Determination of the effective zero point of contact for spherical nanoindentation

Alexander J. Moseson,[a] Sandip Basu, and Michel W. Barsoum
Department of Materials Science and Engineering, Drexel University, Philadelphia, Pennsylvania 19104

(Received 2 August 2007; accepted 25 September 2007)

Accurate determination of the "zero point," the first contact between an indenter tip and sample surface, has to date remained elusive. In this article, we outline a relatively simple, objective procedure by which an effective zero point can be determined accurately and reproducibly using a nanoindenter equipped with a continuous stiffness measurement option and a spherical tip. The method relies on applying a data shift, which ensures that curves of stiffness versus contact radius are linear and go through the origin. The method was applied to fused silica, sapphire single crystals, and polycrystalline iron with various indenter sizes to a zero-point resolution of ≈2 nm. Errors of even a few nanometers can drastically alter plots and calculations that use the data, including curves of stress versus strain.

I. INTRODUCTION

Instrumented indentation is a valuable method for characterizing the mechanical behavior of materials, especially that of single crystals and thin films. Field and Swain,[1,2] and Oliver and Pharr[3] have made significant progress in developing the technique, but recognize the significant hurdles yet to overcome. One such obstacle is the accurate and reliable determination of the zero point, where the indenter tip makes first contact with the sample surface.[3–9] At this point, both the applied indentation load, P, and the total displacement, h_t, or indentation depth, should be zero, though the sample stiffness, S, may appear positive. Inset 1 in Fig. 1(a) shows the conundrum, wherein a curve of P versus h_t has no clear zero point.

To date, methods of various sophistication have been proposed to qualitatively or quantitatively determine the zero point.[10–17] For example, one is to simply plot P versus h_t and choose a point where P first exceeds a certain threshold, while another uses a video camera.[14,16] For instruments with continuous stiffness measurement (CSM) capabilities, Oliver and Pharr[3] have suggested using the point at which S reaches a local minimum before increasing steadily. Alternatively, they suggested using abrupt changes in CSM harmonic displacement or phase angle if they are clearer, but all three options require some subjectivity, however small. They claim the method has an accuracy of 2 nm, but we have found that it identifies the zero point too early, sometimes by as much as 14 nm. This may seem insignificant, but as we show below, the difference can be quite important. Chudoba and colleagues[10,11] and Ullner[17] have suggested using regression on the P-versus-h_t plots. The former advocates an iterative numerical method to fit the data to a variation of the Hertzian model, replacing the conventional parameters of tip radius and effective modulus with an optimized proportionality constant, and forcing the data to go through zero.[10,11] The latter advocates optimizing the terms of a second-order polynomial to fit the data.[17] Finally, a variation on the method of Oliver and Pharr[3] is to locate the zero point where S first meets or exceeds 200 N/m. This is based on an assumption that while the tip is hanging free, other factors such as vibration produce values below 200 N/m, and that this small value first appears when the tip makes contact with the surface. This method indeed works in some cases, but in the authors' experience it can falsely locate the surface up to hundreds of nanometers away from the actual zero point. In our deductive data analysis, we found that this is most likely the method used by the MTS instrument (MTS Nano Instruments, Oak Ridge, TN) that we used, and all δ values (see below) in this article are given with respect to the original zero assumed by the instrument. In summary, and while some of these methods may be somewhat successful, they are limited in that both load and displacement values are noisy at low loads and are greatly impacted by the zero point.

