Understanding nanoindentation unloading curves

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

A. Bolshakov
Houston Technology Center, Baker Atlas/INTEQ, Houston, Texas

(Received 17 June 2002; accepted 24 July 2002)

Experiments have shown that nanoindentation unloading curves obtained with Berkovich triangular pyramidal indenters are usually well-described by the power-law relation $P = \alpha(h - h_f)^m$, where h_f is the final depth after complete unloading and α and m are material constants. However, the power-law exponent is not fixed at an integral value, as would be the case for elastic contact by a conical indenter ($m = 2$) or a flat circular punch ($m = 1$), but varies from material to material in the range $m = 1.2$–1.6. A simple model is developed based on observations from finite element simulations of indentation of elastic–plastic materials by a rigid cone that provides a physical explanation for the behavior. The model, which is based on the concept of an indenter with an "effective shape" whose geometry is determined by the shape of the plastic hardness impression formed during indentation, provides a means by which the material constants in the power law relation can be related to more fundamental material properties such as the elastic modulus and hardness. Simple arguments are presented from which the effective indenter shape can be derived from the pressure distribution under the indenter.

I. INTRODUCTION

Load and depth-sensing indentation, also referred to as nanoindentation, has been developed over the past two decades as a technologically important tool for measuring the mechanical properties of materials, especially at small scales.[1–7] The technique relies on high-resolution instruments that continuously monitor the loads and displacements of an indenter as it is pushed into and withdrawn from a material. The load–displacement data obtained during one or more cycles of loading and unloading can be analyzed to derive a variety of mechanical properties, most commonly, the hardness and elastic modulus.[3,4] Such analyses are frequently based on solutions to the problem of indentation of an elastic half-space by a rigid, axially symmetric punch.[8,9]

Accurate measurement of mechanical properties by nanoindentation methods requires a detailed understanding of the information contained in the indentation loading and unloading curves.[4,9] Obtaining such an understanding is not an easy task due to the complex elastic and plastic deformation processes that occur during indentation, as well as the nonuniformity of the stress and deformation fields in the vicinity of the contact. For this reason, many methods for measuring properties by nanoindentation rely heavily on empirical observations that do not have solid theoretical underpinnings.[3,4,7]

In this work, a conceptual framework is developed to explain the experimentally observed mathematical form of nanoindentation unloading curves obtained with sharp, geometrically self-similar indenters like pyramids and cones. A typical set of nanoindentation data is presented in Fig. 1. The data were obtained by indenting fused silica with a Berkovich diamond indenter, a triangular pyramid with an area-to-depth relationship identical to that of the four-sided Vickers pyramid used commonly in microhardness testing. Because of the geometric self-similarity of pyramidal indenters, it is often convenient to model their behavior by that of a cone with a half-included angle that gives the same area-to-depth relationship. For the Berkovich and Vickers pyramids, the equivalent cone angle is 70.3°, and the area-to-depth relationship, also known as the area function, is given by

$$A = 24.56d^2 \quad , \tag{1}$$

where A is the cross-sectional area of the indenter at a distance d back from its tip.

Experiments have shown that indentation loading curves obtained with Berkovich indenters are usually well-described by the relation

$$P = \beta h^2 \quad , \tag{2}$$

a)Address all correspondence to this author.
e-mail: pharr@utk.edu

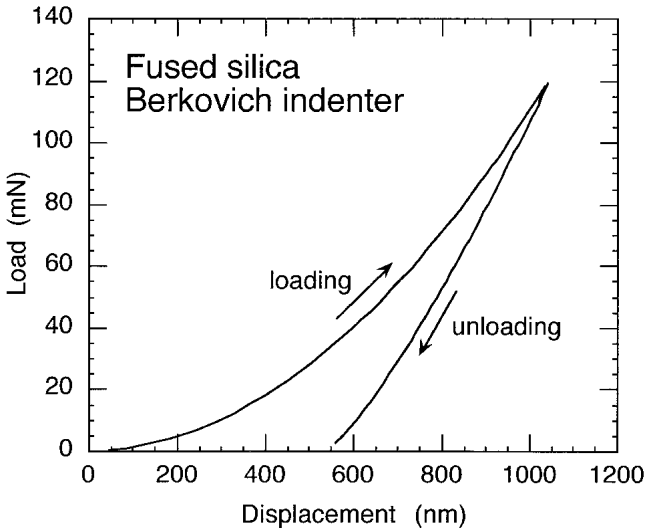

FIG. 1. Nanoindentation load–displacement data for fused silica obtained with a Berkovich indenter.[4]

where P is the indentation load, h is the measured depth, and β is a material constant related to the elastic and plastic properties of the material.[4] The simple quadratic form of this relation is a natural consequence of the geometric self-similarity of the indenter; since the load scales with the projected contact area and the indenter area scales with the square of the depth, the simple P–h relation of Eq. (2) results. A recent paper by Hainsworth *et al.* provides a rationale for this relation and relates the constant β to the elastic modulus and hardness of the material.[10] Deviations from the behavior of Eq. (2) are sometimes observed at small depths due to rounding of the indenter tip, which destroys the geometric similarity,[5,7] or to the indentation size effect.[11–14] However, such deviations are significant only at small depths (typically less than a micrometer for tip rounding and of the order of a micrometer for the indentation size effect). For larger depths, the quadratic form of Eq. (2) is generally well-obeyed.

The form of the unloading curve is quite different. Experiments conducted on a wide variety of materials including fused silica, soda-lime glass, and single crystals of aluminum, tungsten, and sapphire have revealed that unloading data are usually well-described by the power-law relation:

$$P = \alpha(h - h_f)^m \quad , \tag{3}$$

where h_f is the final displacement after complete unloading and α and m are material constants.[4] Typical values of α and m determined from regression analyses of experimental data are included in Table I, along with correlation coefficients for the curve fits.[4] The table shows not only that Eq. (3) provides a good description of the unloading data (correlation coefficients $R > 0.9999$) but also that the power-law exponents, m, while slightly material dependent, generally fall in the range 1.2–1.6. These exponents are consistent with neither the flat punch indenter geometry, for which $m = 1$,[8] nor the conical geometry, for which $m = 2$.[8] The flat punch geometry is important because it is sometimes invoked to model unloading behavior based on the assumption that the contact area remains constant, at least during the initial stages of unloading.[3] The conical geometry is important because it most closely approximates the self-similar geometry of the Berkovich indenter used to obtain the data. Curiously, an indenter with the geometry of a parabola of revolution, for which $m = 1.5$,[8] comes closest to predicting the experimental observations. However, exactly how a Berkovich indenter approximates to a parabola of revolution, which is neither sharp at its tip nor properly describes the relationship between the depth and the contact area, is not evident.

TABLE I. Values of parameters characterizing unloading curves as observed in nanoindentation experiments with a Berkovich indenter. Data are from Ref. 4.

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

In this paper, results of finite element simulations are presented that reproduce the experimentally observed unloading behavior and from which an understanding of the origin of the power law exponents for the unloading curves is obtained. The understanding is based on the concept of an "effective indenter," the geometry of which is determined by the shape of the plastic hardness impression formed during loading. Simple arguments are presented from which the shape of the effective indenter can be derived on the basis of the way that plasticity influences the pressure distribution under the indenter.

