantity.

Unfortunately, which of these hypotheses is more likely, if either, cannot be determined using the limited data obtained in this study. For sapphire, it could be argued that the reason that the Voigt/Reuss average of 403 GPa is low is because the measured modulus of 441 GPa is strongly influenced by the 499 GPa modulus in the direction of testing (the c-axis). The same, however, cannot be argued for quartz since the experimentally measured modulus (124 GPa) is greater than either the Voigt/Reuss average (95.0 GPa) or the c-axis modulus (105 GPa). The influence of elastic anisotropy on the measurement of modulus using indentation load-displacement methods is an area that requires further study.

VI. CONCLUSIONS

(1) Careful examination of indentation load-displacement data obtained for six materials using a Berkovich indenter reveals that unloading curves can be accurately described by the power law relation

$$P = A(h - h_f)^m$$

where P is the load, $(h - h_f)$ is the elastic displacement, and A and m are material constants. The fact that the power law exponents in the above relation are always greater than 1 (they vary from material to material in the range 1.25 to 1.51) suggests that the flat punch method of analysis for determining hardness and modulus from indentation load-displacement data is not entirely adequate. This is corroborated by dynamic measurements of contact stiffness which show that contact area changes continuously as the indenter is withdrawn.

(2) The curvature in the unloading data can be accounted for by assuming that the indenter behaves as a punch whose geometry is such that the contact area changes continuously during unloading. Conical and paraboloid punches satisfy this condition and are attractive for two other reasons: (1) analytical solutions for the elastic contact problem are available for each, and (2) each has features which model the Berkovich geometry in different ways. Careful examination of the

TABLE IV. Comparison of measured moduli with values in the literature.

Material	Experimental modulus (GPa)	Standard deviation (GPa)	Literature modulus (GPa)	Poisson's ratio	Reference
Aluminum	68.0	0.93	70.4	0.347	35
Quartz	124	0.54	95.0 (c-axis = 105)	0.077	35
Soda-lime glass	69.9	0.22	70.0	0.23	36
Sapphire	441	4.70	403 (c-axis = 499)	0.234	35
Fused silica	69.3	0.39	72	0.17	37
Tungsten	410	4.70	409.8	0.280	35

shape of unloading curves suggests that the paraboloid is the preferred geometry.

(3) Using a new analytical procedure, contact areas, hardnesses, and moduli have been computed from the indentation load-displacement data for a variety of peak loads. Comparison of predicted contact areas with values measured independently in the SEM shows that the contact area determined by the analysis procedure is a good measure of the size of the residual contact impression. As a consequence, there is good correlation between hardness measured from the load-displacement data and that measured using conventional techniques. The moduli predicted by the method are within 4% of values reported in the literature for the materials with isotropic elastic properties. For anisotropic materials. a question still remains as to what the modulus measured by load and displacement sensing indentation techniques represents.

(4) The procedure works well only when several key quantities are known with some accuracy and precision. These include the unloading stiffness, the load frame compliance, and the area function of the diamond. Procedures for determining each are outlined. The procedure for determining the area function is attractive because it does not require imaging of indentations.

ACKNOWLEDGMENTS

REFERENCES

1. J. B. Pethica, R. Hutchings, and W. C. Oliver, Philos. Mag. A **48**, 593 (1983).
2. W. C. Oliver, R. Hutchings, and J. B. Pethica, in ASTM STP 889, edited by P. J. Blau and B. R. Lawn (American Society for Testing and Materials, Philadelphia, PA, 1986), pp. 90–108.
3. M. F. Doerner and W. D. Nix, J. Mater. Res. 1, 601 (1986).
4. J. B. Pethica, in *Ion Implantation into Metals*, edited by V. Ashworth, W. Grant, and R. Procter (Pergamon Press, Oxford, 1982), pp. 147–156.
5. J. L. Loubet, J. M. Georges, O. Marchesini, and G. Meille, J. Tribology **106**, 43 (1984).
6. D. Newey, M. A. Wilkens, and H. M. Pollock, J. Phys. E: Sci. Instrum. **15**, 119 (1982).
7. D. Stone, W. R. LaFontaine, P. Alexopoulos, T-W. Wu, and Che-Yu Li, J. Mater. Res. **3**, 141 (1988).
8. J. J. Gilman, in *The Science of Hardness Testing and Its Research Applications*, edited by J. H. Westbrook and H. Conrad (American Society for Metals, Metals Park, OH, 1973), pp. 51–74.

