Pergamon

Acta mater. 49 (2001) 4089–4101

Acta MATERIALIA

www.elsevier.com/locate/actamat

MULTISCALE SIMULATIONS OF SILICON NANOINDENTATION

G. S. SMITH, E. B. TADMOR‡, N. BERNSTEIN§ and E. KAXIRAS†

Department of Physics and Division of Engineering and Applied Sciences, Harvard University, Oxford Street, Cambridge, MA 02138, USA

(Received 4 April 2001; received in revised form 19 June 2001; accepted 19 June 2001)

Abstract—Nanoindentation experiments are an excellent probe of micromechanical properties, but their interpretation is complicated by the multiple length scales involved. We report simulations of silicon nanoindentation, based on an extended version of the local quasicontinuum model, capable of handling complex crystal structures. This method embeds an interatomic force law within a finite element framework. We identify which features of the simulation are robust by investigating the effect of different interatomic force laws and different finite element meshes. We find that our simulations qualitatively reproduce the experimental load vs. displacement curves of indented silicon and provide information on the microscopic aspects of the phase transformations that take place during indentation. This information is linked to the macroscopic electrical resistance, providing a simple physical picture that gives a satisfactory explanation of experimental results. *© 2001 Acta Materialia Ltd. Published by Elsevier Science Ltd. All rights reserved.*

Keywords: Indentation; Theory & modeling; Semiconductors; Mechanical properties (constitutive equations); Electrical properties (electronic conductivity)

1. INTRODUCTION

Nanoindentation is an important experimental technique used to probe the mechanical properties of small volumes of a material. For example, nanoindentation is routinely used to measure the moduli and hardness of thin films [1, 2]. However, nanoindentation experiments are difficult to interpret at a fundamental level. The response of the system is very complicated due to the large stresses generated in the vicinity of the indenter, which can create defects or cracks, or cause a phase transformation to occur in the material.

There is a great deal of experimental evidence indicating that when silicon is indented a conducting phase or phases are formed underneath the indenter. In one of the earliest works, Gridneva, Milman, and Trefilov observed a change in resistance of a silicon sample during indentation [3]. Clarke *et al.* [4] found that the resistance change is reversible on subsequent loading and unloading, and Pharr *et al.* [5–7] observed extrusion near the indenter, and hysteretic load vs. displacement and load vs. electrical resistance curves. Spherical indenter studies by Weppelmann *et al.* [8, 9] and Williams *et al.* [10] observed discontinuities in the load vs. displacement curves that they attributed to sudden phase transformations. Using Berkovich indenters, Mann *et al.* [11] observed discontinuities in load vs. displacement and resistance vs. displacement curves above a critical load that were also attributed to sudden phase transformations. Finally, Kailer *et al.* [12] utilized micro-Raman spectroscopy to detect transformed phases that remain in the transformed state after full unloading, and detected several phases of silicon in the indented region. Several experimental groups also reported that indentation can create an amorphous phase, a process which appears to depend on the unloading rate and the size of indentation [4, 11, 12].

The experimental inference that a reversible phase transformation to a metallic form of silicon occurs during indentation relies on macroscopic measurements, such as load vs. displacement curves, to interpret what changes take place underneath the indenter. Probes such as Raman spectroscopy are available to identify more directly the types of phases that appear, but presently it is not possible to use them *in situ* during indentation. A number of important questions cannot be fully answered from experimental measurements alone: What phases are formed during indentation? What is the size and geometry of the transfor-

---

† To whom all correspondence should be addressed.
E-mail address: kaxiras@cmt.harvard.edu (E. Kaxiras)

‡ Present address: Faculty of Mechanical Engineering, Technion—Israel Institute of Technology, 32000 Haifa, Israel.

§ Present address: Naval Research Lab., Washington, DC 20375, USA.

PII: S1359-6454(01)00267-1

med region? How do the phase transformations give rise to the hysteresis and discontinuity in the load vs. displacement curves? How can the resistance measurements be interpreted? How does cracking affect the macroscopic measurements? Some of these open questions have motivated computational work on this interesting system. However, the large size of the sample, coupled with the very large strains in the immediate vicinity of the indenter makes the problem extremely challenging from a computational point of view.