II. FINITE ELEMENT SIMULATION

Elastic–plastic indentation was simulated using the axisymmetric capabilities of the ABAQUS (Hibbitt, Karlsson, & Sorensen, Inc., Pawtucket, RI) finite element code. As detailed elsewhere,[15,16] the indenter was modeled as a rigid cone with a half-included angle of 70.3°, the angle that gives the same area-to-depth ratio as the Berkovich triangular pyramid used commonly in nanoindentation experiments. The specimen was modeled as a large cylinder 100,000 nm in height and 100,000 nm in radius represented by approximately 3000 four-node axisymmetric elements. Most simulations were performed to a depth of 500 nm, for which the specimen dimensions were found to be large enough to approximate the

behavior of a semi-infinite half-space as evidenced by an insensitivity of results to increases in specimen size. Roller boundary conditions were applied along the centerline and bottom (no interfacial friction or displacements normal to the boundary), a free surface was modeled at the outside of the cylinder, and the interface between the indenter and the specimen was assumed to be frictionless.

In the region of contact, very fine elements 20 nm in width were used to achieve accurate determination of the contact profiles. Away from the contact, a progressively coarser mesh was employed both at the surface and in the interior of the specimen. Plasticity was modeled using the large strain formulation of ABAQUS with the material behaving as a von Mises solid. The output of the finite element simulations included indentation load–displacement curves during one or more cycles of loading and unloading, the shapes of contact impressions at full load and after complete unloading, and the sizes and shapes of plastic zones.

Two different material constitutive behaviors were examined. The first was that of a real material—aluminum alloy 8009—whose stress–strain characteristics in uniaxial tension as well as its nanoindentation behavior have been well-documented in experiment.[17] The elastic–plastic constitutive behavior was described by piecewise fitting the results of a tension test. After yielding at a stress of 353.1 MPa, the material exhibits a small amount of work hardening before the flow stress saturates at 425.6 MPa. The strain at saturation is 2.44%. The fact that the flow stress reaches a constant level at such a small strain implies that the material behaves much like an elastic–perfectly-plastic solid with a flow stress of 425.6 MPa. The elastic constants used in the simulations were Young's modulus $E = 82.12$ GPa and Poisson's ratio, $\nu = 0.31$. As documented elsewhere, the simulated indentation load–displacement curves were found to agree well with real experimental data, despite the fact that the simulations were conducted for a conical indenter while the experiments were performed with a Berkovich pyramid.[15] This implies that the edges of the pyramid do not significantly influence the load–displacement behavior of the material. Simulation results for aluminum alloy 8009 were used to prove and develop the concept of an effective indenter based on the behavior of a real material.

The second constitutive behavior modeled in the study was that of an elastic–perfectly-plastic solid. For these simulations, the yield strength, σ_y, was varied systematically from 0.114 GPa to 5.32 GPa, while the modulus and Poisson's ratio were held constant at $E = 70$ GPa and $\nu = 0.25$. This produced modulus-to-yield strength ratios in the range $13.2 \leqslant E/\sigma_y \leqslant 614$, corresponding to modulus-to-hardness ratios in the range $7.13 \leqslant E/H \leqslant 237$ (the E/H values were computed using hardnesses deduced from the finite element results). These simulations were used to examine the behavior of a wide variety of materials, albeit materials that do not work harden. The E/H range spans a wide variety of metals and ceramics.[16]

III. SIMULATION RESULTS: ALUMINUM ALLOY 8009

The unloading curve generated by elastic–plastic finite element simulation of aluminum alloy 8009 is shown as the solid line in Fig. 2. Of the 500 nm of total displacement at peak load, only 35 nm is recovered during unloading, giving an unloaded final depth $h_f = 465$ nm. Note that the unloading data are distinctly curved, implying that unloading behavior cannot be adequately described by indentation with a flat punch, for which the behavior would be linear.

An important issue in understanding the form of the unloading curve is the extent to which it is elastic, that is, whether any of the recovered displacement is due to reverse plasticity. To explore this, a simulation was undertaken in which the indenter was completely loaded and unloaded five times while plasticity in the vicinity of the indentation and the evolution of the plastic zone were closely monitored. Figure 3, which shows the plastic zone both at full load and after the first unloading, demonstrates that the initial unloading results in a small extension of the plastic zone at the contact edge near the surface. Thus, some plasticity does indeed occur during unloading, even though the simple constitutive relation used in the simulation does not account for kinematic hardening. In subsequent cycles, further changes in the shape and size of the plastic zone also occur but to progressively smaller extents. However, the unloading plasticity has essentially no influence on the load–displacement behavior. This may be seen in Fig. 2, which

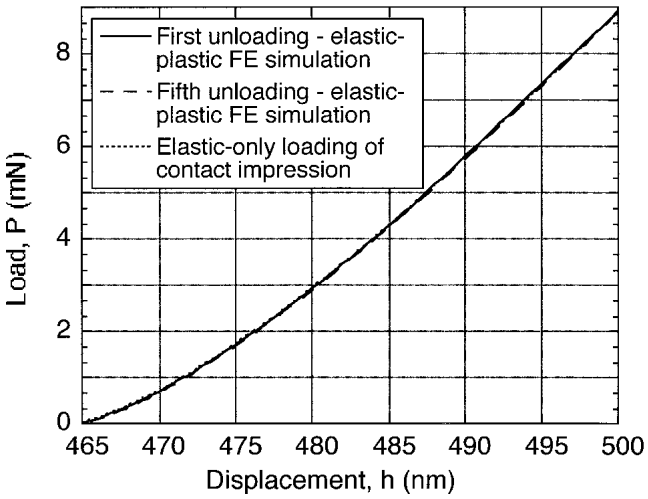

FIG. 2. Unloading curves for aluminum alloy 8009 generated by elastic–plastic finite element simulation.

includes the first and fifth unloading curves from the elastic–plastic simulation, as well as the results of a simulation in which the contact impression generated during the first loading cycle was reloaded using an elastic-only constitutive law. The fact that all three curves are virtually indistinguishable indicates that plastic deformation has a negligible influence on the load–displacement behavior after the contact impression first forms. Another significant observation is that all of the unloading curves are well described by the power-law relation of Eq. (3) with a power law exponent $m = 1.32$ (correlation coefficient $R = 0.99941$). This exponent, which is in excellent agreement with the value $m = 1.36$ measured in experiments with a Berkovich indenter,[17] corroborates the experimental observations that there is something fundamental about a power-law exponent in the range 1.2–1.6.

IV. CONCEPT OF THE EFFECTIVE INDENTER SHAPE

An explanation for why indentation unloading curves behave as they do can be achieved by careful examination of the shape of the contact impression after unloading. As shown schematically in Fig. 4, when a conical indenter is first pressed into a material, both elastic and plastic processes occur, with the contact impression conforming perfectly to the shape of the cone [Fig. 4(a)]. However, during the first unloading, elastic recovery causes the shape of the contact impression to change. Careful examination of the finite element results for alloy 8009 showed that the unloaded impression is not exactly conical in shape; rather, there is a small but important curvature to its surface. For alloy 8009 and similar materials with relatively large *E*/*H* ratios, e.g., soft metals, the curvature is virtually imperceptible in a scaled profile of the contact, giving one the impression that the unloaded contact impression is also conical [the curvature in Fig. 4(b) is grossly exaggerated for the sake of illustration]. The curvature becomes more pronounced and easier to observe in materials with smaller *E*/*H* ratios such as hard metals, ceramics, and glasses.