9. W. C. Oliver, MRS Bulletin XI, 15 (1986).
10. W. C. Oliver, C. J. McHargue, and S. J. Zinkle, Thin Solid Films **153**, 185 (1987).
11. J. Boussinesq, *Applications des Potentiels a l'étude de équilibre et du mouvement des solides élastiques* (Gauthier-Villars, Paris, 1885).
12. H. Hertz, J. reine und angewandte Mathematik 92, 156 (1882).
13. A. E. H. Love, Philos. Trans. A **228**, 377 (1929).
14. A. E. H. Love, Quart. J. Math. **10**, 161 (1939).
15. K. L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge, 1985).
16. I. N. Sneddon, Int. J. Engng. Sci. 3, 47 (1965).
17. J. W. Harding and I. N. Sneddon, Proc. Cambridge Philos. Soc. **41**, 12 (1945).
18. D. Tabor, Proc. R. Soc. A 192, 247 (1948).
19. N. A. Stillwell and D. Tabor, Proc. Phys. Soc. London **78**, 169 (1961).
20. A. P. Ternovskii, V. P. Alekhin, M.Kh. Shorshorov, M. M. Khrushchov, and V. N. Skvortsov, Zavod. Lab. **39**, 1242 (1973).
21. S. I. Bulychev, V. P. Alekhin, M.Kh. Shorshorov, A. P. Ternovskii, and G. D. Shnyrev, Zavod. Lab. **41**, 1137 (1975).
22. S. I. Bulychev, V. P. Alekhin, M.Kh. Shorshorov, and A. P. Ternovskii. Prob. Prochn. **9**, 79 (1976).
23. M. Kh. Shorshorov, S. I. Bulychev, and V. P. Alekhin, Sov. Phys. Dokl. **26**, 769 (1982).
24. S. I. Bulychev and V. P. Alekhin, Zavod. Lab. **53**, 76 (1987).
25. G. M. Pharr, W. C. Oliver, and F. R. Brotzen, J. Mater. Res. **7**, 613 (1992).
26. R. B. King, Int. J. Solids Structures 3, 1657 (1987).
27. A. K. Bhattacharya and W. D. Nix, Int. J. Solids Structures 24, 881 (1988).
28. G. M. Pharr, W. C. Oliver, and D. R. Clarke, Scripta Metall. **23**, 1949 (1989).
29. G. M. Pharr, W. C. Oliver, and D. R. Clarke, J. Elec. Mater. **19**, 881 (1990).
30. G. M. Pharr, W. C. Oliver, and D. S. Harding, J. Mater. Res. **6**, 1129 (1991).
31. J. P. Hirth and J. Lothe, *Theory of Dislocations*. 2nd ed. (John Wiley and Sons, New York, 1982), p. 837.
32. J. B. Pethica and W. C. Oliver, Physica Scripta T19, 61 (1987).
33. J. B. Pethica and W. C. Oliver, in *Thin Films: Stresses and Mechanical Properties*, edited by J. C. Bravman, W. D. Nix, D. M. Barnett, and D. A. Smith (Mater. Res. Soc. Symp. Proc. **130**, 13 (1989).
34. W. C. Oliver and J. B. Pethica, U. S. Patent No. 4848141, July 1989.
35. G. Simmons and H. Wang, *Single Crystal Elastic Constants and Calculated Aggregate Properties: A Handbook*, 2nd ed. (The M. I. T. Press, Cambridge, MA, 1971).
36. G. R. Anstis, P. Chantikul, B. R. Lawn, and D. B. Marshall, J. Am. Ceram. Soc. **64**, 533 (1981).
37. General Electric Fused Quartz Products Technical Data, general catalog number 7705–7725, April 1985.