Herein we present a robust, simple, accurate, and objective method for reliably determining the effective zero point for instrumented spherical indentation equipped with the CSM option. In essence, our method shifts the

[a]Address all correspondence to this author.
e-mail: ajm54@drexel.edu
DOI: 10.1557/JMR.2008.0012

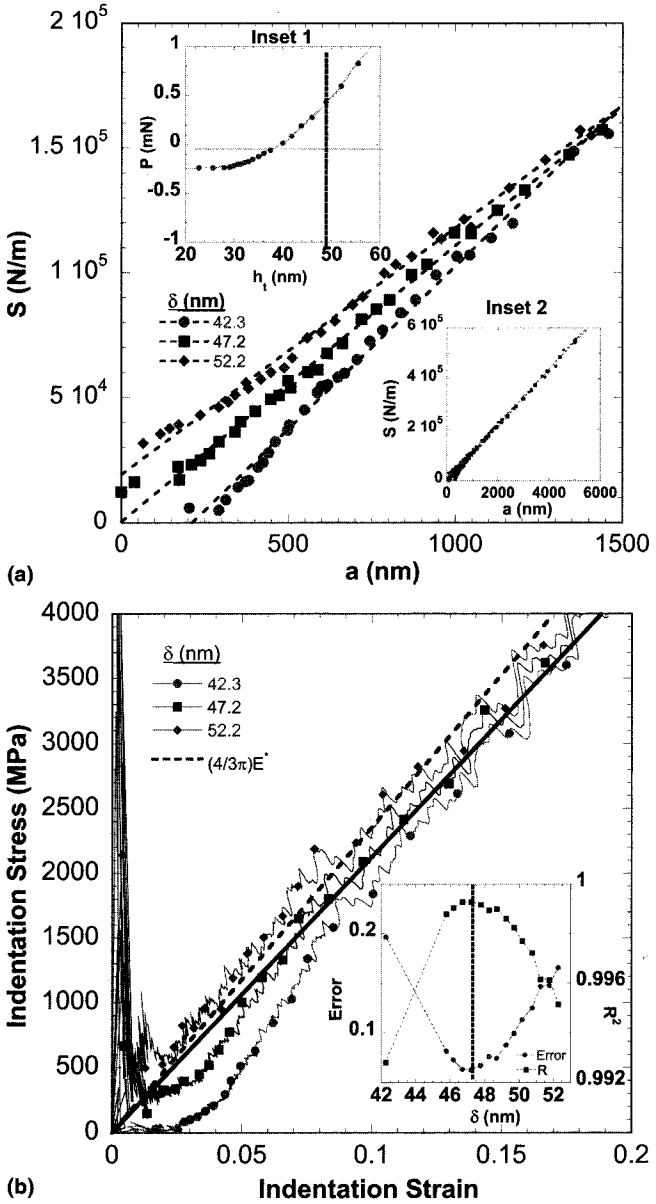

FIG. 1. Fused silica, 13.5-μm indenter. (a) S versus a for various δ values, for a region near the origin. Linear regression is shown. Inset 1: P versus h_t at a region near the supposed S = 200 N/m zero point used by manufacturer to define the zero point. The dashed line denotes the location of the true effective zero point. Inset 2: Full data set. (b) Indentation stress-versus-strain curves; the center curve has the correct δ. The solid line is linear regression of the center curve, and the dashed line is the expected slope as calculated from $4E^*/3\pi$. Inset plots error and R^2 versus δ.

P and h_t columns of a given data set in such a way as to ensure that the curves of S versus the contact radius, a, are straight lines that go through the origin. The power of the technique lies in the fact that, if needed, the results can be linearly back-extrapolated from a region where the signal-to-noise level is high (compared to shallow depths) back through the origin, thus circumventing the myriad problems encountered at low loads.[3] In that sense, our zero point is an "effective" zero point (i.e., the zero point that one would have obtained had the surface been atomically flat and perfectly normal to the loading direction). This effective zero point may, or may not, correspond to the very first point of contact between the indenter and the surface.