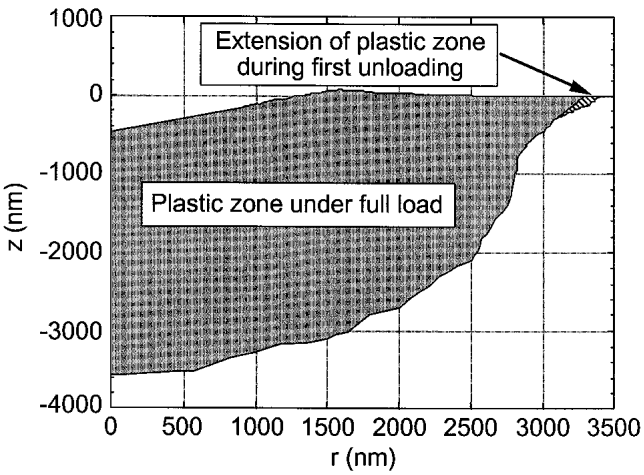

FIG. 3. Contact profiles and plastic zones observed in finite element simulation of aluminum alloy 8009.

The importance of the curvature follows by considering what happens when the contact impression is reloaded. From the geometries in Figs. 4(b) and 4(c), it is apparent that elastic reloading must occur with continuous increases in contact area until full contact is achieved at maximum load. However, because the unloading and reloading processes are elastic, what happens during reloading must be exactly the reverse of what happens during unloading. Thus, the unloading process must be characterized by a continuous decrease in contact area as the contact impression "peels away" from the indenter. Moreover, since the elastic contact stiffness, *S,* is related to the contact area through

$$S = \frac{2}{\sqrt{\pi}} E_{\text{eff}} \sqrt{A} \quad , \tag{4}$$

where E_{eff} is the effective modulus given by

$$\frac{1}{E_{\text{eff}}} = \frac{1 - \nu^2}{E} + \frac{1 - \nu_1{}^2}{E_{\text{i}}} \quad , \tag{5}$$

(E and ν are Young's modulus and Poisson's ratio for the specimen; E_{i} and ν_{i} are the same quantities for the indenter),[9] the slope of the unloading curve must

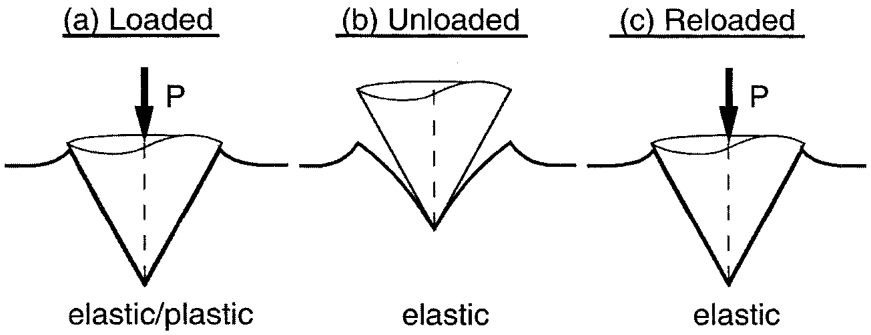

FIG. 4. Schematic representation of deformation processes during loading and unloading of a conical indenter.

continuously decrease as the load is reduced, in agreement with both experimental and finite element observations (see Figs. 1 and 2).

The exact shape and mathematical form of the unloading curve can be understood by introducing the concept of an "effective indenter shape." As illustrated in Fig. 5, the concept accounts for the fact that the unloading process is not properly described by elastic contact between a rigid cone and a *flat* elastic half-space but rather by a half-space whose surface has been locally distorted by plastic deformation during the formation of the hardness impression. The basic idea is to account for the surface distortion by transforming the contact geometry into one for which elastic half-space solutions can be applied. As shown in Fig. 5, this is accomplished by changing the geometry of the indenter to an "effective shape" defined such that when pressed into a flat surface, it gives the same normal surface displacements that would be produced by the conical indenter pressed into the plastically deformed surface of the hardness impression. The problem of elastic indentation of a nonflat surface with complex shape by a geometrically simple indenter is thus replaced by the problem of indentation of a geometrically flat surface by an indenter with complex shape. An analysis developed by Sneddon for indentation of a flat elastic half-space by an axisymmetric indenter of general shape (taken to be the effective shape) can then be used to model the load–displacement behavior.[8] Note that the effective shape concept cannot be applied if reverse plasticity during unloading is large enough to cause the unloading curve to deviate significantly from purely elastic behavior. Such behavior has been observed, for example, in soda-lime silicate glass.[4]

As shown in Fig. 5, the shape of the effective indenter is described by a function $z = u(r)$, where $u(r)$ is the vertical distance between the cone and the unloaded, permanently deformed surface of the hardness impression and r is the radial distance from the center of the contact. Since all finite element simulations conducted in this work showed that contact is last broken at the tip of the indenter, the function is determined with the tip of the cone just touching the surface of the contact impression at its center. Provided the shape of the deformed surface is known, the function $u(r)$ can be determined and the effective indenter can be constructed. From an analytical standpoint, the effective indenter is expected to give a reasonable approximation of the load–displacement behavior when the surface of the plastic hardness impression is not greatly displaced from the initially flat surface, i.e., for indenters with relatively large included angles.

The profile of an unloaded hardness impression made to a depth of 500 nm in alloy 8009 by a rigid 70.3° cone as determined in finite element simulation was carefully analyzed to establish the function $z = u(r)$. The resulting effective indenter shape is shown in Fig. 6. The shape is terminated at a radius of 1576 nm corresponding to the position of the contact edge at full load. Three features of the effective indenter are worthy of special consideration. First, at the contact edge ($r = 1576$ nm), the vertical separation between the flat surface and the indenter is only 12 nm, thus indicating how subtle the curvature really is. On a scaled plot of the profile of the hardness impression, this curvature would be imperceptible, and the impression would appear to match perfectly the shape of the 70.3° cone. Second, in sharp contrast to the case of a cone indenting a flat half-space, the effective indenter has a smooth, rounded profile at its tip. The reason for this is that the slope of the deformed hardness impression at its center exactly matches the slope of the conical indenter; i.e., the two surfaces perfectly conform at the tip of the indenter. Thus, the plastic deformation that produces the hardness impression has the interesting effect of removing the elastic singularity at the tip of the effective indenter. Third, as shown as the solid line through the data in Fig. 6, the shape of the effective indenter can be conveniently approximated by the simple power-law relation

$$z = Br^n \quad , \tag{6}$$

where the best power-law fitting constants have the values $B = 4.34 \times 10^{-8}$ nm$^{-1.63}$ and $n = 2.63$. The effective indenter is thus not at all conical in shape but similar to

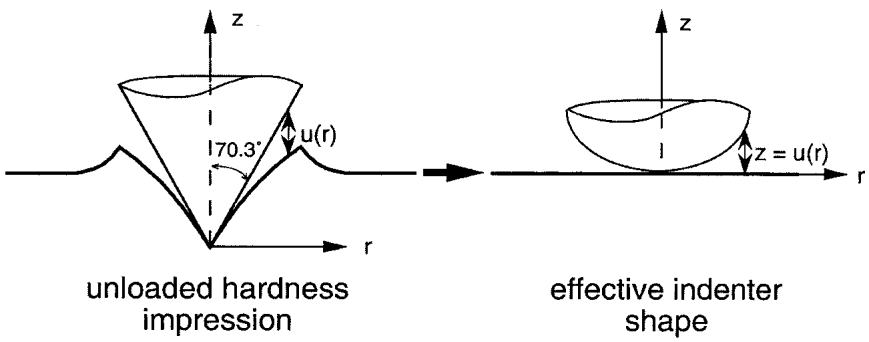

FIG. 5. Concepts used to define the effective indenter shape.

a *parabola of revolution* with a slightly higher power law exponent ($n = 2.63$ rather than $n = 2$).

