REVIEW

This section of Journal of Materials Research is reserved for papers that are reviews of literature in a given area.

Measurement of hardness and elastic modulus by instrumented indentation: Advances in understanding and refinements to methodology

W.C. Oliver
MTS Systems Corporation, Oak Ridge, Tennessee, 37830

G.M. Pharr[a)]
The University of Tennessee and Oak Ridge National Laboratory, Department of Materials Science and Engineering, Knoxville, Tennessee 37996

(Received 15 June 2003; accepted 23 September 2003)

The method we introduced in 1992 for measuring hardness and elastic modulus by instrumented indentation techniques has widely been adopted and used in the characterization of small-scale mechanical behavior. Since its original development, the method has undergone numerous refinements and changes brought about by improvements to testing equipment and techniques as well as from advances in our understanding of the mechanics of elastic–plastic contact. Here, we review our current understanding of the mechanics governing elastic–plastic indentation as they pertain to load and depth-sensing indentation testing of monolithic materials and provide an update of how we now implement the method to make the most accurate mechanical property measurements. The limitations of the method are also discussed.

I. INTRODUCTION

The method we introduced in 1992 for measuring hardness and elastic modulus by instrumented indentation techniques has widely been adopted and used in the characterization of mechanical behavior of materials at small scales.[1,2] Its attractiveness stems largely from the fact that mechanical properties can be determined directly from indentation load and displacement measurements without the need to image the hardness impression. With high-resolution testing equipment, this facilitates the measurement of properties at the micrometer and nanometer scales.[3–5] For this reason, the method has become a primary technique for determining the mechanical properties of thin films and small structural features.[6–23] Films with characteristic dimensions of the order 1 μm are now routinely measured, and with good technique, the method can be used to characterize, at least in a comparative sense, the properties of films as thin as a few nanometers.

During the past decade, we have made several important changes to the method that both improve its accuracy and extend its realm of application. These changes have been developed both through experience in testing a large number of materials and by improvements to testing equipment and techniques. For example, the measurement of contact stiffness by dynamic techniques allows for continuous measurement of properties as a function of depth and also facilitates more accurate identification of the point of first surface contact.[24] We have also developed improved methods for calibrating indenter area functions and load frame compliances. Thus, the primary purpose of this article is to provide the reader with an update on how we implement the method in practice now and the improvements we have made to it during the past 10 years.

We have also developed a much better understanding of the contact mechanics on which the method is based, mostly through finite element simulation.[25–36] Thus, a second objective of the article is to review what we now know about the mechanics of elastic–plastic contact and how this impacts the measurement methods. For instance, we now have a much better grasp of the physical origin of some of the empirical constants needed in the method,[35] and we also have a much better understanding of the limitations of the method, particularly in materials that pile-up.[25,34,37]

[a)]This author was an editor of this focus issue during the review and decision stage. For the *JMR* policy on review and publication of manuscripts authored by editors, please refer to http://www.mrs.org/publications/jmr/policy.html

To keep the article to a reasonable length, we have intentionally limited its focus to the behavior of monolithic materials that can be described as a semi-infinite, elastic–plastic half spaces. There are, of course, important issues that must be addressed when applying the method to thin films to account for the influences of the substrate. In fact, much of the recent literature on load and depth-sensing indentation testing deals with this subject, and there are numerous important unresolved issues remaining to be explored.[11–23] We also limit the discussion to materials in which there are no time-dependent deformation mechanisms such as creep or viscoelasticity. Here, too, there are important issues under investigation.[38–42]

To set the stage, we begin with a brief review of the method as it was originally developed. This is then followed by a discussion of advances in our understanding of the contact problem and then details of the method as we now use and apply it.

II. THE METHOD

A. Basic principles

The method was developed to measure the hardness and elastic modulus of a material from indentation load–displacement data obtained during one cycle of loading and unloading. Although it was originally intended for application with sharp, geometrically self-similar indenters like the Berkovich triangular pyramid, we have since realized that it is much more general than this and applies to a variety of axisymmetric indenter geometries including the sphere. A discussion of why the method works for spherical indentation is given at the end of this section.

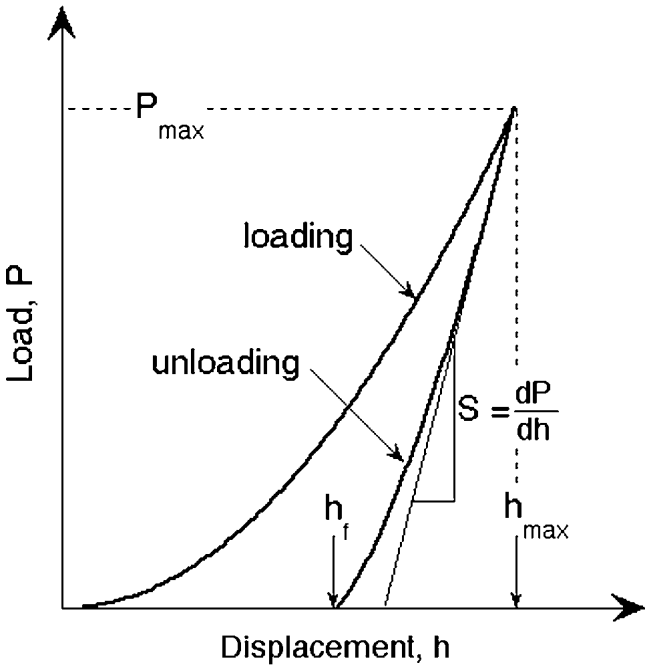

FIG. 1. Schematic illustration of indentation load–displacement data showing important measured parameters (after Ref. 1).

A schematic representation of a typical data set obtained with a Berkovich indenter is presented in Fig. 1, where the parameter P designates the load and h the displacement relative to the initial undeformed surface. For modeling purposes, deformation during loading is assumed to be both elastic and plastic in nature as the permanent hardness impression forms. During unloading, it is assumed that only the elastic displacements are recovered; it is the elastic nature of the unloading curve that facilitates the analysis. For this reason, the method does not apply to materials in which plasticity reverses during unloading. However, finite element simulations have shown that reverse plastic deformation is usually negligible.[35]

There are three important quantities that must be measured from the P–h curves: the maximum load, $P_{\max}$, the maximum displacement, $h_{\max}$, and the elastic unloading stiffness, $S = \mathrm{d}P/\mathrm{d}h$, defined as the slope of the upper portion of the unloading curve during the initial stages of unloading (also called the contact stiffness). The accuracy of hardness and modulus measurement depends inherently on how well these parameters can be measured experimentally. Another important quantity is the final depth, h_f, the permanent depth of penetration after the indenter is fully unloaded.

The analysis used to determine the hardness, H, and elastic modulus, E, is essentially an extension of the method proposed by Doerner and Nix[43] that accounts for the fact that unloading curves are distinctly curved in a manner that cannot be accounted for by the flat punch approximation. In the flat punch approximation used by Doerner and Nix, the contact area remains constant as the indenter is withdrawn, and the resulting unloading curve is linear. In contrast, experiments have shown that unloading curves are distinctly curved and usually well approximated by the power law relation:

$$P = \alpha(h - h_f)^m \quad , \tag{1}$$

where α and m are power law fitting constants.[1] Table I summarizes the values of the constants observed in our

TABLE I. Values of parameters characterizing unloading curves as observed in nanoindentation experiments with a Berkovich indenter.[a]

Material	α (mN/nm^m)	m	Correlation coefficient, R
Aluminum	0.265	1.38	0.999938
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
Silica	0.0215	1.43	0.999985

[a]Data from Ref. 1.

original experiments.[1] The variation of the power law exponents in the range $1.2 \leq m \leq 1.6$ demonstrates not only that the flat punch approximation is inadequate ($m = 1$ for the flat punch), but also that the indenter appears to behave more like a paraboloid of revolution, for which $m = 1.5$.[44] This result was somewhat surprising because the axisymmetric equivalent of the Berkovich indenter is a cone, for which $m = 2$. This discrepancy has since been explained by the concept of an "effective indenter shape," which will be discussed in Sec. III. A.[35,36]

The exact procedure used to measure H and E is based on the unloading processes shown schematically in Fig. 2, in which it is assumed that the behavior of the Berkovich indenter can be modeled by a conical indenter with a half-included angle, ϕ, that gives the same depth-to-area relationship, $\phi = 70.3°$. The basic assumption is that the contact periphery sinks in in a manner that can be described by models for indentation of a flat elastic half-space by rigid punches of simple geometry.[44–48] This assumption limits the applicability of the method because it does not account for the pile-up of material at the contact periphery that occurs in some elastic–plastic materials. Assuming, however, that pile-up is negligible, the elastic models show that the amount of sink-in, h_s, is given by:

$$h_s = \epsilon \frac{P_{max}}{S} \quad , \tag{2}$$

where ϵ is a constant that depends on the geometry of the indenter. Important values are: $\epsilon = 0.72$ for a conical punch, $\epsilon = 0.75$ for a paraboloid of revolution (which approximates to a sphere at small depths), and $\epsilon = 1.00$ for a flat punch.[44] Based on the empirical observation that the unloading curves are best approximated by an indenter that behaves like a paraboloid of revolution ($m = 1.5$), the value $\epsilon = 0.75$ was recommended and has since become the standard value used for analysis. Recent analytical work has provided a more complete picture of why ϵ takes on this value.[35,49] This is also addressed in Sec. III. A.

Using Eq. (2) to approximate the vertical displacement of the contact periphery, it follows from the geometry of Fig. 2 that the depth along which contact is made between the indenter and the specimen, $h_c = h_{max} - h_s$, is:

$$h_c = h_{max} - \epsilon \frac{P_{max}}{S} \quad . \tag{3}$$

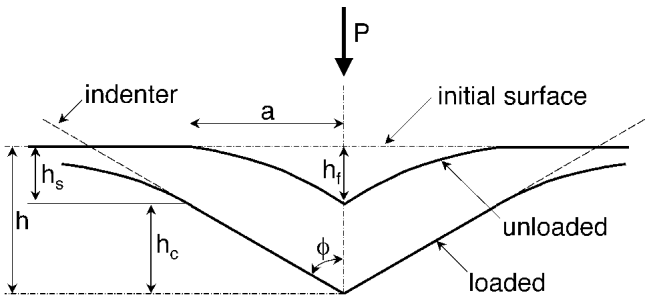

FIG. 2. Schematic illustration of the unloading process showing parameters characterizing the contact geometry (after Ref. 1).

Letting $F(d)$ be an "area function" that describes the projected (or cross sectional) area of the indenter at a distance d back from its tip, the contact area A is then

$$A = F(h_c) \quad . \tag{4}$$

The area function, also sometimes called the indenter shape function, must carefully be calibrated by independent measurements so that deviations from nonideal indenter geometry are taken into account. These deviations can be quite severe near the tip of the Berkovich indenter, where some rounding inevitably occurs during the grinding process. Although a basic procedure for determining the area function was presented as part of the original method, we have made significant changes to it in recent years. A complete description of the process we now use in its place is presented in Sec. IV.

Once the contact area is determined, the hardness is estimated from:

$$H = \frac{P_{max}}{A} \quad . \tag{5}$$

Note that because this definition of hardness is based on the contact area under load, it may deviate from the traditional hardness measured from the area of the residual hardness impression if there is significant elastic recovery during unloading. However, this is generally important only in materials with extremely small values of E/H.[25]

Measurement of the elastic modulus follows from its relationship to contact area and the measured unloading stiffness through the relation

$$S = \beta \frac{2}{\sqrt{\pi}} E_{eff} \sqrt{A} \quad , \tag{6}$$

where E_{eff} is the effective elastic modulus defined by

$$\frac{1}{E_{eff}} = \frac{1 - \nu^2}{E} + \frac{1 - \nu_i^2}{E_i} \quad . \tag{7}$$

The effective elastic modulus takes into account the fact that elastic displacements occur in both the specimen, with Young's modulus E and Poisson's ratio ν, and the indenter, with elastic constants E_i and ν_i. Note that Eq. (6) is a very general relation that applies to any axisymmetric indenter.[2,50] It is not limited to a specific simple geometry, even though it is often associated with flat punch indentation. Although originally derived for elastic contact only,[2] it has subsequently been shown to apply equally well to elastic–plastic contact[50] and that

small perturbations from pure axisymmetry geometry do not effect it either.[51] It is also unaffected by pile-up and sink-in.