II. SPHERICAL INDENTATION MODEL

In a previous article,[18] and in several subsequent ones,[19,20] we outlined a method for producing nanoindentation stress–strain curves from load-displacement curves obtained with spherical indenters.[18] Inherent in that method are the tools necessary for our zero-point method. For an isotropic elastic solid indented with a spherical indenter, the sample stiffness S and contact radius a are related by[21]

$$S = 2E_{\text{eff}}a \quad , \tag{1}$$

where S is the sample stiffness, given by

$$\frac{1}{S} = \frac{1}{S_s} - \frac{1}{S_f} \quad , \tag{2}$$

where S_s is the stiffness value of the system, reported by the CSM, and S_f is the load-frame stiffness, as given by the instrument manufacturer. In our case, this was 5.5 MN/m. E_{eff} is the composite modulus, defined by[3]

$$\frac{1}{E_{\text{eff}}} = \frac{1-\nu^2}{E} + \frac{1-\nu_i^2}{E_i} \quad , \tag{3}$$

where E_i and ν_i, respectively, refer to the Young's modulus and Poisson's ratio of the diamond indenter (1140 GPa and 0.07, respectively); the E and ν values are those of the sample. The contact depth, h_c, compared to the distance from the circle of contact to the maximum penetration depth, is given by[1,3,18]

$$h_c = h_t - \frac{3}{4}\frac{P}{S} \quad . \tag{4}$$

Once h_c is calculated, a is determined from

$$a = \sqrt{2R_t h_c - h_c^2} \quad , \tag{5}$$

where R_t is the indenter tip radius. Finally, making use of the Hertz equation[1,3,22,23] and work by Sneddon[24] and Johnson,[21] it can be shown that:

$$\frac{P}{\pi a^2} = \frac{4}{3\pi}E^*\left(\frac{a}{R_t}\right) \quad . \tag{6}$$

The left side of Eq. (6) is defined as the indentation stress, mean contact hardness, or Meyer hardness.[18,23] The expression in parentheses, to the right, is defined as the indentation strain.[18,23] The modulus E^* in Eq. (6)

relates to the slope of the elastic portion of the stress–strain curve, and would equal E_{eff} as found from the slope of the S–a curve [Eq. (1)] if and only if all assumptions made[21] in its derivation (i.e., atomically flat surface, surface perpendicular to tip, isotropic material, and perfectly spherical shape) are valid. Note that in our method, Eqs. (4) and (5) are used to find a, not Eq. (1) (see below).

Originally, Eq. (4) included a small correction, δ, that helped ensure that the early regions of the stress–strain curves behaved in a linear elastic fashion. The origin, or need for, δ was not clear at that time.[18] We have since discovered that δ arose from using an incorrect zero point, leading to the method described herein.

III. EFFECTIVE ZERO POINT DETERMINATION

For a properly zeroed sample, according to Eq. (1), a plot of S versus a should be a straight line that goes through the origin, with a slope of 2 E_{eff}. The essence of our method lies in finding the datum point, which, if it is an effective zero point, would make the S-versus-a curves go through the origin as predicated by Eq. (1). To create the S-versus-a plots, the former is known from the CSM, and we use Eqs. (4) and (5) to find a. We have chosen to express the shift as δ, the difference in h_t between the correct effective zero point, X_z, and the first point X_o where S is 200 N/m, compared to the one given by the instrument. (Note that X_o is defined as such here for convenience, as our instrument was programmed to choose the zero point based on S = 200 N/m; this may vary by instrument. The δ value itself is not so important as identifying the correct effective zero point.) To start, we chose points X_j near (±10 nm) where P definitely becomes positive. To find the correct effective zero point, we simply shift the P and h_t columns by subtracting (even if the values are negative) P_j and $h_{t,j}$, respectively, from the entire column. Data points with negative values of h_t are discarded, and S versus a is plotted. The shift that results in a S-versus-a line that best goes through the origin is the sought-after effective zero point.