That the elastic deformation of a flat surface by the effective indenter closely approximates the elastic deformation of the deformed hardness impression by the conical indenter is shown in Fig. 7, which compares load–displacement curves generated in several ways. The upper solid line is the first unloading curve determined in the full elastic–plastic finite element analysis for the 70.3° conical indenter (i.e., the same curve as in Fig. 2). This is to be compared to load–displacement curves for the effective indenter determined in three different ways. The first two were generated by elastic finite element simulation (no plasticity) of the deformation of a flat elastic half-space by the effective indenter; in one, the exact effective indenter shape was used, and in the other, the indenter was constructed from the power law approximation of the effective indenter shape [Eq. (6)]. The similarity of these two curves to that for the full elastic–plastic analysis shows that the effective shape concept does indeed closely model the unloading behavior. The third curve for the effective indenter was computed using an analytical approach based on Sneddon's analysis for the deformation of an elastic half-space by a rigid indenter of arbitrary axisymmetric shape.[8] For an indenter with shape described by the general power law of Eq. (6) characterized by parameters B and n, Sneddon's method yields

$$P = \frac{2E_{\text{eff}}}{(\sqrt{\pi}B)^{1/n}} \left(\frac{n}{n+1}\right) \left[\frac{\Gamma(n/2+1/2)}{\Gamma(n/2+1)}\right]^{1/n} h^{1+1/n} \quad , \tag{7}$$

where h is the elastic displacement, $E_{\text{eff}} = E/(1-\nu^2)$, and Γ is the factorial (gamma) function. The curve generated from this equation, though slightly lower than those determined in the finite element simulations, is still in reasonably good agreement. Recent work has shown that the analytical result of Eq. (7) slightly underestimates actual behavior because it does not consider influences of radial material displacement at the surface.[18]

One important feature of Eq. (7) is that it directly links the exponent n describing the effective indenter shape in Eq. (6) to the exponent m describing the shape of the unloading curve in Eq. (3); specifically,

$$m = 1 + 1/n \quad . \tag{8}$$

Since $n = 2.63$ for the effective indenter, Eq. (8) predicts that $m = 1.38$ in good agreement with the elastic/plastic finite element simulation ($m = 1.32$) and the results of actual nanoindentation experiments in alloy 8009 ($m = 1.36$).[17] Thus, the effective shape concept explains why the unloading behavior is more like deformation with a parabola of revolution ($m = 1.5$) than a cone ($m = 2$), at least in this material.

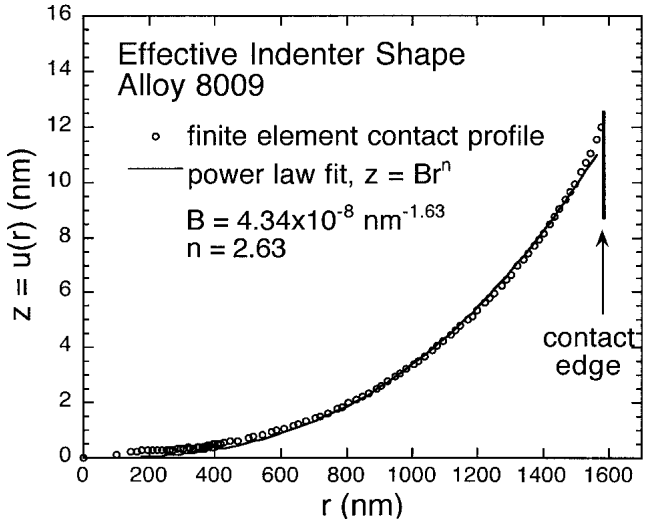

FIG. 6. Effective indenter shape for aluminum alloy 8009 indented by a 70.3° cone.

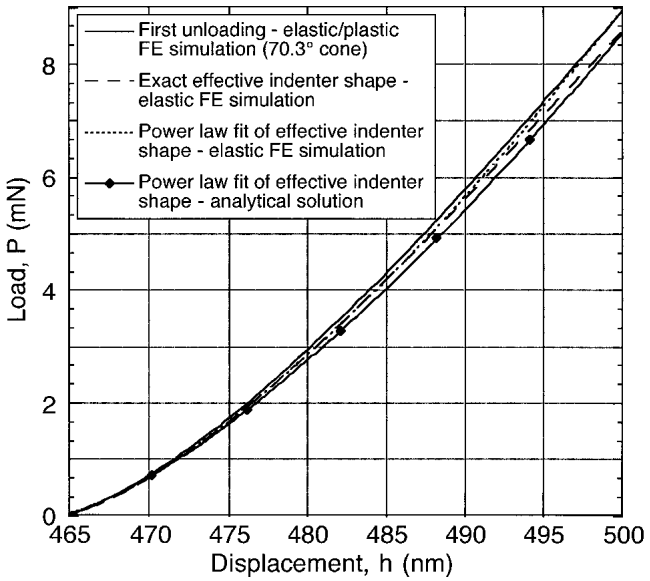

FIG. 7. Unloading curves for aluminum alloy 8009 computed in several different ways demonstrating that the effective indenter accurately models the behavior.

V. UNDERSTANDING THE EFFECTIVE INDENTER SHAPE

Further insight into the origin of the effective shape can be gleaned by considering the pressure distribution under the indenter, as illustrated by the simple thought processes in Fig. 8. During loading of a conical indenter into an elastic–plastic material [Fig. 8(a)], the distribution of pressure is determined by both elastic and plastic processes. The exact form of the pressure distribution is quite complex and not generally amenable to closed form

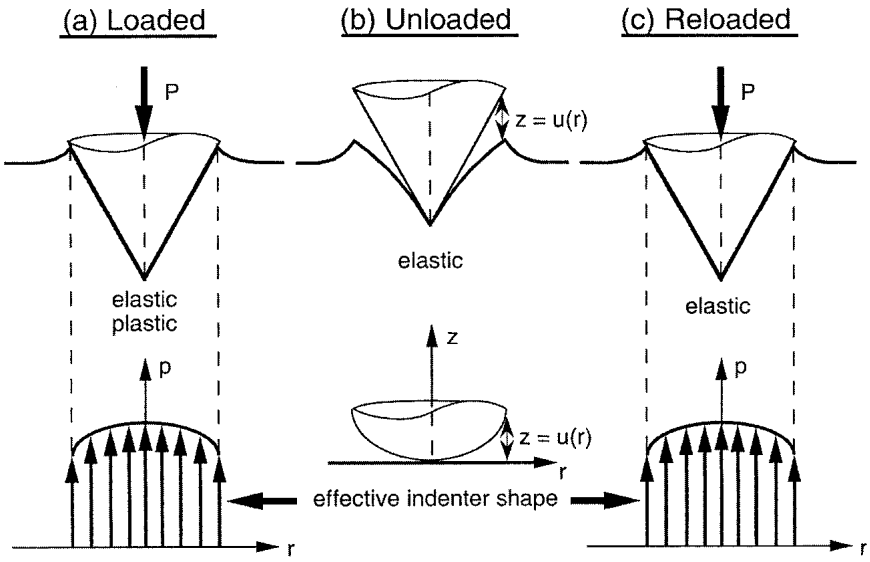

FIG. 8. Schematic representation of pressure distributions under the indenter and how they can be used to deduce the effective indenter shape.

analysis. As the pressure is reduced during unloading [Fig. 8(b)], the elastic component of the deformation recovers, but the plastic component does not, and the residual hardness impression changes its shape by elastic processes only. Since subsequent reloading of the hardness impression is also elastic [Fig. 8(c)], the pressure distribution after reloading must exactly match that produced by the elastic–plastic processes during the initial formation of the hardness impression. In other words, matching the pressure distribution at peak load serves to couple the elastic–plastic deformation during initial loading to the elastic processes during unloading. As shown in the lower portion of Fig. 8, the effective indenter shape is then found by determining the indenter geometry that produces this same pressure distribution by elastic deformation of a flat half space. Ideas similar to these have been used by Johnson[19,20] and Hirst and Howse[21] to deduce peak load pressure distributions from the shapes of residual contact profiles.