In the original method for measuring hardness and modulus, the dimensionless parameter β was taken as unity. Traditionally, β has been used to account for deviations in stiffness caused by the lack of axial symmetry for pyramidal indenters. However, it has been shown that even for indentation of an elastic half-space by axisymmetric rigid cone, β can deviate significantly from unity.[26] In this work, we therefore choose to redefine β to account for any and all physical processes that may affect the constant in Eq. (6). In Sec. III. E, other values for β that may be important in precise measurements will be considered along with the physical nature of their origins.

B. Application to spherical indentation

Although we have discussed the method primarily as it pertains to indentation with a sharp Berkovich or conical indenter, it is not generally recognized that the method applies equally well to a sphere and can be used without modification to determine the hardness and modulus from spherical indentation. To demonstrate this, it is useful to consider the contact depth for spherical indentation determined by means of Eq. (3) when applied to Hertzian elastic contact by a spherical indenter of radius R_1 pressed into a spherical hole of radius R_2 that represents the hardness impression. Note that Hertzian contact analysis is restricted to the condition that the depth of penetration is small relative to the radius of the sphere. Letting $R = (1/R_1 + 1/R_2)^{-1}$, the load–displacement relation is[52,53]

$$P = \tfrac{4}{3}\sqrt{R}E_{\text{eff}}(h - h_{\text{f}})^{3/2} \quad . \tag{8}$$

The depth that appears in this relation is $h - h_{\text{f}}$ rather than h because the displacements are for the elastic unloading curve only (Fig. 1). The stiffness during unloading is then found by differentiating this expression with respect to h, or

$$S = \frac{\mathrm{d}P}{\mathrm{d}h} = 2\sqrt{R}E_{\text{eff}}(h - h_{\text{f}})^{1/2} \quad . \tag{9}$$

Letting $h = h_{\max}$ to evaluate the expressions at the maximum depth and noting that ε for a sphere is 0.75, substituting Eqs. (8) and (9) into Eq. (3) yields

$$h_{\text{c}} = \frac{h_{\max} + h_{\text{f}}}{2} \quad , \tag{10}$$

or the contact depth determined from the method is simply the average of the final and maximum depths. This is precisely the value recommended by Field and Swain in their method for analyzing load and depth-sensing data obtained with spherical indenters.[54,55] The contact area and hardness determined by the two methods will thus be the same as will be the elastic modulus because Eq. (6) applies to any indenter geometry. Note that hardnesses measured with a spherical indenter are not necessarily the same as those for the Berkovich.

III. ADVANCES IN UNDERSTANDING

Although the method has extensively been used and verified for numerous materials, certain aspects of it have always been of concern to us, either because the physical processes were not completely understood or because some of the constants needed to apply it were empirically derived. In particular, we have always been interested in understanding: (i) why unloading curves are well described by the power law relation of Eq. (1); (ii) why the power law exponent in Eq. (1) falls roughly in the range $1.2 \leqslant m \leqslant 1.6$ rather than taking on the value $m = 2$ as would be expected for a conical indenter; and (iii) why the best value for the geometric parameter ε is about 0.75. Moreover, from a more fundamental standpoint, it was not entirely clear to us how the elastic contact solutions used to derive Eqs. (3) and (6) could be used to model accurately the elastic unloading process as they apply to the indentation of a *flat* elastic half-space, whereas the real problem involves a half space whose surface has severely been distorted by the formation of a hardness impression. These questions have been answered through the concept of an "effective indenter shape."[35,36]

A. The effective indenter shape

The ideas underlying the effective indenter shape are outlined in Fig. 3. The basic principles are derived from observations gleaned from finite element simulations of indentation of elastic–plastic materials by a rigid conical indenter with a half included angle of 70.3°.[35] During the initial loading of the indenter [Fig. 3(a)], both elastic and plastic deformation processes occur, and the indenter conforms perfectly to the shape of the hardness impres-

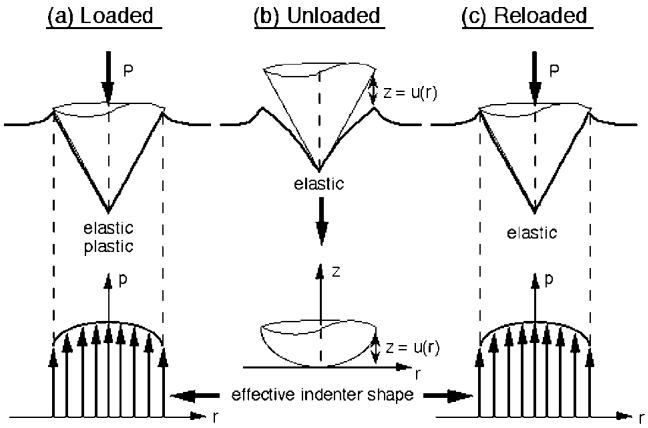

FIG. 3. Concepts used to understand and define the effective indenter shape (from Ref. 35).

sion. However, during unloading [Fig. 3(b)], elastic recovery causes the hardness impression to change its shape. A key observation is that the unloaded shape is not perfectly conical, but exhibits a subtle convex curvature that has been exaggerated in Fig. 3(b) for the purposes of illustration. The importance of the curvature is that as the indenter is elastically reloaded [Fig. 3(c)], the contact area increases gradually and continuously until full load is again achieved, a process which must be the reverse of what happens during unloading because both processes are elastic. It is this continuous change in contact area that produces the nonlinear unloading curves. Furthermore, the relevant elastic contact problem is not that of conical indenter on a flat surface, but a conical indenter pressed into a surface that has been distorted by the formation of the hardness impression.

Experimental evidence for the curvature in the unloaded hardness impressions is given in Fig. 4, where topographical images of a Berkovich indentation in fused silica are presented. The images were obtained by scanning the specimen laterally under the indenter tip with a very light load applied to the indenter to keep it in contact. The three-dimensional (3D) image in Fig. 4(a) clearly shows the curvature of the indentation faces as does the cross section in Fig. 4(b).

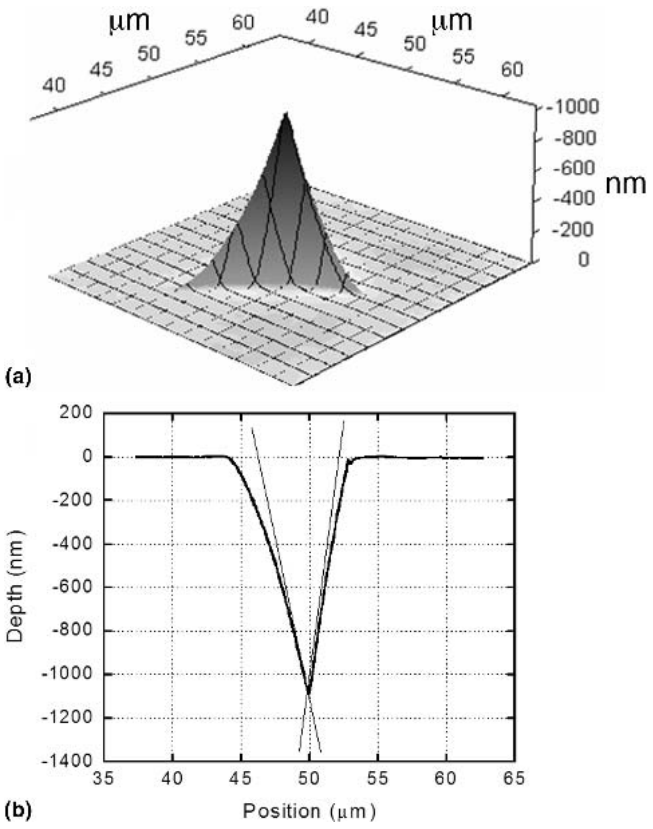

FIG. 4. Topographic image of a Berkovich indentation in fused silica obtained using a quantitative imaging system that uses the indenter itself as the imaging tip: (a) complete 3D image showing the curvature of the faces; (b) cross section through an edge and the opposing face with straight lines included to accentuate the curvature.

The mathematical form of the unloading curve can be understood by introducing an "effective indenter shape." As shown in the lower portion of Fig. 3, the shape of the effective indenter is that which produces the same normal surface displacements on a flat surface that would be produced by the conical indenter on the unloaded, deformed surface of the hardness impression. As such, the shape is described by a function $z = u(r)$, where $u(r)$ is the distance between the conical indenter and the unloaded deformed surface and r is the radial distance from the center of contact. Thus, provided the shape of the deformed surface is known, the function $u(r)$ can be found and the effective indenter can be constructed. Finite element simulations have verified that constructing the effective indenter in this way provides an accurate representation of the unloading data.[35,36] Although the exact shape of the effective indenter depends in complicated ways on the elastic and plastic deformation characteristics of the material, it always has a smooth, rounded profile at its tip. The reason for this is that the slope of the unloaded hardness impression at its center exactly matches the slope of the conical indenter (i.e., the two surfaces conform perfectly at the tip of the indenter). Thus, the plastic deformation producing the hardness impression has the interesting effect of removing the elastic singularity at the tip of the indenter. Finite element studies have also shown that the effective indenter shape is well approximated by the power-law relation

$$z = Br^n \quad , \tag{11}$$

where B is a fitting constant and the exponent n varies in the range 2–6 depending on the material properties.[35] The lower end of this range corresponds to an indenter with the shape of a paraboloid of revolution, thus explaining why indentation with a rigid cone is more like that of a paraboloid of revolution or a sphere.

One can also use the effective shape concept to understand why unloading curves are well described by the power-law relation of Eq. (1) with power-law exponents m in the range 1.2–1.6. Sneddon[44] has shown that for indentation of an elastic half-space by an axisymmetric rigid indenter described by the power-law relation of Eq. (11), the load–displacement relation is

$$P = \frac{2E_{\text{eff}}}{(\sqrt{\pi}B)^{1/n}}\left(\frac{n}{n+1}\right)\left[\frac{\Gamma(n/2+1/2)}{\Gamma(n/2+1)}\right]^{1/n} h^{1+1/n} \quad , \tag{12}$$

where Γ is the factorial or "gamma" function. Comparison with Eq. (1) shows that

$$m = 1 + 1/n \quad . \tag{13}$$

Thus, given that n values range from 2–6, one would expect ms in the range 1.2–1.5, in accordance with the experimental observations in Table I.

From a more fundamental standpoint, the shape of the effective indenter can be estimated by making simple assumptions about the distribution of pressure under the indenter. As schematically illustrated in the lower portion of Fig. 3, the pressure that develops during initial loading is determined by complex elastic and plastic deformation processes that are generally not amenable to closed-form analysis. However, during unloading, the pressure decreases by elastic processes only, and the shape of the hardness impression changes to produce the curved surface. Upon reloading, this pressure distribution must be recovered by elastic processes only. Thus, the pressure distribution at peak load serves to link the elastic–plastic processes during initial loading to the elastic processes during unloading and reloading. In this context, the shape of the effective indenter must be that which produces this same pressure distribution by elastic deformation of flat elastic half-space.