Atomistic studies of silicon nanoindentation by Kallman *et al.* [13], were based on molecular dynamics simulations using the Stillinger–Weber interatomic potential [14]. They reported evidence for amorphization beneath the indenter, but detected no phase transformations to a metallic crystalline form. Due to the computational intensity of the simulation, these studies were restricted to a very small simulation cell with sides ~20 nm. In addition, the loading rate was on the order of 10^3 m/s, six to ten orders of magnitude faster than the experiments mentioned above. No simulated load vs. displacement or load vs. resistance curves were reported, making it difficult to compare with experiment. Pérez *et al.* [15] performed zero-temperature, quasistatic density functional theory (DFT) simulations of silicon nanoindentation. They observed the formation of interstitial defects and extrusion. The computationally intensive nature of these simulations restricted the cell to ~2 nm on a side, and necessitated the use of an atomically sharp rigid indenter. They did not attempt to directly compare their results with experiment.

At the other extreme, continuum finite element simulations using phenomenological constitutive models have been routinely used to study indentation. Bhattacharya and Nix [16], in one of the earliest works, studied nanoindentation into silicon, nickel and aluminum using a 2D axisymmetric model. The constitutive model used was that of an isotropic elastic–plastic material with linear strain hardening. More recent approaches, such as that of Knapp *et al.* [17] are fundamentally similar. The main limitation of these approaches, in contrast to the present study, is the use of the simple phenomenological models which leave out the possibility of structural phase transformations. None of these computational and theoretical approaches addressed the issue of the observed load vs. resistance behavior, while the simulated load vs. displacement curves do not capture important features of experimental results, such as hysteretic loops or discontinuities.

In this paper we apply the local quasicontinuum method, extended to handle complex crystal structures, to silicon nanoindentation. Our approach combines some of the strengths of the atomistic and continuum approaches. After briefly describing the method in Section 2, we proceed to examine a series of simulations in order to establish which features of the simulation are robust, and which are influenced by biases in the underlying force law, coarseness of the finite element mesh, or limitations due to other approximations made in the local quasicontinuum method. In Section 3 we report simulations that use a two-dimensional mesh, and in Section 4 we report simulations that use a three-dimensional mesh. Through these simulations we explore aspects of the microscopic behavior of the system, which are difficult to obtain from experiment. Finally, in Section 5, we compare our simulation results with available experimental data in an effort to understand the underlying physics of the system.

2. METHOD

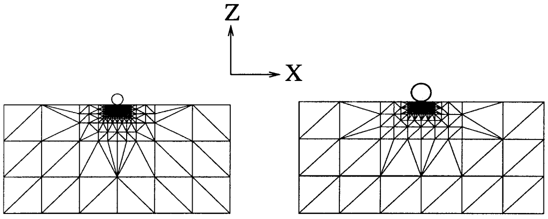
The local version of the quasicontinuum method of Tadmor, Ortiz, and Phillips [18, 19] will be the basis of simulations in the present work. In this approach a continuum finite element formulation is used to characterize the mechanical response of a given system. The difference from standard finite element methodologies is that the constitutive response of the system is obtained from an atomistic calculation rather than an empirical phenomenological rule. The basic idea is that every point in a continuum corresponds to a very large region on the atomic scale. Thus, the constitutive response at that point, i.e. the local stress–strain relation, may be obtained by deforming the underlying crystal structure by the local deformation gradient to obtain the local state of stress. This is referred to as the “local” approximation of the quasicontinuum method. This approximation will be adequate as long as the variation of the continuum fields is slow on the atomic scale. When this breaks down, for example near defects like dislocations and grain boundaries, the nonlocal limit of the quasicontinuum method must be used, where the explicit positions of the atoms are taken into account [18, 19]. Here we will focus entirely on the local version of the quasicontinuum method.

In its original version the quasicontinuum method was formulated for simple crystals, with one atom per unit cell. Many materials used in practical applications (silicon being one example) have complex crystal structures with more than one atom per primitive unit cell. Extending the quasicontinuum method to simulate such materials results in a method that maintains the advantages of traditional finite element approaches, including correct treatment of the far field boundary conditions, while at the same time it naturally incorporates the anisotropy and symmetries of the material, the nonlinearities associated with finite deformation, and the possible occurrence of uniform structural phase transformations. Finally, the efficiency of the finite element formulation permits the application of atomistic energy functionals, which are computationally intensive, to significantly larger systems than would otherwise be possible. A complete discussion of the method as extended to describe complex crystals is given in [20].

There are several limitations of the local Quasi-Continuum method extended to complex crystals (subsequently referred to as the "QCCC" method). In particular, cracks and defects are not allowed to form, and since the simulation is carried out at zero temperature, thermally-activated processes are not included. Moreover, interface energies between different phases are not taken into account.† There are several details of the implementation specific to the simulation of silicon indentation that we describe below, as appropriate.