Figure 9 shows the peak load pressure distributions determined by finite element analysis for a wide variety of elastic–perfectly-plastic materials characterized by different E/σ_y ratios. Each of the indentations was made to a maximum depth, h_{max}, of 500 nm. The fact that the contact radius is different for each material results from the differing degrees of elastic and plastic deformation. Since in a typical nanoindentation experiment one does not necessarily know the value of E/σ_y, the figure also includes values of h_f/h_{max}, the ratio of the final depth after unloading to the maximum depth of penetration. Because this parameter is experimentally measurable, it is potentially a useful indicator of the type of pressure distribution that can be expected for an elastic–perfectly-plastic material.

Inspection of Fig. 9 reveals that over a fairly significant range of material behavior, the pressure under the indenter is roughly constant and independent of radial position. This is especially true for materials with $E/\sigma_y < 100$. For larger values, there is a tendency for the pressure to decrease with increasing radius, but even for the most extreme material, $E/\sigma_y = 613$, the reduction in pressure from the center to the periphery of the contact is less than 50%.

Assuming, then, that the pressure distribution at maximum load is relatively flat and characterized by a constant pressure, p, one can deduce the effective shape of the indenter from simple elastic contact theory. (The influence of an approximately linear decrease in pressure with increasing radius will be examined later). Johnson has shown that for a uniform pressure applied over a circular contact region of a flat, semi-infinite elastic half-space, the vertical surface displacements $u_z(r)$ are given by

$$u_z = \frac{4(1-\nu^2)pa}{\pi E}\,\mathbf{E}(r/a), \quad (r < a) \quad , \tag{9}$$

where a is the radius of the circle and $\mathbf{E}(r/a)$ is the complete elliptic integral of the second kind evaluated at r/a.[19] The maximum displacement $u_z = 2(1-\nu^2)pa/E$ occurs at the center of the contact, while the displacement at the edge of the contact is given by $u_z = 4(1-\nu^2)pa/(\pi E)$. These displacements would be produced by an indenter with a profile described by

$$z(r) = \frac{4(1-\nu^2)pa_{max}}{\pi E}\left[\frac{\pi}{2} - \mathbf{E}(r/a_{max})\right] \quad . \tag{10}$$

Note that the contact radius appearing in this expression is not "a" but "a_{max}," the contact radius at maximum load. This is because the constant pressure distribution is achieved only when the effective indenter reaches the maximum load.

The effective indenter shape defined by Eq. (10) is plotted in nondimensional form in Fig. 10 where $\bar{r} = r/a_{max}$ and $\bar{z} = z/[4(1 - \nu^2)pa_{max}/(\pi E)]$. Comparison with Fig. 6 shows that the shape is qualitatively similar to that predicted by the finite element simulation of alloy 8009. Also shown in the figure is the power-law approximation of the shape $\bar{z} = B\bar{r}^n$. On the basis of the regression analysis, the best fit is achieved with $B = 0.548$ and $n = 2.61$. The shape of the effective indenter is thus described by

$$z(r) = 0.548\,\frac{4(1-\nu^2)pa_{max}}{\pi E}\left(\frac{r}{a_{max}}\right)^{2.61} \quad . \qquad (11)$$

The power-law fit provides a reasonable description. Note that the power-law exponent that describes the shape, $n = 2.61$, is in excellent agreement with that derived from finite element simulation of alloy 8009 ($n = 2.63$). Moreover, since $m = 1 + 1/n$, the simple analysis suggests that the value of the exponent m describing the unloading curve in $P = \alpha(h - h_f)^m$ should be 1.38, once again in good agreement with the finite element results for alloy 8009 ($m = 1.32$) and experimental data obtained with a Berkovich indenter ($m = 1.36$). Thus, to the extent that the peak load pressure distribution is constant, there is a simple rationale for why the effective indenter takes on the shape that it does and why indentation unloading curves obtained with conical and Berkovich indenters can be described by a simple power-law relations like Eq. (3) with power-law exponents near $m = 1.38$. Although not explicitly addressed in this work, the same principles could be applied to spherical indentation. Since it is known that pressure distributions for elastic–plastic contact with a sphere are relatively flat,[19] one might expect a similar effective indenter shape to describe the unloading behavior.

One can use these results to derive an equation that describes the entire unloading curve for a material in which the peak load pressure distribution is uniform. The relation follows directly from Eq. (7). Using $B = 0.548$ and $n = 2.61$, noting that the pressure, p, is equivalent to the hardness, H, and noting that the contact radius at peak load is related to the hardness through

$$a_{max} = \left[\frac{P_{max}}{\pi H}\right]^{1/2} \quad , \qquad (12)$$

Eq. (7) reduces to

$$P = 0.858\,(P_{max})^{0.31}\left[\frac{E_{eff}}{\sqrt{H}}\right]^{1.38}(h - h_f)^{1.38} \quad . \qquad (13)$$

This equation can also be expressed in the convenient nondimensional form:

$$\frac{P}{P_{max}} = 0.858\left[\frac{E_{eff}}{\sqrt{P_{max}H}}(h - h_f)\right]^{1.38} \quad . \qquad (14)$$

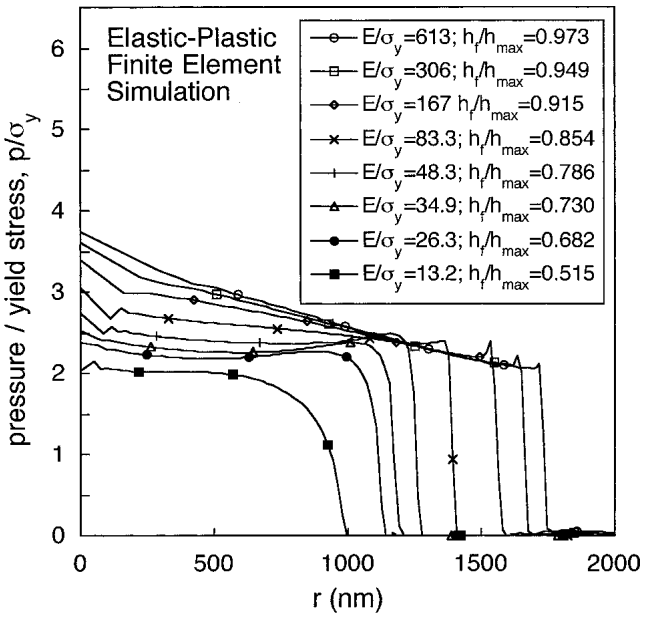

FIG. 9. Peak load pressure distributions for a wide variety of materials determined by finite element simulation.