Linear regression is used to quantitatively determine the degree to which each S-versus-a curve is linear and goes through the origin (i.e., a straight line forced through the origin). Note that none of the data sets interact. We used two criteria for quantifying the curve fits, though one would suffice. The first is the standard error, defined as the average vertical difference between each datum point and the best-fit line forced through the origin. The second is the well-known correlation coefficient, R^2, again, with respect to the same best-fit line forced through the origin. The value of δ_j that minimizes the error or maximizes R^2 is the correct δ. A detailed example of the procedure can be found in Appendix 2 of Ref. 25.

IV. EXPERIMENTAL DETAILS

For this work, we used a Nanoindenter XP system (MTS) with a CSM attachment. All tests were carried out with a load rate over a load factor of $(dP/dt)/P = 0.1$ and an allowable drift rate of 0.05 nm/s. The harmonic displacement for the CSM was 2 nm with a frequency of 45 Hz. Two diamond spherical tips, with radii of 13.5 and 1 μm, were used. The three materials used were: fused silica (GM Associates Inc., Oakland, CA); C-orientation sapphire single crystal (Kyocera Industrial Ceramics, Vancouver, WA); and iron (99.65%, SurePure Chemetals, Florham Park, NJ). The maximum load was 690 mN for fused silica, and 50 mN for iron and sapphire.

V. RESULTS AND DISCUSSION

Figure 1 shows results for fused silica with the 13.5-μm indenter. The problem at hand is apparent in inset 1 of Fig. 1(a), where the P-versus-h_t plot shows no clear zero point. The correct δ, by our method, is indicated as a vertical dashed line. Note its location in contradistinction to the more intuitive point where P begins increasing, at a δ of around 30 nm. Figure 1(a) shows S versus a plotted for three δ values over a span of 10 nm. Inset 2 shows the entire data set, wherein it is apparent that after ≈2500 nm, the value of δ is no longer of consequence. The inset of Fig. 1(b) plots the linear regression R^2 values and the above-defined standard error from the data forced through zero, at various δ values. In this case, δ_z is clearly ≈47.2 nm. Figure 1(b) shows a plot of indentation stress versus strain, as defined by Eq. (6).

Figures 2 and 3 show similar results for iron and sapphire, respectively. Here again, the instrument value for the zero point is slightly off. Note the early spike in the results that appears in Figs. 1(b), 2(b), and 3(b), especially at strains <0.01. This spike, to our knowledge not previously highlighted or understood, arises primarily because at low h_t, the stress is quite sensitive to the values of P, through a^2. Its effect on the early region of the stress–strain curve is best seen in inset 2 in Fig. 2(b), which is a replot of the center curve with δ = 7.5. If the as-received P = 0.04 mN, is used, the spike is huge (solid circles). When P is zeroed according to P_z (the P values for the chosen δ), the curve is well behaved and the spike disappears (solid squares). Surprisingly, adding just 0.02 mN results in a huge spike (open circles); subtracting 0.02 mN results in negative stresses (solid triangles). This result was unanticipated; 0.02 mN is only 0.08% of the maximum load for this test, and 0.003% of the full scale of the hardware. For this reason, discarding the first few early outliers (i.e., the spike at strain ≤0.1) in the stress–strain plots is permissible, and, because the stress–strain curves merge at strains >0.15, would have no effect on the final results or conclusions. Herein, they