To implement these ideas, one must have some knowledge of what the actual pressure distribution is. Finite element simulations of conical indentation of elastic–plastic materials have shown that to a first approximation, the pressure is roughly uniform.[35] This occurs because plasticity tends to diminish the effects of the elastic singularity at the tip of the cone and more evenly distribute the pressure. Assuming, then, a perfectly flat pressure distribution, elastic contact theory gives the effective indenter shape as

$$z(r) = \frac{4pa_{\max}}{\pi E_{\text{eff}}}\left[\frac{\pi}{2} - \mathbf{E}(r/a_{\max})\right] \quad , \qquad (14)$$

where p is the pressure, $a_{\max}$ is the radius of the contact circle at peak load, and $\mathbf{E}(r/a_{\max})$ is the complete elliptic integral of the second kind evaluated at $r/a_{\max}$.[35,53] Curve-fitting procedures show that this shape is very well approximated by the power-law relation:[35]

$$z(r) = 0.548\frac{4pa_{\max}}{\pi E_{\text{eff}}}(r/a_{\max})^{2.61} \quad . \qquad (15)$$

Note that this equation is exactly of the form of Eq. (11) with $n = 2.61$, for which the corresponding value of m according to Eq. (13) is $m = 1.38$. Thus, to the extent that the pressure distribution is constant, there is a simple explanation for why the effective indenter behaves more like a paraboloid of revolution (with $n = 2.61$) than a cone and why unloading curves obtained with conical indenters can be described by power-law relations like Eq. (1) with $m \approx 1.38$.

The effective shape concept can also be used to develop an equation that describes the entire unloading curve in terms of fundamental material properties. Combining Eqs. (11), (12), and (15) and noting that the pressure p is equivalent to the hardness, H, the unloading curve is given by

$$\frac{P}{P_{\max}} = 0.858\left[\frac{E_{\text{eff}}}{\sqrt{P_{\max}H}}(h - h_f)\right]^{1.38} \quad , \qquad (16)$$

which compares relatively well to experimental data.[35] Note that although the concepts discussed here were developed specifically for conical indentation, they probably apply to other indenter geometries as well, as plasticity has the effect of flattening the pressure distribution. Thus, one might expect unloading curves for spherical and other indenters to be well approximated by this same equation.

Lastly, the effective indenter shape concept is also valuable in the way it provides a physically justifiable procedure for determining the value of the geometric parameter ϵ needed for accurate measurement of H and E.[35,49] Using Sneddon's method for determining the surface displacement at the contact perimeter for indentation with a rigid punch with profile given by Eq. (11) in conjunction with Eq. (2) yields[35]:

$$\epsilon = m\left[1 - \frac{2\Gamma\left(\frac{m}{2(m-1)}\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{2(m-1)}\right)}(m-1)\right] \quad . \qquad (17)$$

This relation is plotted in Fig. 5. Note that over the range of expected ms ($m = 1.2$ to 1.6), ϵ varies only mildly between 0.74 and 0.79 with an average value of 0.76. Thus, the value typically used in experiment, $\epsilon = 0.75$, is a reasonable estimate, although one could easily rationalize slightly different values. Alternatively, a better, more self-consistent approach would be to measure experimentally the exponent m from the unloading curve and then use Eq. (17) to determine the relevant value of ϵ. This could easily be implemented in experiment.

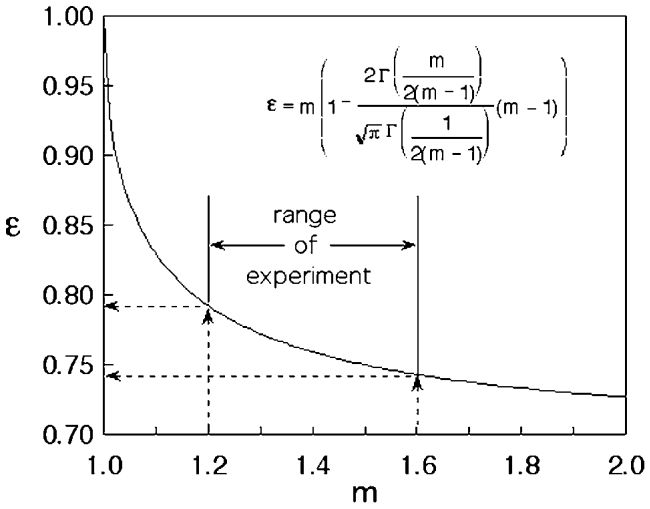

FIG. 5. The relation between ϵ and m given by Eq. (17) (from Ref. 35).

B. Errors due to pile-up for conical and Berkovich indenters

As noted in our discussion of Eq. (2), one significant problem with the method is that it does not account for pile-up of material around the contact impression, as is observed in many elastic–plastic materials. When pile-up occurs, the contact area is greater than that predicted by the method, and both the hardness estimated from Eq. (5) and the modulus from Eq. (6) are overestimated, sometimes by as much as 50%.[25] This inability to deal with pile-up is a direct consequence of using an elastic contact analysis to determine the contact depth. Because materials deformed elastically always sink-in, pile-up cannot properly be modeled.

The types of materials and conditions for which pile-up is most likely to occur have been examined by finite element simulation.[25] The fundamental material properties affecting pile-up are the ratio of the effective modulus to the yield stress, E_{eff}/σ_y, and the work-hardening behavior. In general, pile-up is greatest in materials with large E_{eff}/σ_y and little or no capacity for work hardening (i.e., "soft" metals that have been cold-worked prior to indentation). The ability to work harden inhibits pile-up because as material at the surface adjacent to the indenter hardens during deformation, it constrains the upward flow of material to the surface.

Results of finite element studies that illustrate the influence of material parameters on pile-up and sink-in are presented in Fig. 6,[25] which shows contact profiles under load for indentation by a 70.3° rigid cone. The materials examined all had a Young's modulus $E = 70$ GPa and Poisson's ratio $\nu = 0.25$, but the yield stresses were systematically varied between $\sigma_y = 0.114$ GPa and $\sigma_y = 26.62$ GPa to examine different plastic behaviors. Two separate cases of work hardening were considered: one with no work hardening, that is, a work-hardening rate $\eta = d\sigma/d\epsilon = 0$, (an elastic–perfectly plastic material) and the other with a linear work-hardening rate $\eta = 10\sigma_y$. These yield strengths and work-hardening rates cover a wide range of metals and ceramics.

In the course of the study, it was found that there is a convenient, experimentally measurable parameter that can be used to identify the expected indentation behavior of a given material.[25] The parameter is the ratio of final indentation depth, h_f, to the depth of the indentation at peak load, $h_{\max}$. The ratio $h_f/h_{\max}$ can be extracted easily from the unloading curve in a nanoindentation experiment. Furthermore, because conical and Berkovich indenters have self-similar geometries, $h_f/h_{\max}$ does not depend on the depth of indentation. The natural limits for the parameter are $0 \leq h_f/h_{\max} \leq 1$. The lower limit corresponds to fully elastic deformation and the upper limit to rigid–plastic behavior. A similar approach based on the ratio of the slopes of the loading and unloading curves, a quantity that can be measured continuously during loading by dynamic stiffness measurement, has also been developed.[56]

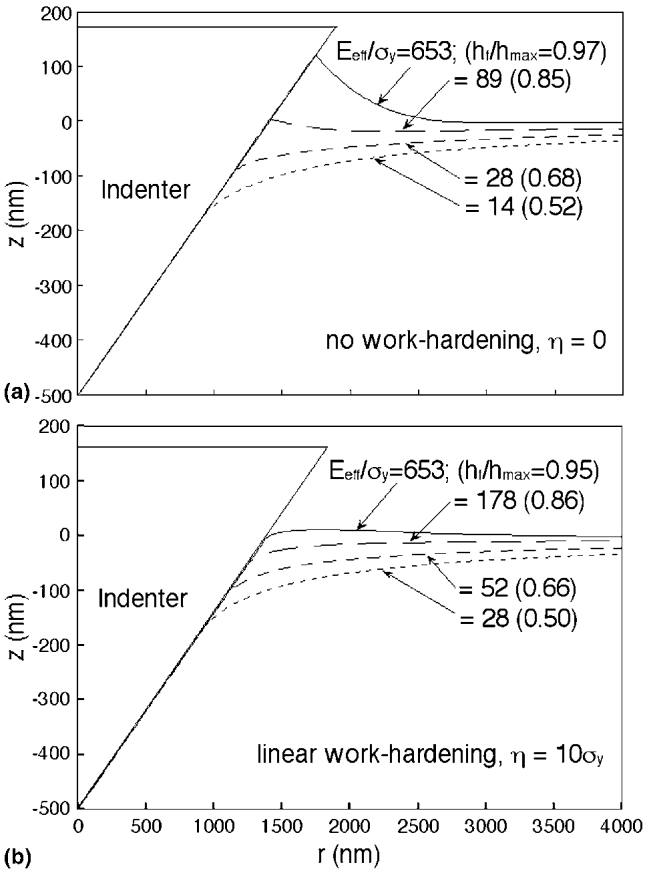

FIG. 6. Finite element simulation of contact profiles under load for conical indentation of (a) non-work-hardening materials and (b) linear work-hardening materials with work-hardening rate $\eta = 10\ \sigma_y$ (after Ref. 25).

The results in Fig. 6 show that the amount of pile-up or sink-in depends on $h_f/h_{\max}$ and the work-hardening behavior. Specifically, pile-up is large only when $h_f/h_{\max}$ is close to 1 and the degree of work hardening is small. It should also be noted that when $h_f/h_{\max} < 0.7$, very little pile-up is observed no matter what the work-hardening behavior of the material.

These observations are particularly important when considered in relation to the contact areas shown in Fig. 7. This figure includes two separate measures of the contact area: one obtained by direct examination of the contact profiles in the finite element mesh, A_{true}, and the other by applying Eqs. (3) and (4) to the simulated indentation load–displacement data to obtain the area deduced from the data analysis method, A_{expt}. To generalize, each area has been normalized with respect to A_{af}, the area given by the indenter area function evaluated at the maximum indentation depth, $h_{\max}$. Since A_{af} is the area that would occur in the absence of pile-up or

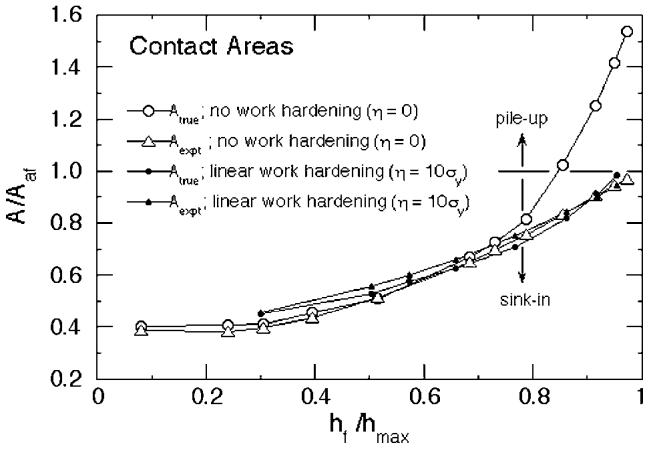

FIG. 7. Dependence of normalized contact areas on the experimental parameter h_f/h_{max}. The areas are normalized with respect to A_{af}, the area given by the area function of the indenter assuming no pile-up or sink-in (after Ref. 25).

sink-in, values of A/A_{af} greater than 1 indicate pile-up, whereas values less than 1 indicate sink-in. From this figure it is seen that the data analysis method, because it is based on an elastic analysis, significantly underestimates the true contact area for materials in which pile-up is important. Furthermore, when $h_f/h_{max} > 0.7$, the accuracy of the method depends on the amount of work hardening in the material. If the material is elastic–perfectly plastic, the method underestimates the contact area by as much as 50%. On the other hand, contact areas for materials that work harden are predicted very well by the method.