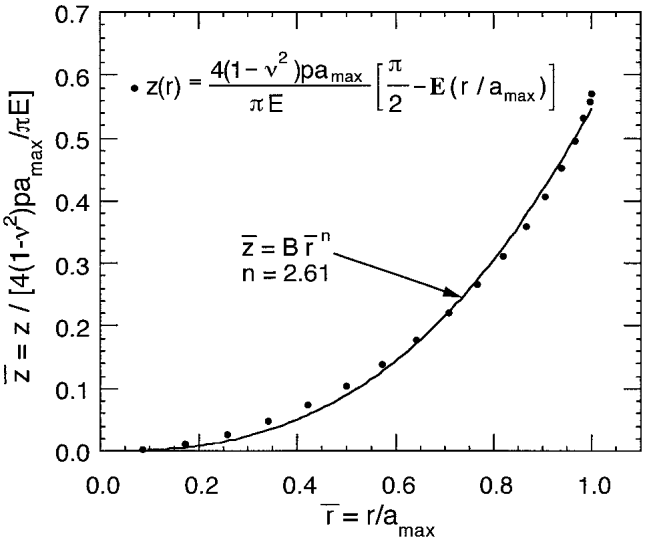

FIG. 10. Effective indenter shape that produces a constant distribution of pressure at peak load.

How well Eq. (13) works in describing real nanoindentation unloading curves is illustrated in Fig. 11, where the predictions of Eq. (13) are compared with experimental data obtained with a Berkovich indenter.[4] The predictions of the equation are based on known values of E, ν, H, and P_{max} given in Table II. Given the approximate nature of the derivation, the agreement between theory and experiment is relatively good.

VI. EXTENSION TO OTHER MATERIALS

To further explore and expand on the utility of the effective indenter shape concept, the results of finite element simulations for a wide variety of elastic–perfectly-plastic materials were analyzed using the concepts developed in the previous sections. The findings are summarized in Table III. The first column gives the ratio of the elastic modulus to yield strength, E/σ_y, which can be used to assess the type of material behavior. Very soft metals have E/σ_y near the high end of the range, while the low end is characterized by hard glassy materials with relatively low elastic moduli such as fused silica. Most metals and ceramics fall in between. The other columns in the table are quantities derived from the load–displacement data and/or contact profiles generated in the finite element simulations.

Figure 12 shows for several values of E/σ_y the effective indenter shape deduced from the shape of the unloaded hardness impression using the same procedures as those for alloy 8009 in Sec. IV. Also included in the figure are curve fits for the effective shape according to the power law relation of Eq. (6). As for the case of alloy 8009, the power law provides a good description of the effective shape. As shown in Table III, the power-law exponents derived from the fits vary over the range $2.10 \leqslant n \leqslant 5.28$. The corresponding unloading curve exponents computed from Eq. (8) fall in the range $1.16 \leqslant m \leqslant 1.48$. On the basis of these results, it can thus be concluded that unloading curve exponents should be slightly material dependent through the way the relative elastic and plastic properties influence the pressure distribution. Significantly, however, all the values of m fall in the range 1.16–1.48, in close agreement with the experimental observations in Table I.

In Fig. 13, the unloading curve exponents are plotted as a function of the ratio of the final depth of penetration to the maximum depth, h_f/h_{max}. This experimentally

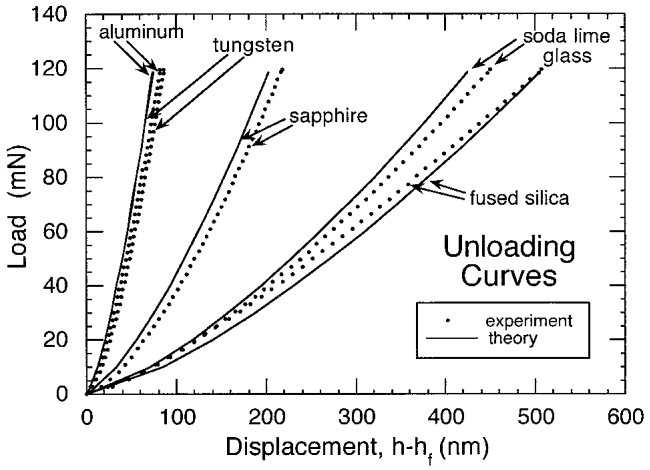

FIG. 11. Comparison of theoretical predictions for unloading curves with experimental data obtained in nanoindentation experiments with a Berkovich indenter. Experimental data are from Ref. 4.

TABLE II. Material properties and experimental parameters used to predict unloading curves using Eq. (13). Values for the effective modulus E_{eff} were computed from Eq. (5) using $E_i = 1141$ GPa and $\nu = 0.07$ for the diamond indenter.[4]

Material	P_{max} (mN)	E (GPa)	ν	E_{eff}	H (GPa)
Aluminum	118.32	70.4	0.347	74.8	0.21
Soda-lime glass	118.37	70.0	0.23	69.4	5.9
Sapphire	118.50	403	0.234	310.8	26.9
Fused silica	118.43	72.0	0.17	69.6	8.4
Tungsten	118.43	409.8	0.28	320.4	3.8

TABLE III. Results of finite element simulation for a variety of elastic–perfectly-elastic materials.

E/σ_y	E/H	h_f/h_{max}	n	$m = 1 + 1/n$
613	237	0.973	2.10	1.48
306	118	0.949	2.28	1.44
167	63.7	0.915	2.70	1.37
83.3	32.6	0.854	3.81	1.26
48.3	19.3	0.786	5.11	1.20
34.9	14.5	0.730	5.78	1.17
26.3	12.0	0.682	6.40	1.16
13.2	7.31	0.515	5.28	1.19

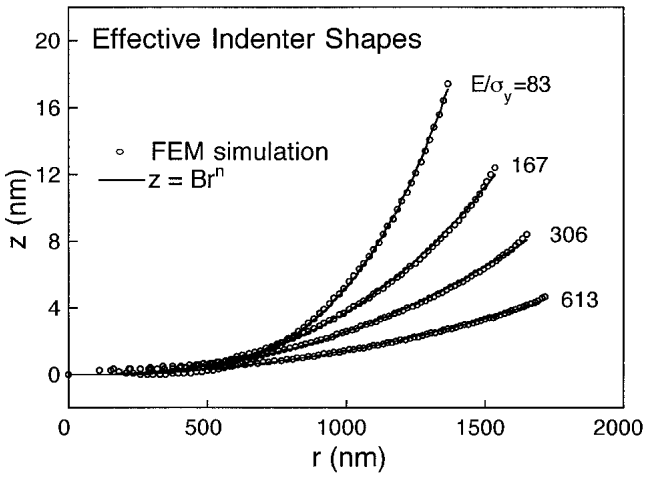

FIG. 12. Effective indenter shape for several elastic–perfectly-plastic materials with different values of E/σ_y.

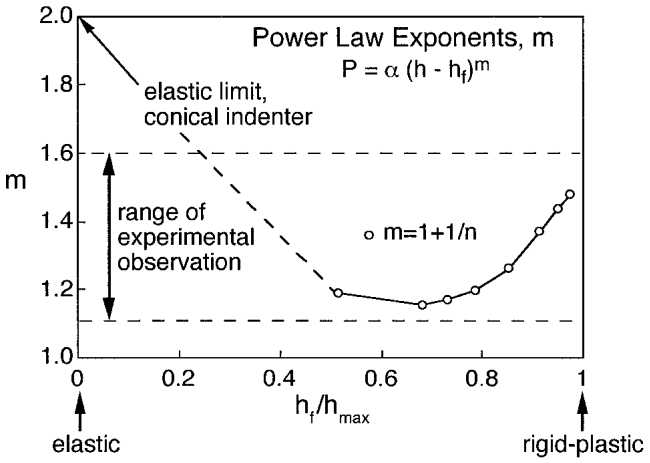

FIG. 13. Dependence of unloading curve exponents, *m,* on the parameter, h_f/h_{max}, as determined by finite element simulation of elastic–perfectly-plastic materials.

measurable parameter provides a useful gauge of the relative contributions of elasticity and plasticity to the indentation displacements. When $h_f/h_{max} = 0$, the displacement is entirely recovered during unloading, implying a perfectly elastic contact, whereas when $h_f/h_{max} = 1$, none of the displacement is recovered and the material is rigid–plastic. Intermediate values of h_f/h_{max} represent the continuum of elastic–perfectly-plastic behavior. The two limiting values for h_f/h_{max} are shown as the extremes of the abscissa in Fig. 13.