APPENDIX I: Continuous measurement of contact stiffness by a dynamic technique

The continuous measurement of stiffness during indentation is accomplished by applying a small oscillation to the force signal at a relatively high frequency (69.3 Hz). The amplitude of the force oscillation is kept sufficiently small that the deformation process is not significantly affected by its addition. The corresponding displacement oscillation is monitored at the excitation

frequency using a two-channel, phase-sensitive detector (lock-in amplifier) which can accurately measure displacements as small as 0.001 nm using frequency specific amplification. The phase difference between the displacement signal and the force signal can also be measured. A time constant of 0.33 s provides a good combination of low noise and dynamic response for the results described in this work.

Determination of the stiffness of the contact requires an understanding of the response of the entire system, a dynamic model for which is shown in Fig. A1. An analysis of this model reveals that the stiffness of the contact, S, can be calculated from the amplitude of the displacement signal from

$$\left| \frac{P_{os}}{h(\omega)} \right| = \sqrt{\left\{ (S^{-1} + C_f)^{-1} + K_s - m\omega^2 \right\}^2 + \omega^2 D^2}$$

(A1)

or from the phase difference between the force and displacement signals from

$$\tan(\phi) = \frac{\omega D}{(S^{-1} + C_f)^{-1} + K_s - m\omega^2}$$

(A2)

where

C_f = the compliance of the load frame ($\sim$1.13 m/MN)
K_s = the stiffness of the column support springs ($\sim$60 N/m)
D = the damping coefficient ($\sim$54 N s/m)
P_{os} = the magnitude of the force oscillation

DYNAMIC MODEL

FIG. A1. A dynamic model for the system.

$h(\omega)$ = the magnitude of the resulting displacement oscillation
ω = frequency of the oscillation (69.3 Hz)
ϕ = the phase angle between the force and displacement signals
m = mass ($\sim$4.7 gms)

The constants that must be known for the analysis are C_f, K_s, and D. The constants K_s and D are experimentally determined by examining the motion of the system according to the model in Fig. A1 when the indenter is not touching the surface ($S = 0$). The technique for determining K_m is an important part of the analysis technique described in the body of this work and is fully described there. Details of the damping coefficient measurement and the calibration of the system have been described elsewhere.[33,34]

For a complete understanding of the technique, several characteristics of the system should be noted. First, due to the double gap in the displacement sensing capacitive gage, the damping coefficient D varies as the fourth power of the displacement of the moving plate and is thus not constant for large plate motions. However, since the positional derivative of D near the center of travel is small, D is relatively constant over the range of positions used in a normal indentation experiment.

Second, by defining a reduced spring constant $K = (S^{-1} + C_f)^{-1} + K_s$, the model can be reduced to the simpler system shown in Fig. A2. The natural frequency for this system is
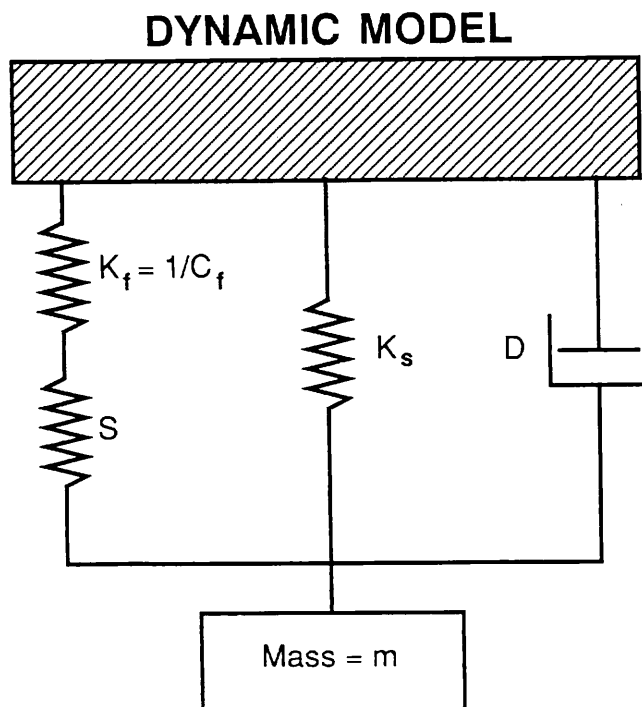
$$\omega_n = \sqrt{\frac{K}{m}}$$