Note that from an experimental point of view, it is not possible to predict if a material work hardens based solely on the load–displacement data. Therefore, in an indentation experiment, care must be exercised when $h_f/h_{max} > 0.7$, as use of the method can lead to large errors in the contact area. On the other hand, when the pile-up is small (i.e., $h_f/h_{max} < 0.7$), the contact areas given by the method match very well with the true contact areas obtained from the finite element analyses, independent of the work-hardening characteristics. Because the hardness of the material is given by $H = P/A$, errors in the contact area will lead to similar errors in the hardness, and according to Eq. (6), the modulus will be in error by a factor that scales as $\sqrt{A}$.

As a practical matter, if there is suspicion that pile-up may be important based on the value of h_f/h_{max} and/or other independent knowledge of the properties of the material, indentations should be imaged to examine the extent of the pile-up and establish the true area of contact. For Berkovich indenters, indentations with a large amount of pile-up can be identified by the distinct bowing out at the edges of the contact impression.[57] If pile-up is large, accurate measurements of H and E cannot be obtained using the contact area deduced from the load–displacement data; rather, the area measured from the image should be used to compute H and E from Eqs. (5) and (6).

C. Correcting for pile-up

Developing a method that can be used to correct for pile-up in a manner that does not involve imaging the contact impression has been one of the "holy grails" of instrumented indentation research. The basic problem in achieving this goal is that the amount of pile-up and sink-in depends on the work-hardening characteristics of the material, so without some independent knowledge of the work-hardening behavior, one does not know what correction to apply.

A good example of work in this area is that of Cheng and Cheng,[29–31] who used finite element simulations of a wide variety of elastic–plastic materials with different work-hardening behaviors to examine pile-up during indentation with a 68° cone (68° is the centerline-to-face angle for a Vickers indenter; although a 70.3° cone is more representative of Berkovich indentation, the basic results of this work are still relevant). The method they proposed to account for pile-up is based on the work of indentation,[31] which can be measured from the areas under indentation loading and unloading curves. Letting W_{tot} be the total work of indentation (the area under the loading curve) and W_u the work recovered during unloading (the area under the unloading curve), they found that the ratio of the irreversible work to the total work, $(W_{tot} - W_u)/W_{tot}$, *appears* to be a unique function of E_{eff}/H, independent of the work-hardening behavior.[31] Although their results are presented only in graphical form, the relation can be approximated as:

$$\frac{W_{tot} - W_u}{W_{tot}} \cong 1 - 5\frac{H}{E_{eff}} \quad . \tag{18}$$

Combining Eqs. (5) and (6) and taking $\beta = 1$ leads to another equation involving H and E_{eff}:

$$\frac{4}{\pi}\frac{P_{max}}{S^2} = \frac{H}{E_{eff}^2} \quad . \tag{19}$$

Because W_{tot}, W_u, P_{max}, and S are all measurable from load–displacement data, Eqs. (18) and (19) represent two independent relations that can be solved for H and E_{eff} in a manner that does not directly involve the contact area. However, the contact area could be computed from the derived value of H by means of Eq. (5). Presumably, this area would be the true area including the effects of pile-up.

To the best of our knowledge, this method has never been tested experimentally. One potential problem is that although pile-up is large and potentially an issue only when E_{eff}/H and E_{eff}/σ_y are relatively large (Fig. 6) (i.e.,

in soft metals), close inspection of the data used to derive Eq. (18) shows that $(W_{tot} - W_u)/W_{tot}$ may not be entirely independent of the work-hardening behavior in this regime. If this is the case, then the method suffers from exactly the same problem as others, that is, one must have an independent knowledge of the work-hardening behavior to apply the corrections. Further experimental examination of this issue is warranted.

D. Pile-up for spherical indenters

We would also like to comment briefly on the issue of pile-up during spherical indentation, as the method can be applied to this important indenter geometry as well. Spherical indentation differs from conical or pyramidal indentation in that there is no elastic singularity at the tip of the indenter to produce large stresses. As a consequence, deformation at small loads and displacements is entirely elastic and then transitions to plastic as the indenter is driven further into the material. This feature of spherical indentation is what allows stress–strain curves to be approximated from indentation data using the classical approach of Tabor.[58] Field and Swain[54,55] have applied Tabor's approach to instrumented indentation and have developed a method that uses the indentation load–displacement data to approximate the stress–strain curve and the work-hardening exponent. The approach requires that the deformation be "fully plastic"; that is, beyond the transition from elastically dominated to plastic-dominated deformation. Also, it is assumed that in the fully plastic regime, the pile-up geometry is uniquely related to material deformation parameters.

A point we would like to make is that the pile-up geometry can change considerably during the course of spherical indentation and it is therefore not possible to predict the pile-up based on the mechanical properties of the material alone, even when "fully plastic" deformation is achieved. The results of a recent finite element study help to elucidate how the pile-up changes with the depth of penetration.[34] Figure 8 shows the pile-up profiles for an elastic–perfectly plastic material with $E/\sigma_y = 200$, which is fairly typical for a metal. To present the profiles on a single set of axes, the z coordinates have been normalized with respect to the depth, h, the r coordinates with respect to the contact radius, a, and the depth of penetration h with respect to the radius of the indenter, R. The pile-up and sink-in behavior are conveniently characterized by the parameter s/h, where s is the height of the contact periphery relative to the undeformed surface ($s/h < 0$ for sink-in and $s/h > 0$ for pile-up).

At small depths, the material deforms only elastically, with the indentation profile corresponding to that of Hertzian contact with sink-in at the contact periphery such that $s/h = -0.5$. First yielding occurs at $h/R \approx 1.4 \times 10^{-4}$ and a mean pressure $p_m \approx 1.1\sigma_y$ in a region on the axis of symmetry below the surface at $z \approx 0.5a$, also in accordance with Hertzian contact theory.[52,53] As the load on the indenter is increased, the plastic zone grows and spreads upward, and the sink-in diminishes. At h/R of about 0.08, material begins to pile-up, and with further penetration, the pile-up grows larger. It should be noted, however, that an upper limiting value for s/h was not reached in the simulations, even at the maximum depth of penetration, $h/R = 0.34$, corresponding to $a/R = 0.75$. This is significant because it shows that the pile-up geometry can continue to change even up to contact radii in excess of $0.75R$, a value near the upper end of most indentation experiments. Thus, unlike the constraint factor, p_m/σ_y, which plateaus at an upper value of approximately 3 at the onset of the "fully plastic" deformation regime,[33,34] pile-up can continue to grow well after full plasticity is achieved. Therefore, it does not follow that just because the constraint factor has become constant that the pile-up geometry can also be assumed to be constant. This ever-changing nature of the pile-up geometry should carefully be considered when interpreting results from spherical indentation experiments in materials with a tendency to pile-up.

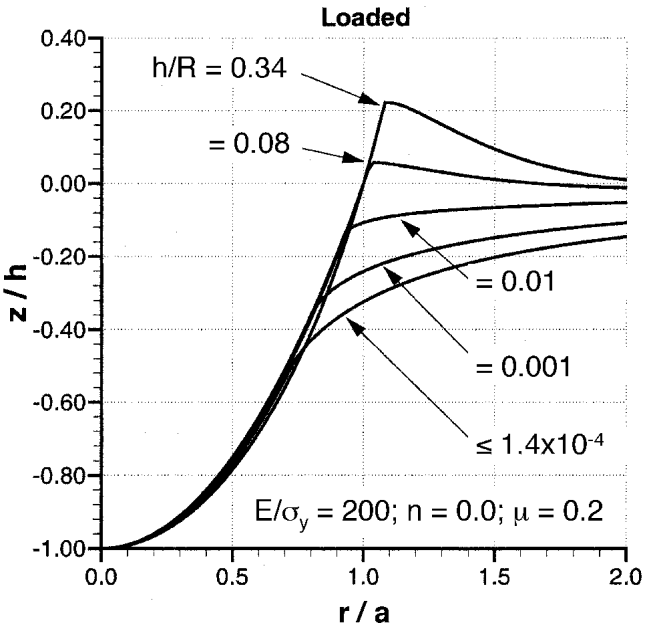

FIG. 8. Finite element simulation of normalized contact profiles under load for spherical indentation of an elastic–perfectly-plastic material with $E/\sigma_y = 200$ showing the influence of the normalized penetration depth, h/R, on the pile-up/sink-in behavior (after Ref. 34).

E. The correction factor β

The correction factor β appearing in Eq. (6) plays a very important role when accurate property measurements are desired. This constant affects not only the elastic modulus calculated from the contact stiffness by means of Eq. (6), but the hardness as well because procedures for determining the indenter area function are

also based on Eq. (6), and area functions can be in error if the wrong value of β is used (area function calibration methods are addressed in Sec. IV). Thus, without a good working knowledge of β, one is limited in the accuracy that can be achieved in the measurement of both H and E.

For the case of small deformation of an elastic material by a rigid axisymmetric punch of smooth profile, β is exactly 1 (note that small deformation is achieved only if the half-included angle of the indenter ϕ is close to 90°). However, because real indentation experiments are conducted with non-axisymmetric indenters and involve large strains, other values for β may be more appropriate. Here, we review the factors that affect β for Berkovich indentation and attempt to draw some conclusions concerning what value is best. The complex nature of elastic–plastic deformation with pyramidal indenters makes this a very difficult exercise with results that are not entirely satisfying.

The importance of correcting Eq. (6) was first recognized by King,[59] who used numerical methods to explore how noncircular geometries effect the elastic contact stiffness of materials tested with rigid, flat-ended punches. He found that that $\beta = 1.012$ for a square-based indenter and $\beta = 1.034$ for a triangular punch. The latter value has widely been adopted for instrumented indentation testing with a Berkovich indenter (a triangular pyramid). Vlassak and Nix later conducted independent numerical calculations for the flat-ended triangular punch using a more precise method and found a higher value, $\beta = 1.058$.[60] Hendricks, noting that the pressure distribution for elastic deformation by a flat punch is not representative of the real elastic–plastic problem, adopted another approach to estimate β.[61] Assuming the pressure profile is perfectly flat, he used simple elastic analysis procedures to show that $\beta = 1.0055$ for a Vickers indenter (square-based pyramid) and $\beta = 1.0226$ for the Berkovich.

Finite element simulations have also been used to evaluate β and explore the factors that influence it.[25,26,29,30,32] Finite element simulations have the advantage of including the effects of plasticity, but are subject to inaccuracies caused by inadequate meshing and convergence. Larson et al.[32] conducted full 3D finite element calculations of true Berkovich indentation for a purely elastic material and for four different elastic–plastic materials that simulate the behavior of aluminum alloys. For the purely elastic material, they found that β is slightly dependent on Poisson's ratio, ν, through the relation

$$\beta = 1.2304\ (1 - 0.21\nu - 0.01\nu^2 - 0.41\nu^3) \quad . \qquad (20)$$

Assuming $\nu = 0.3$, this gives $\beta = 1.14$. Note, however, that this solution applies to indentation of a flat, elastic half-space and thus inherently ignores the fact that plasticity severely distorts the surface during the formation of the hardness impression. Because of this, one might expect better results from the elastic–plastic simulations because they account for plasticity. For the four elastic–plastic materials examined, the value $\beta = 1.034$ suggested by King was found to work well, producing errors of no more than 6.5% in the contact area and hardness.