The data in Fig. 13 suggest that there may be a systematic relationship between m and h_f/h_{max}, with m going through a minimum near $h_f/h_{max} = 0.7$. From a theoretical perspective, the value of m should rise to 2 at $h_f/h_{max} = 0$, since this exponent is what is expected for purely elastic contact by a rigid cone.[8] This being the case, one might expect a minimum in the curve and thus a lower limit on m. The behavior of real materials is probably somewhat more complex due to influences of work-hardening that have not been taken into account in the finite element simulations. Note that in the rigid–plastic limit, $h_f/h_{max} = 1$, the value of m should theoretically tend to 1, but there is no evidence for this in the finite element simulations. The decrease must therefore occur for material parameters outside the range examined here.

VII. OTHER PRESSURE DISTRIBUTIONS

The finite element results presented in Fig. 9 suggest that although the pressure distribution for many elastic–perfectly-plastic materials is approximately constant, there are some materials, particularly those with high E/σ_y values, for which the pressure decreases with increasing distance from the center of the contact. To explore the effect this may have on the results and conclusions, an analysis was undertaken to determine the effective indenter shape for a pressure distribution $p(r)$ that decreases linearly with the radial distance, r, from the center of the contact circle according to the relation

$$p(r) = p_0 - \Delta p \frac{r}{a} \quad . \tag{15}$$

In this expression, p_0 is the maximum pressure at the center of the contact and Δp is the reduction in pressure at the contact edge relative to the maximum value.

Using a methodology outlined in the Appendix, the effective indenter shape for this pressure distribution is

$$z(r) = \frac{4(1-\nu^2)p_0 a_{max}}{\pi E}\left\{\left(1 - \frac{1}{2}\frac{\Delta p}{p_0}\right)\left[\frac{\pi}{2} - \mathbf{E}(r/a_{max})\right] + \frac{r^2}{4{a_{max}}^2}\frac{\Delta p}{p_0}\int_0^{\pi/2}\sin^2\phi \ln\left[\frac{1+\sqrt{1-(r/a_{max})^2\sin^2\phi}}{1-\sqrt{1-(r/a_{max})^2\sin^2\phi}}\right]d\phi\right\} \quad . \tag{16}$$

Note that the for a constant pressure distribution, $p_0 = p$, $\Delta p = 0$, and Eq. (16) reduces to the simple closed-form of Eq. (10). For the more general case, numerical evaluation of the integral on the right hand side of Eq. (16) is required.

To explore the implications of this result, the pressure distribution in Fig. 9 determined by finite element simulation for the material with $E/\sigma_y = 613$ was fitted according to Eq. (15) to establish appropriate values for p_0 and Δp. This material was chosen because it exhibits the largest decrease in pressure and also because the decrease is very nearly linear. Numerical integration of Eq. (16) gave the effective shape, which once again was found to be well-described by the power-law relation of Eq. (6) with $n = 1.89$. The corresponding value of the unloading curve exponent computed from Eq. (8) is $m = 1.53$. This value is considerably larger than the value $m = 1.38$ derived in Sec. V for the effective indenter that produces a constant pressure distribution, thereby indicating that the effect of a linearly decreasing pressure distribution is to increase the value of m (and decrease the value of n). This is consistent with the results in Table III and Fig. 9 that show that materials with higher m values are generally those with the larger gradients in pressure.

VIII. IMPLICATIONS FOR NANOINDENTATION TESTING

In addition to providing a general physically justifiable rationale for the mathematical form of nanoindentation unloading curves, the concepts developed here can be used to provide a theoretical basis for an important aspect

of nanoindentation data analysis that has heretofore been based largely on empiricism. In particular, in the frequently used method of Oliver and Pharr for measuring hardness and elastic modulus from nanoindentation load–displacement data,[4] a simple expression is used to compute the contact depth, h_c (the depth along which contact is made between the indenter and the specimen), from experimentally measured quantities. The expression is

$$h_c = h_{max} - \epsilon \frac{P_{max}}{S} \quad , \tag{17}$$

where P_{max} is the maximum load, h_{max} is the maximum displacement, $S = dP/dh$ is the unloading contact stiffness, and ϵ is a constant that depends on the geometry of the indenter. The value of ϵ is formally defined by the relation

$$\epsilon = S \frac{h_s}{P_{max}} = 2E_{eff}a_{max} \frac{h_s}{P_{max}} \quad , \tag{18}$$

where h_s is the depth to which the surface sinks in at the perimeter of the contact.[4] For simple indenter geometries, values of ϵ can be computed from results given by Sneddon;[8] specifically, $\epsilon = 0.72$ for a conical indenter, $\epsilon = 0.75$ for a parabola of revolution, and $\epsilon = 1.0$ for a flat cylindrical punch.[4] Given that a Berkovich indenter is geometrically more like a cone than a parabola of revolution or a flat punch, one might expect the value $\epsilon = 0.72$ to be most applicable. However, on the basis of a large number of experimental obser- vations, Oliver and Pharr concluded that the best value for ϵ is 0.75.

The effective shape concept provides a means by which this value of ϵ can be understood. In simple terms, because the elastic unloading process is described by an effective indenter shape that is more like a parabola of revolution than a cone, the value $\epsilon = 0.75$ is more appropriate than $\epsilon = 0.72$. However, a more precise estimate of ϵ may be achieved using the methods developed here in conjunction with Sneddon's method for determining the surface displacement at the contact perimeter for indentation of an elastic half-space by a rigid axisymmetric indenter of general power-law profile.[8] Substituting Sneddon's result for h_s into Eq. (18) yields

$$\epsilon = \left(1 - \frac{2\Gamma(1/2 + n/2)}{n\sqrt{\pi}\Gamma(n/2)}\right) \frac{1+n}{n} \quad , \tag{19}$$

which uniquely relates ϵ to the power-law exponent n for the shape of the effective indenter. Noting that the power-law exponent that best describes the effective shape in the analysis presented in Sec. V is $n = 2.61$, Eq. (19) yields $\epsilon = 0.761$. This theoretically justifiable value for ϵ is in good agreement with the empirical value established by Oliver and Pharr.[4] Alternatively, using Eq. (8), Eq. (19) can be rewritten in terms of the unloading curve power law exponent m as

$$\epsilon = m \left(1 - \frac{2\Gamma\left(\frac{m}{2(m-1)}\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{2(m-1)}\right)} (m-1)\right) \quad , \tag{20}$$

which is plotted in Fig. 14. Note that over the range of most experimental observations, i.e., $1.2 \leqslant m \leqslant 1.6$, the value of ϵ varies mildly between 0.79 and 0.74, with 0.76 being a reasonable average. However, since it was shown in Sec. VI that there may be some variability in the value of m depending on the elastic and plastic behavior of the material, a better value of ϵ might be obtained by experimentally measuring m and then using Eq. (20) to determine the corresponding value of ϵ. This procedure could be easily implemented in experimental methods.