We use two-dimensional (2D) and three-dimensional (3D) finite element meshes with constant-strain finite elements; the finite elements are triangles in the 2D mesh [shown in Fig. 1(a)], and tetrahedra in the 3D mesh [a slice is shown in Fig. 1(b)]. The meshes are finest near the indenter, and coarser farther away, and do not adapt in the course of a simulation. The boundary conditions of the finite element meshes are periodic on the sides of the mesh, free on the top surface (except where the indenter is in contact with the mesh), and fixed at the bottom. All simulations are quasistatic: the center of the indenter is incrementally lowered toward the surface of the sample, and at each indenter depth the optimal positions of the nodes are determined by minimizing the energy of the system with a conjugate gradient algorithm.

For convenience and computational efficiency we use the primitive two atom unit cell to describe the ground state diamond structure of silicon, as well as all other phases that might appear in the course of a simulation. We discuss the implications of this constraint in Section 5. For a given deformation gradient field the energy of each finite element must be minimized with respect to the relative position of the two atoms in the unit cell, as discussed in [20]. In the course of a simulation some of the finite elements become unstable, and the relative positions of the basis atoms change by a large amount. This usually indicates that a phase transformation has occurred. To verify this, and to identify unambiguously the new phase, the deformation gradient and positions of the basis atoms of the candidate structure corresponding to the finite element are used to create an infinite silicon crystal. Separate from the finite element simulation, this crystal is allowed to relax to its local minimum energy well, and the phase of the structure at this local minimum is the phase associated with the finite element.‡



(a) (b)

Fig. 1. The finite element meshes used: (a) For the 2D simulations, corresponding to an infinite cylindrical indenter; (b) For the 3D simulations, corresponding to a spherical indenter: a slice on the *xy* plane, at $y = 0$, is shown; the corresponding slice on the *yz* plane at $x = 0$ is identical.

In the indentation simulations there is no absolute length scale, so the indenter radius R_I is used to scale the results. In the 2D simulations, a cylindrical indenter is used which is rigidly in contact with the surface of the finite element mesh. In the 3D simulations we use a spherical force field to mimic a friction-free indenter [22]. We find that in three dimensions the spherical force field is numerically more stable over repeated load/unload cycles, and simpler to implement. With the indenter modelled as a force field, the energy of the system to be minimized is a sum of the strain energy of the finite element mesh, U_M, and the energy due to the indenter, U_I:

$$U_{\mathrm{Tot}} = U_M + U_I = U_M + \frac{\alpha}{2}\sum_{j=1}^{N}\left(\frac{R_I - |\mathbf{r}_I - \mathbf{r}_j|}{R_I}\right)^n \theta(R_I|\mathbf{r}_I - \mathbf{r}_j|) \tag{1}$$

where θ is the Heavyside step function, R_I is the indenter radius, N is the number of nodes in the finite element mesh, and $\mathbf{r}_I$ and $\mathbf{r}_j$ are the position vectors of the center of the indenter and the node *j*, respectively. There is no choice of the coefficient α and the exponent n that can represent exactly an elastic indenter. The indenter in equation (1) is like a sphere of independent springs, each with one end attached to the center of the indenter and the other end attached to a node located less than a distance R_I away from the center of the indenter. The springs are free to rotate around the center of the indenter at constant length, but resist with a $r^{(n-1)}$ force law when moved in the normal direction. There is no coupling between the springs, which is unphysical, since a compressed spring should make it easier to compress a neighboring spring. Since we do not seek quantitative accuracy, modelling the indenter more accurately with a finite element mesh would be a waste of computational effort.

---

† Neglecting interface energies sets a minimum length scale for the physical size of the finite elements, above which the strain energy of each finite element is much larger than the interfacial energy. Based on DFT calculations [21], we estimate this length scale to be on the order of 10 nm, two orders of magnitude smaller than the minimum finite element size of simulations scaled to typical experimental length scales.

‡ This method of determining phases assumes that metastable states of silicon that form during indentation do not become unstable when allowed to relax at zero pressure. For the silicon force laws we use here this assumption is justified, but it is not justified for all materials.

For the application to Si indentation, we make the indenter as hard as possible without adversely affecting the numerical performance of the conjugate gradient solver. With the choices $n = 4$, and $\alpha = 4\times10^9$ eV, even in the deepest simulated indentations the maximum amount nodes move into the indenter is about 2% of the indenter radius. By making the indenter very hard, our simulations ought to reproduce the Hertzian load vs. displacement curve, which we use to check our results [23].