Recent studies have shown that even for perfectly elastic contact by a rigid cone, β may deviate from unity due to departures from the small strain approximation. Hay et al.[26] used both finite element simulation and analytical techniques to examine indentation of an elastic half-space by a rigid conical indenter. They showed that the derivation used to obtain Eq. (6) with $\beta = 1$ is not strictly valid for the conical indenter geometry because it improperly accounts for radial material displacements in the region of contact (note: Hay et al. originally designated this correction factor as γ). They developed an approximate correction for the effect that gives

$$\beta = \pi \frac{\left[\frac{\pi}{4} + 0.1548 \cot\phi \frac{(1-2\nu)}{4(1-\nu)}\right]}{\left[\frac{\pi}{2} - 0.8312 \cot\phi \frac{(1-2\nu)}{4(1-\nu)}\right]^2} \quad , \qquad (21)$$

where ϕ is the half included angle of the indenter. For $\nu = 0.3$ and $\phi = 70.3°$, this expression gives $\beta = 1.067$. Although these observations provide physical insight into another important reason why β should be greater than unity, it is not clear that values derived from Eq. (21) are particularly meaningful as they apply only to a flat elastic half-space. However, the results do point out that there may be a dependence of β on the cone angle, with smaller cone angles giving greater βs.

A thorough finite element study of elastic–plastic indentation with a 68° cone has been conducted by Cheng and Cheng.[29,30] Their calculations encompassed a wide range of materials characterized by different E/σ_y ratios, Poisson's ratios, ν, and work-hardening exponents, n. The results are reported in two separate publications.[29,30] The first, which focuses on nonhardening materials, finds that $\beta = 1.05$, independent of E/σ_y and ν.[30] In the second, both hardening and nonhardening materials are considered, and it is reported that $\beta = 1.085$, independent of E/σ_y and n.[29] Note that there is a slight inconsistency in the two papers in that the non-work-hardening results in the second paper give a different value of β from those in the first. Therefore, which of these results is better, if either, is not clear. We have conducted similar finite element simulations ourselves for a 70.3° cone[25] and find that $\beta \cong 1.07$ for most materials, but may rise to higher values in materials with large E/σ_y. Also, β could be slightly higher in materials that work harden.

The wide range of βs reported in these studies makes it difficult to settle in on a single preferred value. What can be concluded, however, is that there are valid reasons

to expect that β for the Berkovich indenter should be slightly greater than unity. The deviation from circular cross-section appears to play an important role, as does the fact that the original analysis from which Eq. (6) was derived ignores the radial displacements of the surface for a conical indenter. If we discard the values derived from elastic analyses of flat half-spaces, which do not consider the formation of the hardness impression, the remaining values fall in the range $1.0226 \leqslant \beta \leqslant 1.085$. Thus, $\beta = 1.05$ is probably as good a choice as any, with a potential for error of approximately ±0.05. Carefully performed experiments and 3D finite element simulations that take care to assure good convergence and accuracy could help to resolve this issue.

IV. REFINEMENTS TO THE METHOD

A. Difficulties with the original method

The ease of use of the method in a well-calibrated testing system has proven to be one of its great advantages. On the other hand, the method we originally proposed for calibrating the load frame compliance, C_f, and the area function of the indenter has proven to be cumbersome and sometimes difficult to implement. We have thus worked over the last several years to develop a simpler, more accurate calibration procedure.

System calibrations are generally based on the relation

$$C = C_f + \frac{\sqrt{\pi}}{2E_{eff}} \frac{1}{\sqrt{A}} \quad , \tag{22}$$

which simply states that the total measured compliance, C (i.e., the inverse of the measured stiffness), is the sum of the compliance of the load frame (the first term) and the contact (the second term); that is, the two act like springs in series. This assumes that the load frame compliance is a constant independent of load. If C_f is known, the area function can then be determined from measurements of compliance as a function of depth. Alternatively, if the area function is known, the load frame compliance can be determined as the intercept of a plot of C versus $A^{-1/2}$. To determine these quantities simultaneously, we originally suggested an iterative procedure in which simple assumptions are made about the starting values of C_f and the area function. Under certain circumstances, however, problems are encountered with this procedure.

At shallow depths, corresponding to small contact areas at the beginning of an indentation experiment, the contact compliance is high and dominates the total measured compliance (Eq. 22). However, as the contact depth (and area) increases, the contact compliance decreases, and at some point, the load frame compliance becomes the more dominant factor. Hence, for measurements at large depths, the load frame compliance must be known with great accuracy. The situation is reverse for the area function. The macroscopic shape of the indenter at large depths can be controlled very precisely during its manufacture. The uncertainty in the area function is due to difficulties in producing a perfectly sharp tip, which are most important at small depths.

If the load frame compliance is low and the tip is relatively sharp, there is an intermediate depth range in which uncertainties in C_f and the area function are inconsequential. In this case, the iterative process for determining the area function and load frame compliance converges after a few cycles, and our original calibration method works well, albeit through a fairly tedious calculation process. However, if the tip is not sharp or the load frame compliance is high, the iterative process converges very slowly and sometimes does not converge to a unique solution. Noise in the data compounds this problem. The method we have developed to circumvent these problems is detailed in Sec. IV. E.

B. The indenter area function

One aspect of the original method that has produced a great deal of confusion concerns the mathematical form of the area function. The form we originally proposed was

$$A = \sum_{n=0}^{8} C_n (h_c)^{2-n} = C_0 h^2 + C_1 h + C_2 h^{1/2} + C_3 h^{1/4} + \ldots + C_8 h^{1/128} \quad , \tag{23}$$

where $C_0 \ldots C_8$ are constants determined by curve-fitting procedures. We wish to make it perfectly clear here that this function was selected *strictly* for its ability to fit data over a wide range of depths and not because it has any physical significance. This being said, the equation is, in fact, quite convenient in describing a number of important indenter geometries. A perfect pyramid or cone is represented by the first term alone. The second term describes a paraboloid of revolution, which approximates to a sphere at small penetration depths, and a perfect sphere of radius R is described by the first two terms with $C_0 = -\pi$ and $C_1 = 2\pi R$.[57] The first two terms also describe a hyperboloid of revolution, a very reasonable shape for a tip-rounded cone or pyramid that approaches a fixed angle at large distances from the tip. One other form with physical significance is that suggested by Loubet et al.,[62] which describes a pyramid with a small flat region on its tip, the so-called tip defect. This geometry is described by the addition of a constant to the first two terms in Eq. (23). In each case, the experimentally determined constants can be compared with the appropriate geometric description to verify the geometry. The higher order terms in Eq. (23) are generally useful in describing

deviations from perfect geometry near the indenter tip and give the experimenter some flexibility in developing an area function that is accurate over several orders of magnitude in depth.

One other important aspect of the area function is the depth range over which it is to be used. There are a number of choices that can be made during the curve-fitting process that affect the applicable range of the function. If the function is to be applied at depths greater than the data used to construct it, one must be concerned about how the function extrapolates. By assigning the constant C_0 associated with the quadratic term to a value determined by the face angles of the pyramid and restricting the remaining constants to positive values, the resulting function will approach the description of the perfect pyramid at large depths and have no local maxima or minima. This forces the results to be reasonable at all depths. The alternative is to allow all of the parameters to be fit and weight the fitting process to the particular depth range of interest. This procedure often yields a function that fits the specified data better than the first approach, but such functions may be highly inaccurate outside the depth range used to construct them. This approach may be preferable when the function is to be used for repetitive measurements; for instance, in quality control applications. Each of these approaches is demonstrated in Sec. IV. E.

C. Load divided by stiffness squared, P/S^2

One of the most important improvements to the calibration procedures is based on an observation originally reported by Joslin and Oliver that the ratio of the load to the stiffness squared, P/S^2, is a directly measurable experimental parameter that is independent of the penetration depth or contact area provided the hardness and elastic modulus do not vary with depth.[63] In a manner similar to Eq. (19), this follows by combining Eqs. (5) and (6) to obtain

$$\frac{P}{S^2} = \frac{\pi}{(2\beta)^2}\frac{H}{E^2} \quad . \tag{24}$$

The utility of the parameter stems from its independence of the contact area. Joslin and Oliver took advantage of this to evaluate comparatively the mechanical properties of materials in which surface roughness led to uncertainties in the contact area. The area-independence of P/S^2 also makes it valuable in the determination of load frame compliances. Stone et al.[64] have developed one such procedure, and the procedure we describe shortly is also based on this principle. The basic idea is that since P/S^2 is not influenced by the area, the machine compliance can be determined in a manner that does require a priori knowledge of the area function. This decouples the measurement of load frame compliance from the area function and eliminates the need for an iterative procedure.

Another useful feature of P/S^2 is that it does not depend on the pile-up or sink-in behavior. Hence, if the modulus of the material is known, Eq. (24) can be used to calculate accurately the hardness, even when there is significant pile-up. Conversely, if the hardness is known (e.g., by direct measurement of the contact area), then the modulus can be determined. Unfortunately, this expression does not allow one to calculate the effect of pile-up on both the hardness and the modulus simultaneously, unless, perhaps, the method proposed by Cheng and Cheng is proven to be experimentally viable[31] (Sec. III. C).

Several other useful applications of P/S^2 as a characterizing parameter have been recognized in experimental studies. Saha and Nix have shown that Eq. (24) is accurate for even films on substrates provided the modulus of the film and substrate are similar.[20] With a known modulus, this facilitates the measurement of film hardness in a manner that fully accounts for pile-up, which can be quite large in many film–substrate systems.[14] In a similar manner, Page et al.[22] have shown that P/S^2 can directly be correlated to the tribological performance of coated systems.

D. Continuous stiffness measurement

Improvements to measurement and calibration procedures have also been facilitated by advances in testing techniques and apparatus. One of the most important of these is the "continuous stiffness measurement" technique (CSM), in which stiffness is measured continuously during the loading of the indenter by imposing a small dynamic oscillation on the force (or displacement) signal and measuring the amplitude and phase of the corresponding displacement (or force) signal by means of a frequency-specific amplifier.[1,24,65,66] As this technique has matured over the last 10 years, it has dramatically reduced our reliance on unloading curves and offers several distinct advantages. First, it has the clear advantage of providing continuous results as a function of depth. Second, the time required for calibration and testing procedures is dramatically reduced because there is no need for multiple indentations or unloadings. Third, at high frequencies, CSM allows one to avoid some of the complicating effects of time-dependent plasticity and thermal drift, which caused so much consternation in our original calibration method. Fourth, the CSM technique allows one to measure the effects of stiffness changes as well as damping changes at the point of initial contact. This facilitates an important new technique for precisely identifying the point of initial contact of the indenter with the sample.

Defining the point of initial contact is the critical starting point in the analysis of almost all instrumented indentation data. The resolution in the load and displacement signals as well as the data acquisition rates clearly affect how well one can determine the location of the surface. However, it is also important to be able to resolve and detect near-surface forces, as these often have an important bearing on the interpretation of the surface location. Even in the simple case of sharp Berkovich diamond contact with the surface of fused silica tested in laboratory air, there is at least one important surface force—water meniscus formation. Another important surface effect is adhesion between the tip and the sample. Because of these, techniques that use preloads or back-extrapolation of the data to determine the point of initial contact can be misleading. The approach we prefer is to observe the entire mechanical response of the system before, during, and after the point of initial contact and identify the point of contact by examining the overall observed behavior. As shown in the next section, information gleaned from CSM measurements has proven invaluable in this regard and has helped significantly in identifying the point of first surface contact.

E. Refined measurement and calibration procedures

The procedure we now use to calibrate the machine compliance and indenter area function for a Berkovich indenter is based on indentation of fused silica using continuous stiffness measurement. Fused silica is chosen for many reasons, one of the most important being that it does not pile-up because of its small E/H value. Here, we detail the procedure as we apply it to measurements with the Nano Indenter XP (MTS Systems Corp., Knoxville, TN). During continuous stiffness measurement with this instrument, several parameters can be controlled and measured. These include the harmonic load (the amplitude of the oscillation in the load signal), the harmonic displacement (the amplitude of the oscillation in the displacement signal), the harmonic frequency (the frequency of the oscillation), the phase angle (the angular phase shift between the load and displacement oscillations), and the harmonic stiffness (which is derived from the other quantities).[1,65,66] The Nanoindenter XP is inherently a force-controlled device, but methods have been developed to keep the harmonic displacement constant by means of feedback control.