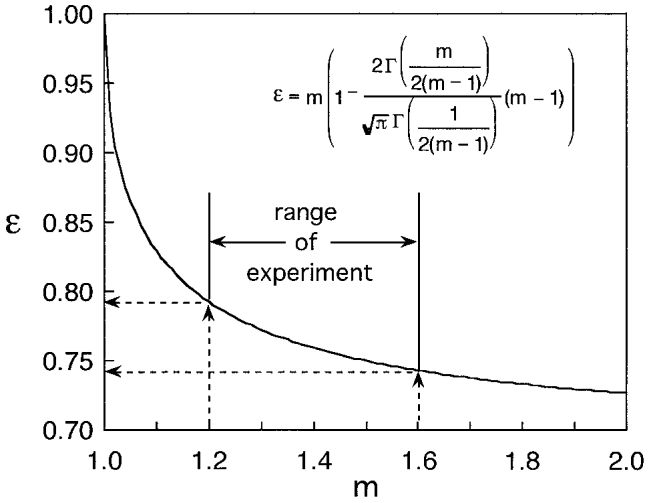

FIG. 14. Relation between ϵ and m predicted by Eq. (20).

IX. CONCLUSIONS

Finite element simulation of elastic/plastic indentation of an aluminum alloy and several elastic–perfectly plastic materials by a rigid conical indenter with a depth-to-area relationship the same as the Berkovich diamond suggests that nanoindentation unloading curves should be well-described by the power-law relation $P = \alpha(h - h_f)^m$. The value of the unloading curve exponent m derived from the simulations is consistent with experimental findings that m varies over the range 1.2–1.6. Although Sneddon's analysis for the deformation of a flat elastic half-space by a rigid conical indenter predicts $m = 2$, the smaller values of m observed in experiment and finite element simulation are explained by the distortion of the surface from a flat plane configuration caused by the

formation of the permanent hardness impression. The surface distortions can be accounted for by defining an effective indenter with a shape such that when pressed into a flat elastic half space, the resulting normal surface displacements are the same as those produced by a conical indenter pressed into the plastically deformed surface of the hardness impression. Simple arguments based on the assumption of a constant pressure distribution under the indenter show that the shape of the effective indenter is well-described by the relation in Eq. (11) (see Sect. V), where p is the pressure, a_{max} is the contact radius at peak load, E is Young's modulus, and ν is Poisson's ratio. The exponent describing the unloading curve consistent with this shape is $m = 1.38$. Slightly larger values of m are expected if the pressure under the indenter decreases with distance from the center of the contact, as is the case for soft metals. The same arguments show that indentation unloading curves (P = load; h = displacement) can be quantitatively approximated by the relation in Eq. (14) (see Sect. V), where P_{max} is the maximum load, H is the hardness, and h_f is the final depth after full unloading. The analysis also reveals that the parameter ϵ used in the Oliver–Pharr method for analyzing nanoindentation data to obtain hardness and elastic modulus should have a value close to 0.76.

ACKNOWLEDGMENT

Research through the Oak Ridge National Laboratory SHaRE User Program was sponsored by the Division of Materials Sciences and Engineering, United States Department of Energy, under Contract DE-AC05-00OR22725 with UT-Battelle, LLC.

REFERENCES

1. J.B. Pethica, R. Hutchings, and W.C. Oliver, Philos. Mag. A **48,** 593 (1983).
2. J.L. Loubet, J.M. Georges, O. Marchesini, and G. Meille, J. Tribol. **106,** 43 (1984).
3. M.F. Doerner and W.D. Nix, J. Mater. Res. **1,** 601 (1986).
4. W.C. Oliver and G.M. Pharr, J. Mater. Res. **7,** 1564 (1992).
5. G.M. Pharr and W.C. Oliver, MRS Bull. **17,** 28 (1992).
6. G.M. Pharr, Mater. Sci. Eng. A **253,** 151 (1998).
7. 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.
8. I.N. Sneddon, Int. J. Eng. Sci. **3,** 47 (1965).
9. G.M. Pharr, W.C. Oliver, and F.R. Brotzen, J. Mater. Res. **7,** 613–617 (1992).
10. S.V. Hainsworth, H.W. Chandler, and T.F. Page, J. Mater. Res. **11,** 1987 (1996).
11. K.W. McElhaney, J.J. Vlassak, and W.D. Nix, J. Mater. Res. **13,** 1300 (1998).
12. N.A. Stelmashenko, M.G. Walls, L.M. Brown, and Y.V. Milman, Acta Metall. **41,** 2855 (1993).
13. Q. Ma and D.R. Clark, J. Mater. Res. **10,** 853 (1995).
14. W.D. Nix and H. Gao, J. Mech. Phys. Solids **46,** 411 (1998).
15. A. Bolshakov, W.C. Oliver, and G.M. Pharr, J. Mater. Res. **11,** 760 (1996).
16. A. Bolshakov and G.M. Pharr, J. Mater. Res. **13,** 1049 (1998).
17. T.Y. Tsui, W.C. Oliver, and G.M. Pharr, J. Mater. Res. **11,** 752 (1996).
18. J.C. Hay, A. Bolshakov, and G.M. Pharr, J. Mater. Res. **14,** 2296 (1999).
19. K.L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge, U.K., 1985).
20. K.L. Johnson, in *Engineering Plasticity,* edited by Heyman and Leckie (Cambridge University Press, Cambridge, U.K., 1968).
21. W. Hirst and M.G.J.W. Howse, Proc. R. Soc. A **311,** 429 (1969).

APPENDIX: EFFECTIVE INDENTER SHAPE FOR A LINEARLY DECREASING PRESSURE DISTRIBUTION

The effective indenter shape for a pressure distribution, $p(r)$, that decreases linearly with distance from the center of the contact, r, can be derived using a method outlined by Johnson.[19] The pressure distribution is assumed to be of the form

$$p(r) = p_0 - \Delta p \frac{r}{a} \quad , \tag{A1}$$

where p_0 is the maximum pressure at the center of the contact, Δp is the reduction in pressure at the contact edge relative to the peak value, and a is the radius of the contact circle. Following Johnson,[19] the position of each point in the contact circle can be defined by a set of coordinates (s, ϕ), for which the linear distribution of pressure can be expressed as

$$p(s, \phi) = p_0 - \Delta p \sqrt{r^2 + s^2 + 2rs \cos \phi} \quad . \tag{A2}$$

The vertical displacements of the surface are then given by

$$w(r) = \frac{1 - \nu^2}{\pi E} \int_0^{\pi} \int_{s_1}^{s_2} p(s, \phi) \, ds \, d\phi \quad , \tag{A3}$$

where the limits of integration are

$$s_{1,2} = -r \cos(\phi) \pm \{r^2 \cos^2 \phi + (a^2 - r^2)\}^{1/2} \quad . \tag{A4}$$

Substitution of (A1) and (A2) into (A3) and evaluating the displacement at the center of contact circle yields for the effective indenter shape

$$z(r) = \frac{4(1 - \nu^2) p_o a_{max}}{\pi E} \left\{ \left(1 - \frac{1}{2} \frac{\Delta p}{p_o} \right) \left(\frac{\pi}{2} - \mathbf{E}(r/a_{max}) \right) + \frac{r^2}{4 a_{max}^2} \frac{\Delta p}{p_o} \int_0^{\pi/2} \sin^2 \phi \ln \left[\frac{1 + \sqrt{1 - (r/a_{max})^2 \sin^2 \phi}}{1 - \sqrt{1 - (r/a_{max})^2 \sin^2 \phi}} \right] d\phi \right\} \quad . \tag{A5}$$