3. SIMULATIONS USING A 2D MESH

With a 2D finite element mesh we are able to compare simulations using two different interatomic force laws. The simulation details (such as the finite element mesh, loading rate, and parameters for the conjugate gradient algorithm) are exactly the same for both force laws. We have implemented the Stillinger–Weber (SW) empirical potential [14], and the nonorthogonal tight-binding (TB) Hamiltonian of Bernstein and Kaxiras [27] as the two force laws that govern the microscopic behavior of the system. These 2D simulations are not necessarily experimentally relevant. Our principal intent here is to focus on the properties and robustness of simulation results, apart from what these results have to say about the physical system.

The TB Hamiltonian is computationally intensive. Many diagonalizations of the Hamiltonian matrix must be done in order to determine the energy of each finite element: no special *k*-points can be used for sampling reciprocal space since deformations in the course of a simulation can be arbitrary. As a result, we are restricted to the modest simulations that use the mesh shown in Fig. 1(a). The 2D mesh has 556 elements; the indenter is circular with perfect stick boundary conditions. Close to the indenter the finite element size is about 1/5 of the indenter radius, and the total size of the mesh is about 20 times the indenter radius. Nodes not constrained by boundary conditions or stuck to the indenter are free to move in the *xz* plane, but are not permitted to move in the *y* direction (i.e. plane strain conditions); simulations in which these nodes were also permitted to move in the *y*-direction showed no substantial difference.

The crystallographic $[\bar{3}12]$ direction of silicon was oriented along the *x*-axis in Fig. 1(a), and the [111] direction was directed along the *z*-axis. We explored other crystallographic orientations, but find that our conclusions do not depend on the crystallographic orientation.

Before phase transformations occur, both force laws give similar load vs. displacement curves, as shown in Fig. 2. Since there is no absolute length scale in the simulations, the results are scaled to the indenter radius R_I. In the 2D simulations, the onset of phase transformations is marked by discontinuities in the load vs. displacement curves. The phase transformations begin at the same indentation depth and load. Snapshots of the mesh, after high pressure phases have begun forming, are shown in Fig. 3. Figures 3(a) and (b) show the shear strain, ϵ_{xz} for the two force laws. The two plots are qualitatively similar, both showing two concentrations of shear strain on either side of the *z* axis beneath the surface. For both force laws, these concentrations occur at the edges of the phase transformed region. Figures 3(c) and (d) show the finite elements that have transformed. The atomic structure of these phases is shown in Fig. 4.

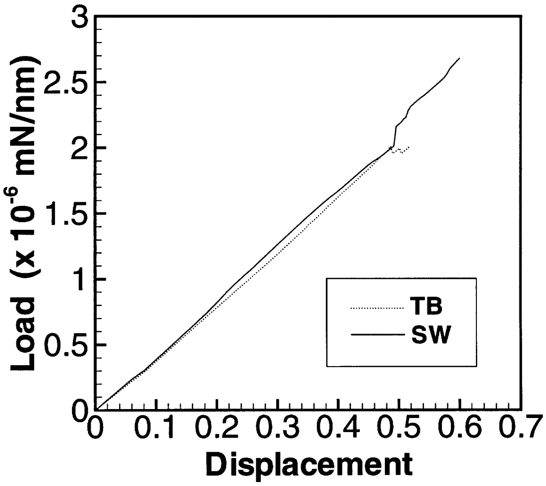


Fig. 2. Comparison of the load vs. displacement curves obtained from the 2D simulations using the Stillinger–Weber (SW) potential and a tight-binding (TB) Hamiltonian. The initial phase transformations occur where the discontinuities appear. Both the displacement and the load are scaled by the indenter radius R_I.

An obvious difference between the two simulations is that the phases underneath the indenter are different. The energy vs. volume curves for selected phases comparing the TB Hamiltonian and the SW interatomic potential are shown in Fig. 5. For most of the phases, the positions of the minima and the curvature are different for the TB and SW force laws, indicating that the energy surface of the two-atom unit cell is also different. Given the inhomogeneous stress fields due to the indenter and the elastic coupling between finite elements, it is not surprising that the differences in the energy surfaces of the two force laws result in finite elements moving into different energy wells in the course of a simulation.