The basic experiment begins with a slow approach to the surface to identify carefully the point of first contact based on measurements of load, displacement, and stiffness. The surface approach velocity is 10 nm/s, with a 3-nm oscillation in the harmonic displacement at a frequency of 80 Hz. Clearly, a small displacement oscillation is desirable to avoid effects on the total measured displacement, but larger oscillations improve the signal-to-noise ratio. A 3-nm oscillation was found to be an appropriate compromise for fused silica. Once the surface is detected, the indenter is loaded at a constant value of $(dP/dt)/P = 0.3$ (the loading rate divided by the load), which has the advantage of logarithmically scaling the data density so that there is just as much data at low loads as high. Constant $(dP/dt)/P$ tests also have the advantage of producing a constant indentation strain rate, $(dh/dt)/h$, provided the hardness is not a function of the depth.[42] For the first 400 nm of displacement, the harmonic displacement is kept constant by feedback control at 3 nm. However, when the displacement reaches 400 nm, the experiment is paused for 10 s during which the harmonic displacement is increased and stabilized at 8 nm to reduce the noise levels in subsequent measurements (Note: The interruption and increase in harmonic displacement has advantages in hard materials like fused silica, but may lead to measurement inaccuracies in soft materials like aluminum). The loading is then continued until the maximum load is achieved (approximately 630 mN), after which the indenter is unloaded at a constant loading rate $dP/dt = 19$ mN/s to 10% of the maximum load. After holding for 200 s to establish the thermal drift rate for correction of the displacement data using the technique described in the original method, the indenter is completely unloaded. During the experiment, the load on the sample, the displacement into the surface, the harmonic load, the harmonic displacement, and the phase angle are all measured at a data acquisition rate of 5 Hz. For the results reported here, the experiment was repeated 10 times for statistical averaging.

We also generated images of the residual hardness impressions used in the calibration experiments to compare the areas deduced from the indentation data to the actual contact areas. To do this, the indenter was raster scanned across the indented area at 2 μm/s at a small load $P = 5$ μN to produce a fully quantitative topographic image of the deformed surface with a resolution of 512 × 256 pixels. The scanning was accomplished by a monolithic, piezoelectrically driven nanopositioning stage with 100 × 100 μm of travel and feedback control based on capacitive displacement sensors. The resolution, accuracy, linearity, flatness of travel, and settling time for the stage are 2 nm, 2 nm, 10 nm, and 2 μs, respectively. The resulting images, shown in Figs. 4 and 9, do not suffer from the distortion typically associated with conventional scanning probe microscopy.

A typical load versus displacement curve for the calibration experiment is shown in Fig. 10. The curve, which has been corrected for thermal drift, exhibits parabolic loading and power-law unloading. Note that neither the pause during loading to increase the harmonic displacement amplitude nor the pause during unloading to measure the thermal drift are apparent in the curves, thus

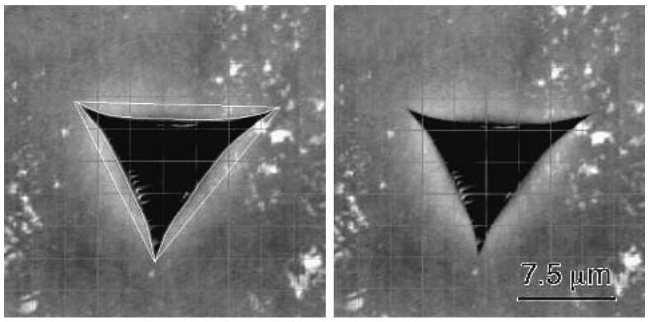

FIG. 9. An indentation in fused silica made by raster scanning with a nanopositioning stage. Data obtained from this indentation were used to calibrate the load frame compliance and area function. (a) The image on the left shows the borders defining the two contact areas discussed in the text. (b) The image on the right is included for clarity.

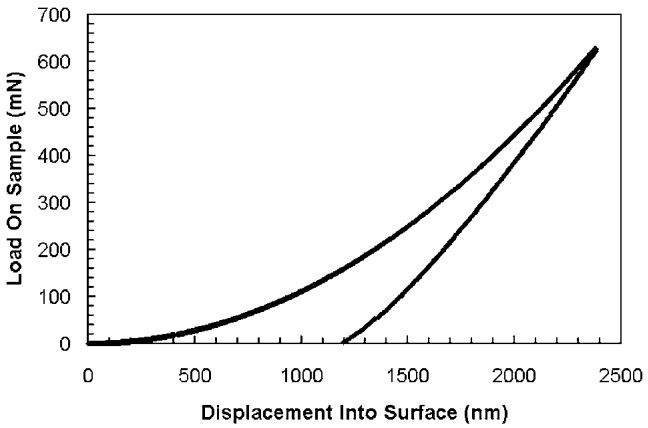

FIG. 10. Typical load versus displacement curve for a calibration experiment in fused silica.

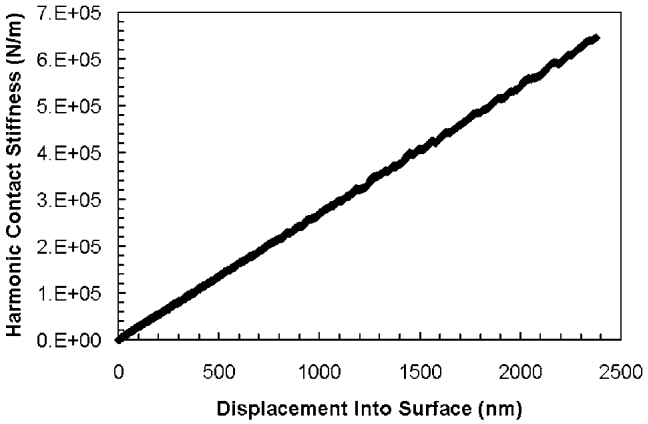

FIG. 11. A typical stiffness versus displacement curve for the calibration experiment. Stiffnesses were determined using the continuous stiffness measurement technique.

demonstrating that these interruptions do not influence the data. The stiffnesses measured by CSM during loading are shown in Fig. 11. Here again, the brief pause at 400 nm is not evident.

Figure 12 shows a plot of load versus displacement enlarged to examine the region of first contact. Figures 13 and 14 are the corresponding behavior for the harmonic displacement and phase angle, respectively. In all three cases, there is a clear change in behavior corresponding to contact with the surface. Note, however, that the change is smooth and continuous rather than abrupt, resulting in some uncertainty in defining the point of first contact. In contrast, the data in Fig. 15 shows that the harmonic stiffness rises rapidly when contact is achieved, thereby providing the best estimate of the surface location. The effects of a very thin adsorbed layer (most likely water) are also apparent; meniscus formation results in a measurable reduction in stiffness just prior to contact. What actually constitutes the point of first contact is debatable when such effects occur; however, it can certainly be identified to within ±2 nm.

Although the harmonic stiffness is the preferred locator of the surface in this particular situation, it should be noted that which of the measurements in Figs. 12–15 provides the best surface definition depends on the properties of the material *and* the characteristics of the testing system. The resonant frequency of the testing system is particularly important (12 Hz for the system used here). As the stiffness of the contact increases, the resonant frequency of the combined instrument and contact sweeps toward higher values. The exact nature of the changes observed in the dynamic measurements in Figs. 13–15 then depends on whether the system moves toward resonance (as in our experiments) or away from it as contact is made. In fact, it is because the system moves toward and through resonance that the harmonic displacement in Fig. 13 actually increases rather than decreasing when contact is made. Often, the most sensitive frequency for surface detection is the resonant frequency of the testing system just prior to contact; however, this is generally also the frequency at which the system is most vulnerable to noise (vibration, and so forth).

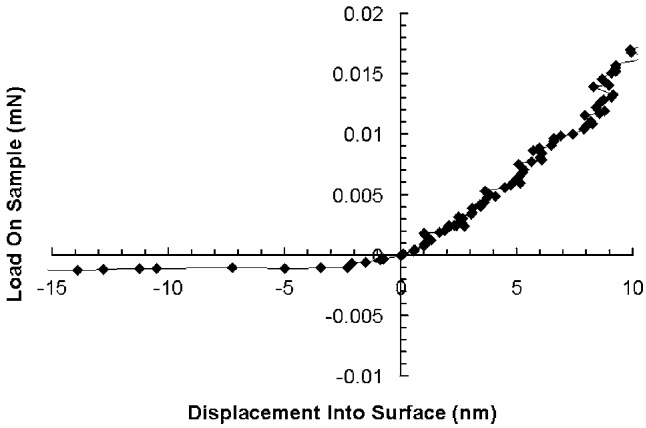

FIG. 12. An expanded view of the load versus displacement curve at the point of contact.

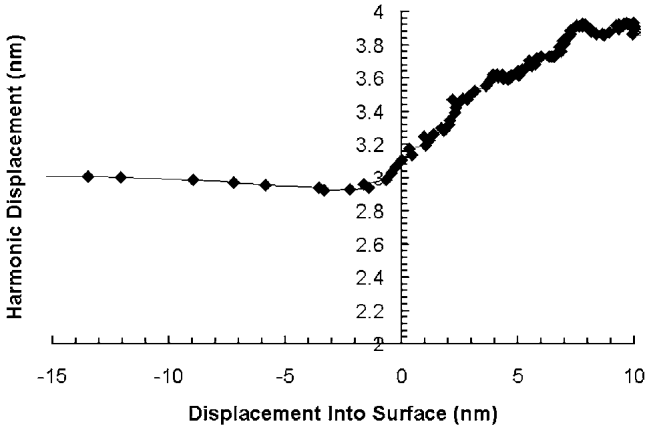

FIG. 13. An expanded view of the harmonic displacement amplitude versus displacement curve at the point of contact.

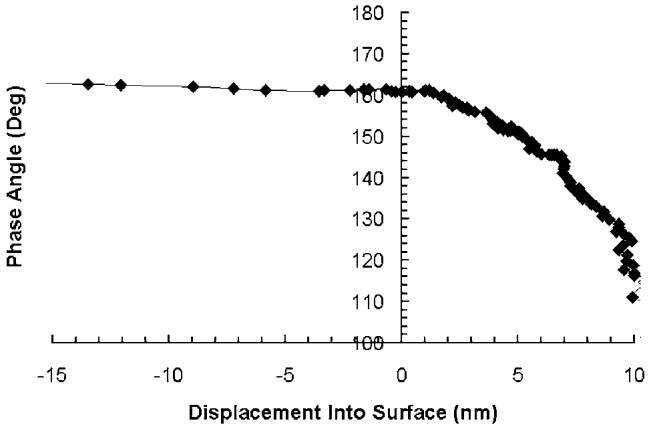

FIG. 14. An expanded view of the phase angle versus displacement curve at the point of contact.

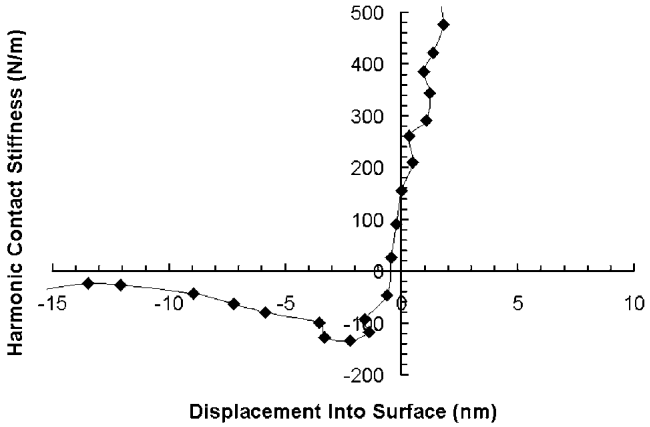

FIG. 15. An expanded view of the harmonic stiffness versus displacement curve at the point of contact. Note the minimum observed prior to full contact and the steep rise in stiffness at the point of contact.