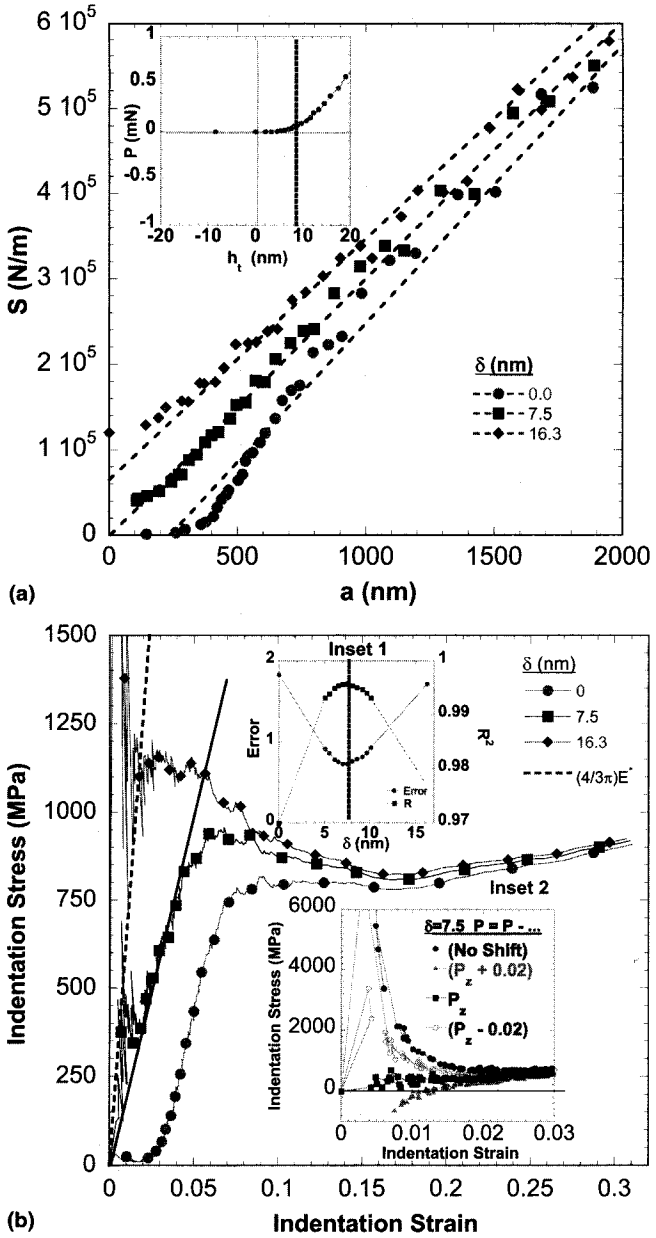

FIG. 2. Iron, 13.5-μm indenter. (a) S versus a for various δ values, for the region near the origin. Linear regression is shown. Inset: Raw P versus h_t at the region near the supposed S = 200 N/m zero point. The dashed line denotes the true location of the effective zero point as determined herein. (b) Indentation stress-versus-strain curves; the center curve has the correct δ. The solid line is a linear regression of the center curve; the dashed line is the expected slope as calculated from $4E^*/3\pi$. Inset 1 plots error and R^2 versus δ. Inset 2 plots the effect of changing the values of the P data column on the stress–strain curves for δ = 7.5 nm. The filled circles use as-received data, P_0; filled squares, P_0–P(δ = 7.5). Open circles, value for filled squares + 0.02 mN; filled triangles, value for filled squares –0.02 mN.