(A3)

Since the critical damping coefficient is 4 Nm/s, the system is over-critically damped. This means that the displacement amplitude is single valued and no resonant peak is observed. Typically, the characteristics of the components of an oscillating system are fixed and the system response is examined by changing the frequency of oscillation. In the case considered here, the excitation frequency is held constant while the stiffness of the reduced spring constant K is increased from a very small value ($K_s = 50$ N/m) to a high value (5×10^6 N/m). This means initially the system is operating above the resonant frequency and the phase angle is greater than 90 degrees. As the stiffness S, and therefore K, increases, the resonant frequency of the system increases and approaches the excitation frequency. At the highest values of K measured in this work, the resonant frequency of the system is greater than the excitation frequency and the phase angle approaches zero. As has already been noted, even though the system goes through resonance, there is no peak in the displacement amplitude due to the high damping coefficient. Clearly the complete solutions for the motion of the system given in Eqs. (A1) and
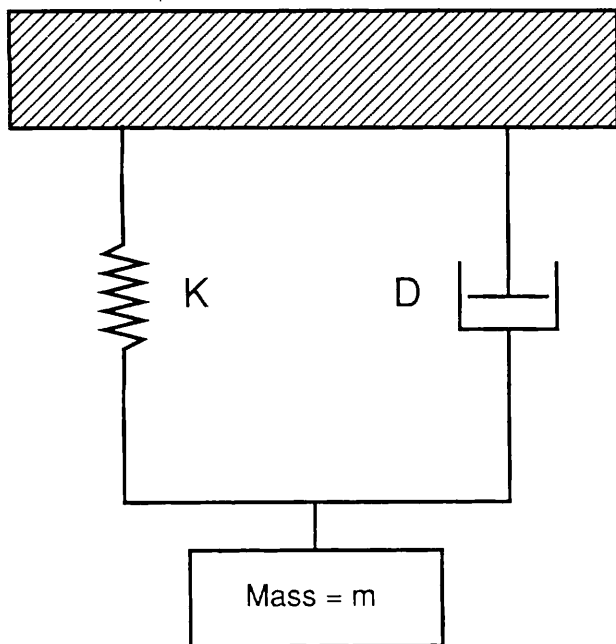
SIMPLIFIED DYNAMIC MODEL

FIG. A2. A simpler model for the system which results by redefining the spring constant.

(A2) must be used with no simplifying assumptions. In practice, the value of S calculated from the phase signal works well only at small stiffnesses where the phase angle is of some magnitude. At high stiffnesses the phase angle becomes less than a few degrees and is too small relative to the noise. The value of stiffness calculated from the amplitude measurement is more accurate, particularly at higher stiffnesses.

APPENDIX II. Indenter area function calibration procedure

The area function of the indenter relates its cross-sectional area to the distance from its tip. In the course of performing this study, a procedure was developed for determining the area function which does not require the imaging of indentations. The procedure involves making a series of indentations in two standard materials—aluminum and fused quartz—and relies on the facts that both these materials are elastically isotropic, their moduli are well known, and their moduli are independent of indentation depth.