Once the point of contact is determined, the data can directly be analyzed to determine the load frame compliance. The procedure we now prefer is based on a plot of P/S^2 versus indenter displacement like that shown in Fig. 16. Note that S in this plot represents the stiffness of the contact. To determine S, the component due to the load frame must first be removed (Eq. 22). The basic principle for the load frame compliance measurement is that at depths greater than a few hundred nanometers where the hardness and modulus of fused silica should be independent of depth, P/S^2 should also be constant according to Eq. (24). The proper load frame compliance is then found by changing the value of C_f until one is found that produces a flat P/S^2 versus h curve at large depths. It is not necessary to know the area function to implement this procedure, so an iterative process is not required.

The data at shallow depths must be ignored in this procedure because there can be, and usually is, a real variation in the measured hardness due to the tip imperfection. Tip blunting results in lower mean pressures because the contact behaves more like a spherical indenter on a flat surface, for which the mean contact pressure falls to zero at zero depth. This is a real change in the mean pressure and not an artifact of an inaccurate area function. As the load increases from zero, the contact will at first be elastic; however, at some depth plasticity begins, and at a still larger depth, full plasticity is established. At very large depths, geometric similarity of the Berkovich pyramid together with the homogeneous properties of fused silica assure that the mean pressure (or hardness) remains essentially constant as the depth increases. These effects are evident in Fig. 16 as the rise in P/S^2 over the approximately first 500 nm of displacement. The extent of the effect depends on the condition of the indenter. For severely worn tips, the region of nonconstant H and P/S^2 can extend to much greater depths.

In Fig. 17, we demonstrate the load frame compliance measurement procedure for three different assumed values of C_f. The correct value, $C_f = 9.62 \times 10^{-8}$ m/N, results in a flat curve for the data between 1000 nm and 2300 nm where the tip rounding is unimportant. However, increasing or decreasing C_f by as little as 20%

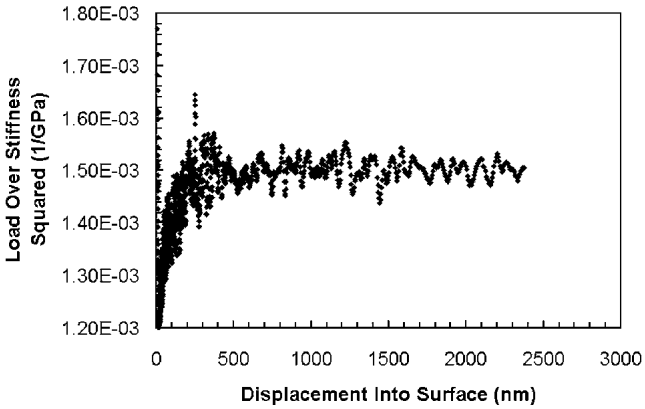

FIG. 16. A typical load over stiffness squared versus displacement for the calibration experiment. Note the constant value beyond 1000 nm.

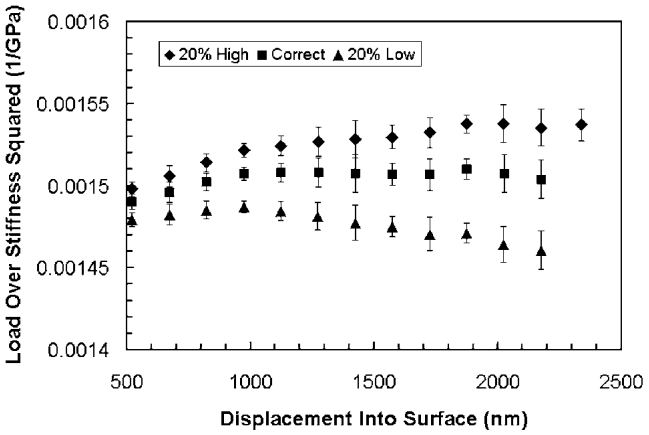

FIG. 17 Load over stiffness squared versus displacement for three different values for the load frame stiffness. The curve with the correct load frame stiffness is flat at large depths.

results in a rising or falling curve. From experience, we have found that the proper value of P/S^2 for fused silica is 0.0015 ± 0.0001 GPa^{-1}. This parameter is quite sensitive to errors in the calibration of load, displacement, and load frame compliance. Hence, the state of the system calibration can quickly be assessed by checking this parameter.

The value of load frame compliance measured in this way is applicable to any indentation experiment for which the contact stiffness does not significantly exceed the largest stiffness used in the measurements. Note, however, that the region of applicability cannot be judged through the depth of the indentation or the load on the indenter; it must be assessed using the stiffness of the contact.

With the load frame compliance established and the measured stiffnesses corrected to remove the load frame contribution from the displacements, the load–displacement–stiffness data can be used to deduce the indenter area function in a fairly straightforward manner. The procedure begins using Eq. (3) with $\epsilon = 0.75$ to determine the contact depths corresponding to each data point on the loading curve. The corresponding contact areas can be generated in two different ways. If the face angles of the indenter have precisely been measured, as we now do routinely, then the constant C_0 in the lead term in the area function can be estimated from the perfect pyramidal geometry. Provided the indenter is relatively sharp, the first term will dominate the area function at large depths and can be used to determine the large-depth contact area. By means of Eq. (6), the quantity βE_{eff} at the largest depth can then be computed from the stiffness and contact area using

$$\beta E_{\text{eff}} = \frac{\sqrt{\pi}}{2}\frac{S}{\sqrt{A}} \quad , \tag{25}$$

which in turn can be used to generate the contact areas at all depths from the measured stiffnesses according to

$$A = \frac{\pi}{4}\frac{S^2}{(\beta E_{\text{eff}})^2} \quad . \tag{26}$$

The area function is then determined by fitting the A versus h_c data to the form of Eq. (23).

Alternatively, if the indenter is not sharp or the face angles have not been measured, an assumed value for β and E_{eff} for fused silica can be used in Eq. (26) to generate the areas. For fused silica, we normally use $E = 72.0$ GPa and $\nu = 0.17$, which in conjunction with the elastic constants for diamond ($E = 1141$ GPa; $\nu = 0.07$) gives $E_{\text{eff}} = 69.6$ GPa. For β, we most often use $\beta = 1.034$, although based on the discussion in Sec. III. E, we may now reconsider this choice. A spherical indenter could also be calibrated in this way.

For the experimental data presented here (Figs. 10 and 11), the first approach was taken to determine the area function. A back-reflection laser goniometer was used to measure the face angles to within 0.025°, giving $C_0 = 24.212$. In precision work, this must be modified if the indenter is not aligned with its axis perfectly perpendicular to the specimen surface, as such misalignments increase the area at a given contact depth. The degree of tilt can be assessed from measurements of the sides of the hardness impression if a good image is available. A simple but effective method is based on the ratio of the long to short side of the hardness impression. Figure 18 shows how the contact area at a given depth depends on this ratio when the indenter is tilted slightly toward an edge or a face. Other tilt directions lie in-between the two

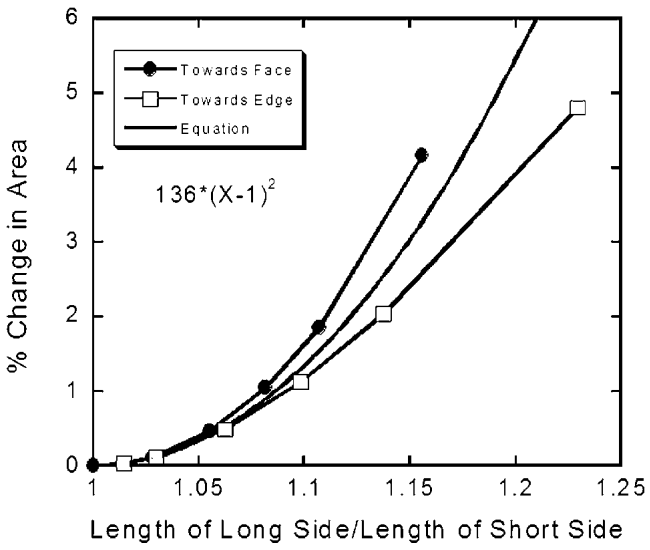

FIG. 18. The change in the projected area of contact for a tilted Berkovich indenter versus the ratio of the lengths of the longest to the shortest side.

curves, and the equation shown in the figure is an approximation based on the average. From the image in Fig. 9, the ratio of the lengths of the sides is 1.14, resulting in a 2% increase in the contact area and a corrected lead term of $C_0 = 24.65$. With this value, the area calculated at the largest depth is 67.0 ± 0.3 μm^2, based on an average of 10 measurements.

It is instructive to compare the calculated areas at the peak depth with those measured from images of the residual contact impressions. In Fig. 9(a), the inner white line traced around the hardness impression encloses exactly the predicted area of 67.0 μm^2. Although there is some ambiguity as to where the true edge of the contact is, it is clear that the predicted area is very close to that in the image. The outer white line in Fig. 9(a) is the triangle defined by the corners of the hardness impression. The enclosed corner-to-corner area is considerably larger, about 97.5 μm^2, indicating that one must be careful to account properly for the bowing in or out of the contact edges in the measurement of contact areas.

To complete the area function calibration procedure, the other coefficients in Eq. (23) must be determined by curve-fitting the A versus h_c data. If the area function is intended for use over a wide range of depth (e.g., 10–1500 nm), it is useful to fix the lead term C_0 using the aforementioned procedures based on the face angles of the indenter and restrict the rest of the coefficients to positive values. Fixing the lead term in this way assures that the area function accurately extrapolates at depths larger than those used in the calibration, and forcing the other terms positive assures a smooth area function. For the experimental data in Figs. 10 and 11, the coefficients for the area function derived in this way are $C_0 = 24.65$, $C_1 = 202.7$, $C_2 = 0.03363$, $C_3 = 0.9318$, $C_4 = 0.02827$, $C_5 = 0.03716$, $C_6 = 1.763$, $C_7 = 0.04102$, and $C_8 = 1.881$. The errors resulting from the fit are shown in Fig. 19, where the fractional difference between the experimentally measured area and that predicted by the area function are plotted as a function of the contact depth. For depths greater than 200 nm, the area function predictions are well within 1% of the correct values. At smaller depths, the error increases to as much as 4%. This can be improved using the second approach for finding the area function, limiting the range over which the function is fit, and placing no restrictions on the coefficients used in the fitting procedures. Letting the depth range be 10–500 nm, the area function derived in this way assuming $E_{eff} = 69.6$ GPa has coefficients $C_0 = 24.261849693995$, $C_1 = 388.715478479561$, $C_2 = -937.723180561482$, $C_3 = 251.535343527613$, $C_4 = 451.330970778406$, $C_5 = 219.019554856779$, $C_6 = -157.740285820129$, $C_7 = -98.1240614964975$, and $C_8 = -72.6226884095761$. More significant digits must be included because the mix of positive and negative coefficients can produce significant round-off errors. Figure 20 shows that this function works well over the contact depth range 10–500 nm, but it should not be used at larger or smaller depths. In practice, one should carefully assess the experimental depth range of interest and tailor the area function around it.