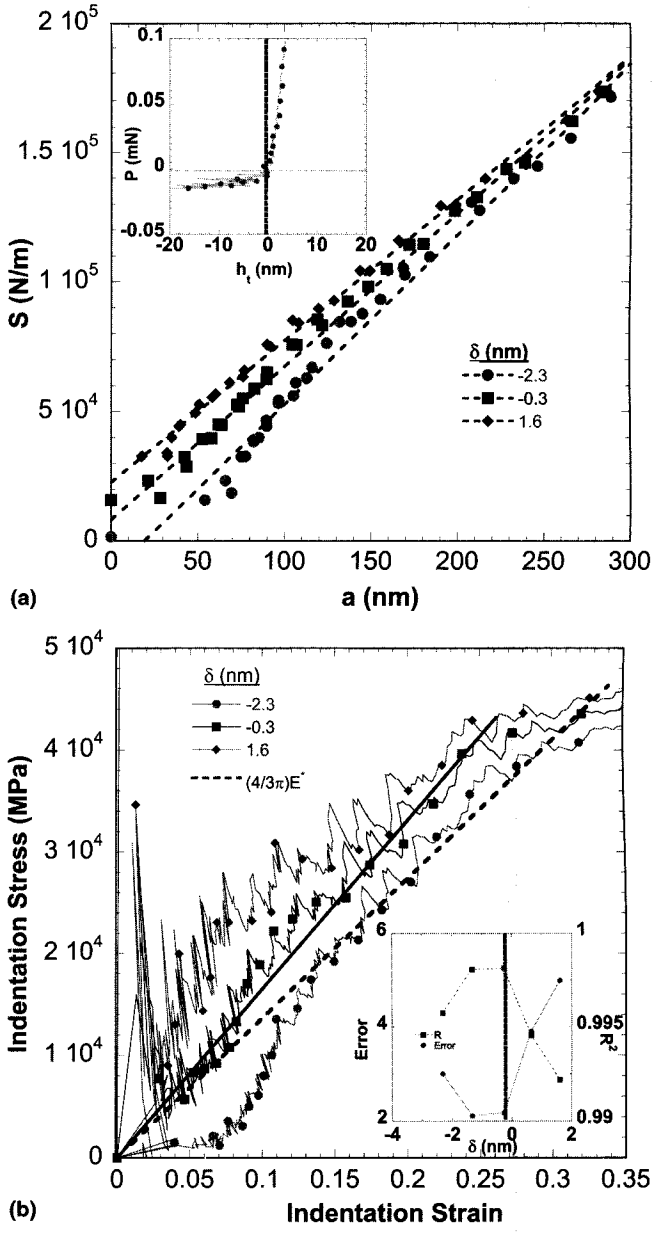

FIG. 3. Sapphire, 1-μm indenter. (a) S versus a for various δ values, for the region near the origin. Linear regression shown. Inset: Raw load versus displacement at the region near the supposed S = 200 N/m zero point. The dashed line denotes the true location of the effective zero point as determined herein. (b) Indentation stress-versus-strain curves; the center curve has the correct δ. The solid line is a linear regression for the center curve, and the dashed line is the expected slope as calculated from $4E^*/3\pi$. Inset plots error and R^2 versus δ.

were left intact for the purpose of discussion, and the complete data sets were used in our calculations. Note that removing these points would not affect the determination of δ nor any of the S-versus-a curves presented here. Removing the early points also greatly mitigates the effects of complex surface phenomena,[3] further bolstering the robustness of our method. These comments notwithstanding, gratifyingly, when the correct zero point was chosen the spike all but disappears, indirectly confirming the correctness of our approach. Given the sensitivity of the curves to P, this result is quite remarkable.

When E_{eff}, obtained from the slopes of the S-versus-a curves divided by 2, for all three samples are plotted as a function of δ (Fig. 4), it is clear that E_{eff} is a function δ. This is an important result because the choice of the wrong zero point can lead to an incorrect determination of E_{eff}. For example, for both Fe and sapphire, an error of only ≈2 nm in the choice of the effective zero point results in an error of 4% or more in E_{eff}.

According to Eq. (6), the slope of the indentation stress-versus-strain curve should equal $4E^*/3\pi$. The inclined dashed lines shown in Figs. 1(b), 2(b), and 3(b) represent the $4E^*/3\pi$ line; the solid inclined lines, on the other hand, represent the least-squares fit of the data points shown in the linear regime forced through zero. The following comments are salient. (i) In the case of silica and sapphire, $E^* \approx E_{eff}$. This is especially true considering that Eq. (6) was derived assuming a perfect sphere indenting a perfectly perpendicular, atomically smooth, elastically isotropic surface. The latter is probably only true here for fused silica. The agreement would have also been more obvious had we chosen to plot the results for larger increments of δ as done in Fig. 2(b). The correlation is also excellent for ZnO,[18,19] Al,[18] and GaN.[20] (ii) For Fe, the dashed line is approximately three times steeper than the solid line (i.e., $E^* \approx 3\ E_{eff}$. As previously discussed,[18] this difference, only manifested when testing metals, is not understood and deserving of future research. (iii) In principle, the linearity of the stress–strain curves, and the need that they pass through the origin, can also be used to find the actual location of the effective zero point instead of, or in addition to, the method outlined herein.