The procedure requires that indentation contact stiffnesses be measured accurately and precisely over a wide range of indentation depths. The first step in determining these stiffnesses is to determine precisely the load frame compliance. This is best accomplished by indenting a material in which large, very stiff contacts can be made at relatively low loads. We have found that well-annealed, high purity aluminum is an excellent choice: it is readily

available, has a low hardness, and is nearly elastically isotropic.

Some care must be exercised in preparing the aluminum to assure that its surface is smooth and unaffected by work hardening. The mechanical polishing procedure we have adopted is as follows: (1) Mount the specimen in epoxy resin to form a standard metallurgical mount with a small section of heavy-walled stainless steel tubing surrounding the aluminum to prevent plastic straining when clamping forces are applied to the mount during testing. (2) Grind the specimen with successively finer grits of SiC abrasive paper through 600 grit, being sure to remove the damage layer at each step. (3) Polish for 8–12 h on a vibratory polisher using 3 μm diamond paste and water on a TEXMET^{TM} cloth followed by 8–12 h using 0.5 μm diamond and water on a MASTERTEX^{TM} cloth. (4) Final polishing is performed using colloidal silica and a MASTERTEX^{TM} cloth for 15–20 min.

A series of indentations are made in the aluminum using the first six peak loads and loading rates shown in Table AI. The load time history we recommend is as follows: (1) approach and contact surface, (2) load to peak load, (3) unload to 90% of peak load and hold for 100 s, (4) reload to peak load and hold for 10 s, and (5) unload completely at half the rate shown in Table AI. The lower hold is used to establish thermal drift and the upper hold to minimize time-dependent plastic effects. The final unloading data are used to determine the unloading stiffnesses using the power law fitting procedure described in Sec. V. B.

The load frame stiffness is determined from the aluminum data using an iterative procedure similar to that outlined in Sec. V. C. The procedure is actually somewhat simpler because the known elastic constants for aluminum, $E = 70.4$ GPa and $\nu = 0.347$, can be

TABLE AI. Peak loads and loading/unloading rates used in the area function calibration procedure.

Indentation numbers	Peak load (mN)	Loading/unloading rate (μN/s)	Nanoindenter load range
1-10	120	12000	High
11-20	60	6000	High
21-30	30	3000	High
31-40	15	1500	High
41-50	7.5	750	High
51-60	3	300	High
61-70	20	2000	Low
71-80	10	1000	Low
81-90	3	300	Low
91-100	1	100	Low
101-110	0.3	30	Low
111-120	0.1	10	Low

used to eliminate the unknown value of E_r in Eq. (18). This results in quicker convergence. The values of the elastic constants we use for the diamond indenter are $E_i = 1141$ GPa and $\nu_i = 0.07$. The procedure also produces A vs h_c data which are used in the determination of the area function at large depths.

The problem with using aluminum to extend the area function to small depths is that because of its low hardness, small indentations in aluminum require very small loads, and a limit is set by the force resolution of the indentation system. This problem can be avoided by making the small indentations in fused quartz, a much harder, isotropic material available in optically finished plate form from the General Electric Corporation. The standard procedure we use for determining the area function involves making a series of indentations in fused quartz using all the peak loads shown in Table AI. The contact areas and contact depths are then determined

using Eqs. (18) and (8) in conjunction with the reduced modulus computed from the elastic constants for fused quartz, $E = 72$ GPa and $\nu = 0.170$. Because the machine compliance is known from the aluminum analysis, there is no need for iteration.

To finally establish the area function, the A vs h_c data obtained for both the aluminum and the fused quartz are fit as a group to Eq. (19). The area function at large depths is then determined primarily by the aluminum data while that at small depths is determined by the fused quartz data. A convenient fitting routine is that contained in the Kaleidagraph™ software for Apple Macintosh™ computers. A weighted procedure can be used to assure that data points with small and large magnitudes are of equal importance. For the loads outlined in Table AI, the minimum contact depth is about 15 nm and the maximum about 4700 nm, so the area function is good in the range in between.