Finally, the results obtained here give us an opportunity to comment on the value of β. In deriving the area function by the first method, a value of $\beta E_{eff} = 69.5 \pm 0.14$ GPa was obtained by means of Eq. (25). Assuming that E_{eff} for fused silica indented by diamond is 69.64 GPa, then $\beta = 0.998 \pm 0.002$. Hence, these experiments suggest that β is unity for all practical purposes, although the experimental uncertainty is about ±5%.

ACKNOWLEDGMENTS

Research sponsored by the Division of Materials Sciences and Engineering, United States Department of Energy, under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC, and by MTS Systems Corporation through a research grant to the University of Tennessee. The authors would like to express their sincere gratitude to Professor W.D. Nix for nearly 30 years of continuous encouragement, support, guidance, and wisdom.

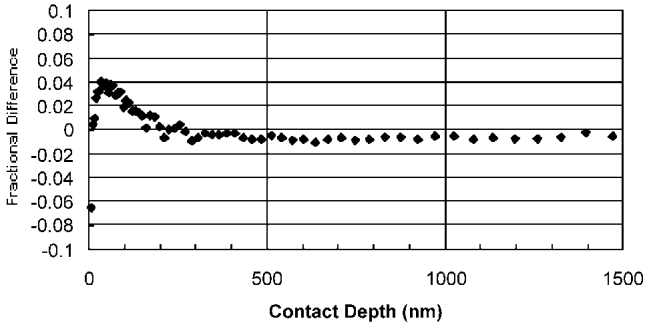

FIG. 19. Difference between the areas predicted by the area function and the data used to obtain the fit. This area function can be used at large depths.

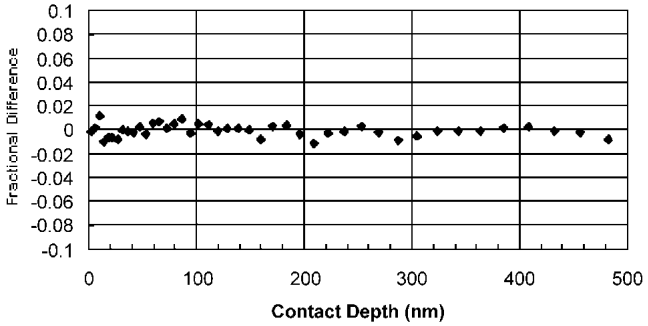

FIG. 20. Difference between the areas predicted by the area function and the data used to obtain the fit. This area function works well at small depths but not at large.

REFERENCES

1. W.C. Oliver and G.M. Pharr, J. Mater. Res. **7,** 1564 (1992).
2. G.M. Pharr, W.C. Oliver, and F.R. Brotzen, J. Mater. Res. **7,** 613 (1992).
3. J.B. Pethica, R. Hutchings., and W.C. Oliver, Philos. Mag. A **48**, 593 (1983).
4. F. Frolich, P. Grau, and W. Grellmann, Phys. Status Solidi **42,** 79 (1977).
5. D. Newey, M.A. Wilkins, and H.M. Pollock, J. Phys. E: Sci. Instrum. **15,** 119 (1982).
6. G.M. Pharr and W.C. Oliver, MRS Bull. **17,** 28 (1992).
7. G.M. Pharr, Mater. Sci. Eng. A **253,** 151 (1998).
8. T.D. Shen, C.C. Koch, T.Y. Tsui, and G.M. Pharr, J. Mater. Res. **10,** 2892 (1995).
9. D. Stone, W.R. LaFontaine, P. Alexopoulos, T.W. Wu, and C-Y. Li, J. Mater. Res. **3,** 141 (1988).
10. W.D. Nix, Metall. Trans. **20A,** 2217 (1989).
11. S.V. Hainsworth, H.W. Chandler, and T.F. Page, J. Mater. Res. **11,** 1987 (1996).
12. M. Tabbal, P. Merel, M. Chaker, M.A. El Khakani, E.G. Herbert, B.N. Lucas, and M.E. O'Hern, J. Appl. Phys. **85,** 3860 (1999).
13. S.V. Hainsworth, M.R. McGurk, and T.F. Page, Surf. Coat. Technol. **102,** 97 (1998).
14. T.Y. Tsui and G.M. Pharr, J. Mater. Res. **14,** 292 (1999).
15. T.Y. Tsui, J.J. Vlassak, and W.D. Nix, J. Mater. Res. **14,** 2196 (1999).
16. S. Bec, A. Tonck, J-M. Georges, E. Georges, and J-L. Loubet, Philos. Mag. A **74,** 1061 (1996).
17. N.X. Randall, Philos. Mag. A **82,** 1883 (2002).
18. Y.Y. Lim and M.M. Chaudhri, J. Mater. Res. **14**, 2314 (1999).
19. K. Miyahara, N. Nagashima, and S. Matsuoka, Philos. Mag. A **82,** 2149 (2002).
20. R. Saha and W.D. Nix, Acta Mater. **50,** 23 (2002).
21. J.A. Knapp, D.M. Follstaedt, S.M. Myers, J.C. Barbour, and T.A. Friedman, J. Appl. Phys. **85,** 1460 (1999).
22. T.F. Page, G.M. Pharr, J.C. Hay, W.C. Oliver, B.N. Lucas, E. Herbert, and L. Riester, in *Fundamentals of Nanoindentation and Nanotribology,* edited by N.R. Moody, W.W. Gerberich, N. Burnham, and S.P. Baker (Mater. Res. Soc. Symp. Proc. **522,** Warrendale, PA, 1998), p. 53.
23. J. Mencik, D. Munz, E. Quandt, E.R. Weppelmann, and M.V. Swain, J. Mater. Res. **12,** 2475 (1997).
24. B.N. Lucas, W.C. Oliver, and J.E. Swindeman, in *Fundamentals of Nanoindentation and Nanotribology,* edited by N.R. Moody, W.W. Gerberich, N. Burnham, and S.P. Baker (Mater. Res. Soc. Symp. Proc. **522,** Warrendale, PA, 1998), p. 3.
25. A. Bolshakov and G.M. Pharr, J. Mater. Res. **13,** 1049 (1998).
26. J.C. Hay, A. Bolshakov, and G.M. Pharr, J. Mater. Res. **14,** 2296 (1999).
27. T.A. Laursen and J.C. Simo, J. Mater. Res. **7,** 618 (1992).
28. H. Gao, C-H. Chui, and J. Lee, Int. J. Solids Structures **29,** 2471 (1992).
29. Y-T. Cheng and C-M. Cheng, J. Appl. Phys. **84,** 1284 (1998).
30. Y-T. Cheng and C-M. Cheng, Int. J. Solids Structures **36,** 1231 (1999).
31. Y-T. Cheng and C-M. Cheng, Appl. Phys. Lett. **73,** 614 (1998).
32. P-L. Larsson, A.E. Giannakopoulos, E. Soderlund, D.J. Rowcliffe, and R. Vestergaard, Int. J. Solids Structures **33,** 221 (1996).
33. S.D.J. Mesarovic and N.A. Fleck, Proc. R. Soc. London A **455,** 2707 (1999).
34. B. Taljat, T. Zacharia, and G.M. Pharr, in *Fundamentals of Nanoindentation and Nanotribology,* edited by N.R. Moody, W.W. Gerberich, N. Burnham, and S.P. Baker (Mater. Res. Soc. Symp. Proc. **522,** Warrendale, PA, 1998), p. 33.
35. G.M. Pharr and A. Bolshakov, J. Mater. Res. **17,** 2660 (2002).
36. A. Bolshakov, W.C. Oliver, and G.M. Pharr, in *Thin Films: Stresses and Mechanical Properties V,* edited by S.P. Baker, C.A. Ross, P.H. Townsend, C.A. Volkert, and P. Børgesen (Mater. Res. Soc. Symp. Proc. **356,** Pittsburgh, PA, 1995), p. 675.
37. K.W. McElhaney, J.J. Vlassak, and W.D. Nix, J. Mater. Res. **13,** 1300 (1998).
38. J-L. Loubet, B.N. Lucas, and W.C. Oliver, *NIST Special Publication 896: International Workshop on Instrumented Indentation* (National Institute of Standards and Technology, San Diego, CA, 1995), pp. 31–34.
39. B.N. Lucas, C.T. Rosenmayer, and W.C. Oliver, in *Thin Films—Stresses and Mechanical Properties VII,* edited by R.C. Cammarata, M. Nastasi, E.P. Busso, and W.C. Oliver (Mater. Res. Soc. Symp. Proc. **505,** Warrebdale, PA, 1998), p. 97.
40. W.H. Poisl, W.C. Oliver, and B.D. Fabes, J. Mater. Res. **10,** 2024 (1995).
41. B.N. Lucas, W.C. Oliver, J-L. Loubet, and G.M. Pharr, in *Thin Films: Stresses and Mechanical Properties VI,* edited by W.W. Gerberich, H. Gao, J.E. Sundgren, and S.P. Baker (Mater. Res. Soc. Symp. Proc. **436,** Pittsburgh, PA, 1997), p. 233.
42. B.N. Lucas and W.C. Oliver, Metall. Mater. Trans. **30A,** 601 (1999).
43. M.F. Doerner and W.D. Nix, J. Mater. Res. **1,** 601 (1986).
44. I.N. Sneddon, Int. J. Eng. Sci. **3,** 47 (1965).
45. A.E.H. Love, Q.J. Math. **10,** 161 (1939).
46. A.E.H. Love, Philos. Trans. A **228,** 377 (1929).
47. J.W. Harding and I.N. Sneddon, Proc. Cambridge Phil. Soc. **41,** 16 (1945).
48. I.N. Sneddon, *Fourier Transforms* (McGraw-Hill, New York, 1951), pp. 450–467.
49. J. Woirgard and J-C. Dargenton, J. Mater. Res. **12,** 2455 (1997).
50. C-M. Cheng and Y-T. Cheng, Appl. Phys. Lett. **71,** 2623 (1997).
51. H. Gao and T-W. Wu, J. Mater. Res. **8,** 3229 (1993).
52. H. Hertz, *Miscellaneous Papers by H. Hertz* (Macmillan, London, 1896).
53. K.L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge, 1985).
54. J.S. Field and M.V. Swain, J. Mater. Res. **8,** 297 (1993).
55. J.S. Field and M.V. Swain, J. Mater. Res. **10,** 101 (1995).
56. J.L. Hay, W.C. Oliver, A. Bolshakov, and G.M. Pharr, in *Fundamentals of Nanoindentation and Nanotribology,* edited by N.R. Moody, W.W. Gerberich, N. Burnham, and S.P. Baker (Mater. Res. Soc. Symp. Proc. **522,** Warrendale, PA, 1998), p. 101.
57. J.L. Hay and G.M. Pharr, in *ASM Handbook Volume 8: Mechanical Testing and Evaluation,* 10th ed., edited by H. Kuhn and D. Medlin (ASM International, Materials Park, OH, 2000), pp. 232–243.
58. D. Tabor, *The Hardness of Metals* (Oxford University Press, London, 1951).
59. R.B. King, Int. J. Solids Struct. **23,** 1657 (1987).
60. J.J. Vlassak and W.D. Nix, J. Mech. Phys. Solids, **42,** 1223 (1994).
61. B.C. Hendrix, J. Mater. Res. **10**, 255 (1995).
62. J.L. Loubet, J.M. Georges, O. Marchesini, and G. Meille, J. Tribology **106,** 43 (1984).
63. D.L. Joslin and W.C. Oliver, J. Mater. Res. **5,** 123 (1990).
64. D.S. Stone, K.B. Yoder, and W.D. Sproul, J. Vac. Sci. Technol. A **9,** 2543 (1991).
65. J.B. Pethica and W.C. Oliver, Phys. Scr. **19,** 61 (1987).
66. 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,** Pittsburgh, PA, 1989), p. 13.