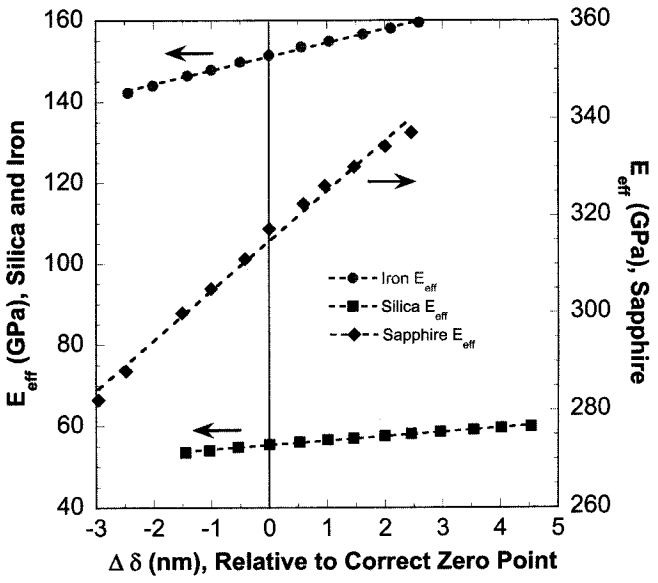

FIG. 4. Dependence of E_{eff} on δ for all three solids tested. Dashed lines are the linear regression for each data set. The change in δ is given with reference to the correct effective zero point for each sample, respectively (i.e., for the correct delta, Δδ = 0). Arrows indicate the corresponding y-axis for data.

VI. SUMMARY AND CONCLUSIONS

The results above demonstrate the method on three different materials, including metal and ceramic, and with two spherical indenter sizes. The importance and sensitivity of correctly identifying the effective zero point is highlighted in the stress-versus-strain curves. For example, in Fig. 3(b), a difference of only 2 nm results in large variations in the indentation stress–strain curves, which previously were left unexplained.

The effective zero-point resolution for our setup is estimated to be ≈2 nm, but this could vary, depending on the hardware and loading schemes. We hope that this work will inspire further research to make instrumented indentation an ever more valuable characterization tool. We suggest, for future investigation, the expansion of this method to other tip geometries, phenomena such as those seen for Fe, full comparison with other methods, and applications of the method.

ACKNOWLEDGMENT

This work was funded by the Army Research Office (ARO) (DAAD19-03-1-0213).

REFERENCES

1. J.S. Field and M.V. Swain: Determining the mechanical-properties of small volumes of material from submicrometer spherical indentations. *J. Mater. Res.* **10,** 101 (1995).
2. J.S. Field and M.V. Swain: The indentation characterisation of the mechanical properties of various carbon materials: Glassy carbon, coke and pyrolytic graphite. *Carbon* **34,** 1357 (1996).
3. W.C. Oliver and G.M. Pharr: Measurement of hardness and elastic modulus by instrumented indentation: Advances in understanding and refinements to methodology. *J. Mater. Res.* **19,** 3 (2004).
4. J.L. Bucaille, E. Felder, and G. Hochstetter: Identification of the viscoplastic behavior of a polycarbonate based on experiments and numerical modeling of the nano-indentation test. *J. Mater. Sci.* **37,** 3999 (2002).
5. P. Grau, G. Berg, W. Fraenzel, and H. Meinhard: Recording hardness testing problems of measurement at small indentation depths. *Phys. Status Solidi A* **146,** 537 (1994).
6. N. Huber and E. Tyulyukovskiy: A new loading history for identification of viscoplastic properties by spherical indentation. *J. Mater. Res.* **19,** 101 (2004).
7. Z. Li, K. Herrmann, and F. Pohlenz: A comparative approach for calibration of the depth measuring system in a nanoindentation instrument. *Measurement* **39,** 547 (2006).
8. B. Rother, A. Steiner, D.A. Dietrich, H.A. Jehn, J. Haupt, and W. Gissler: Depth-sensing indentation measurements with Vickers and Berkovich indenters. *J. Mater. Res.* **13,** 2071 (1998).
9. E. Tyulyukovskiy and N. Huber: Neural networks for tip correction of spherical indentation curves from bulk metals and thin metal films. *J. Mech. Phys. Solids* **55,** 391 (2007).

10. T. Chudoba, M. Griepentrog, A. Dück, D. Schneider, and F. Richter: Young's modulus measurements on ultra-thin coatings. *J. Mater. Res.* **19,** 301 (2004).
11. T. Chudoba, N. Schwarzer, and F. Richter: Determination of elastic properties of thin films by indentation measurements with a spherical indenter. *Surf. Coat. Technol.* **127,** 9 (2000).
12. A.C. Fischer-Cripps: Critical review of analysis and interpretation of nanoindentation test data. *Surf. Coat. Technol.* **200,** 4153 (2006).
13. Y-H. Liang, Y. Arai, K. Ozasa, M. Ohashi, and E. Tsuchida: Simultaneous measurement of nanoprobe indentation force and photoluminescence of InGaAs/GaAs quantum dots and its simulation. *Physica E* **36,** 1 (2007).
14. Y.Y. Lim and M. Munawar Chaudhri: Indentation of elastic solids with a rigid Vickers pyramidal indenter. *Mech. Mater.* **38,** 1213 (2006).
15. V. Linss, N. Schwarzer, T. Chudoba, M. Karniychuk, and F. Richter: Mechanical properties of a graded B–C–N sputtered coating with varying Young's modulus: Deposition, theoretical modelling and nanoindentation. *Surf. Coat. Technol.* **195,** 287 (2005).
16. F. Richter, M. Herrmann, F. Molnar, T. Chudoba, N. Schwarzer, M. Keunecke, K. Bewilogua, X.W. Zhang, H.G. Boyen, and P. Ziemann: Substrate influence in Young's modulus determination of thin films by indentation methods: Cubic boron nitride as an example. *Surf. Coat. Technol.* **201,** 3577 (2006).
17. C. Ullner: Requirement of a robust method for the precise determination of the contact point in the depth sensing hardness test. *Measurement* **27,** 43 (2000).
18. S. Basu, A. Moseson, and M.W. Barsoum: On the determination of spherical nanoindentation stress–strain curves. *J. Mater. Res.* **21,** 2628 (2006).
19. S. Basu and M.W. Barsoum: Deformation micromechanisms of ZnO single crystals as determined from spherical nanoindentation stress–strain curves. *J. Mater. Res.* **22,** 2470 (2007).
20. S. Basu, M.W. Barsoum, A.D. Williams, and T.D. Moustakas: Spherical nanoindentation and deformation mechanisms in free-standing GaN films. *J. Appl. Phys.* **101,** 083522 (2007).
21. K.L. Johnson: *Contact Mechanics* (Cambridge, Cambridge University Press, 1985).
22. J.S. Field and M.V. Swain: A simple predictive model for spherical indentation. *J. Mater. Res.* **8,** 297 (1993).
23. D. Tabor: *Hardness of Metals* (Clarendon, Oxford, UK, 1951).
24. I.N. Sneddon: The relaxation between load and penetration in the axisymmetric Boussinesq problem for a punch of arbitrary profile. *Int. J. Eng. Sci.* **3,** 47 (1965).
25. A.J. Moseson: Spherical nanoindentation: Insights and improvements, including stress–strain curves and effective zero point determination. Master Thesis, Drexel University, 2007.