In the simulation that uses the SW potential all the transformed finite elements are in the bct5 well, which according to Fig. 5 is the well closest in energy and volume to the diamond phase. However, for the simulations using the TB Hamiltonian the behavior is more complicated. In these simulations finite elements have fallen into the body-centered cubic well and the well that corresponds to the bct5* structure (see Fig. 4). Both these wells are higher in energy than other available phases, such as β-Sn and bct5. This suggests that the phases that form in the course

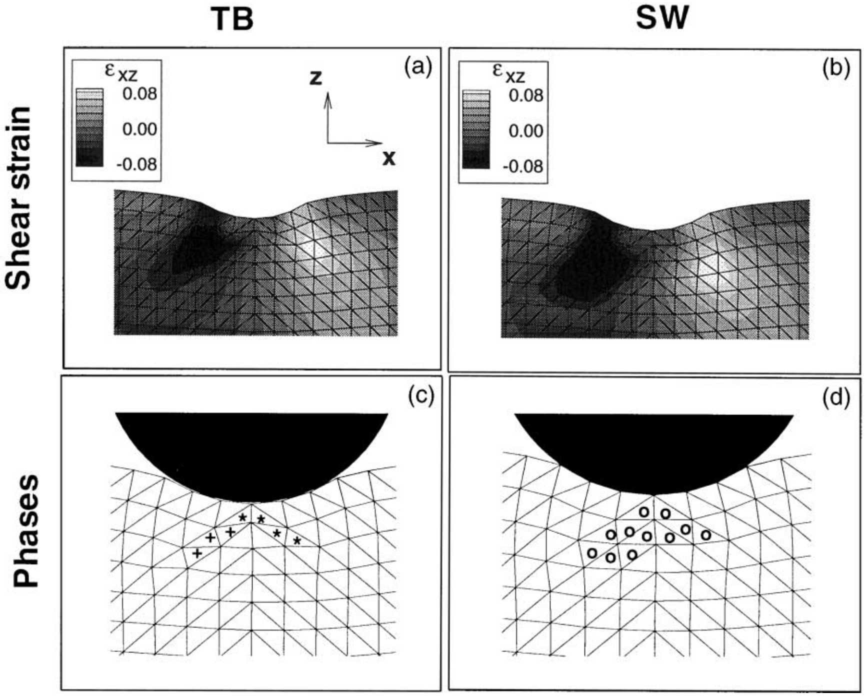


Fig. 3. Comparison of snapshots of the 2D finite element mesh after phase transformations have begun. Results using the Stillinger–Weber (SW) potential and a tight-binding (TB) Hamiltonian are shown. The figures show the shear strain ϵ_{xz} ((a) and (b)), and the phases that have formed ((c) and (d)). In (c) and (d), the symbol ○ indicates the bct5 phase, the symbol $*$ indicates the bcc phase, and the symbol + indicates the bct5* phase (see Fig. 4). Blank finite elements remain in the diamond phase. The indenter is shown as the black regions in (c) and (d).

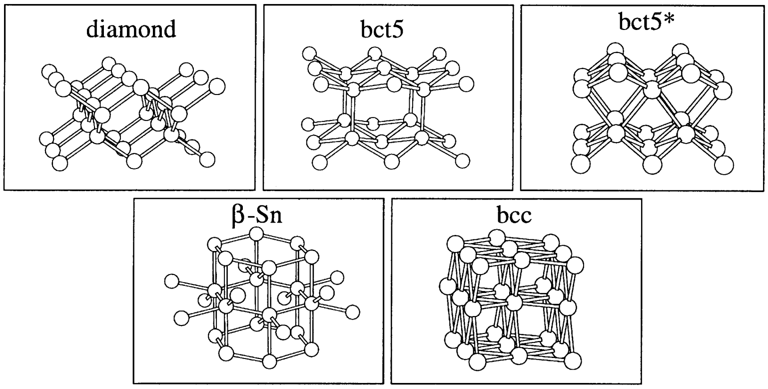


Fig. 4. The structures of some of the phases that form in the course of the simulations of indentation: the equilibrium diamond crystal, a body-centered structure with 5-fold coordination (bct5), a related structure that involves a translation of every other bilayer by half a period in the lateral direction (bct5*), the β-Sn phase with 6-fold coordination and the bcc phase with 8-fold coordination.

of a simulation do not depend simply on the energy or density of the available phases. The interplay between the stress field due to the indenter and the details of the TB energy surface result in finite elements transforming into high energy phases instead of the low-lying phases bct5 or β-Sn. Since the finite elements of the TB simulation transform into denser phases than those in the SW simulation, when transformations begin to occur the macroscopic load suddenly takes different values, as shown in Fig. 2.

Finally, note that from Fig. 5, both the TB and SW curves differ from those calculated through DFT, which has been shown to give quantitative accuracy for a variety of physical systems [28, 29]. In general the SW and TB transformed phases are too low in energy relative to the equilibrium diamond phase. The positions of the minima in the curves in Fig. 5 indi-

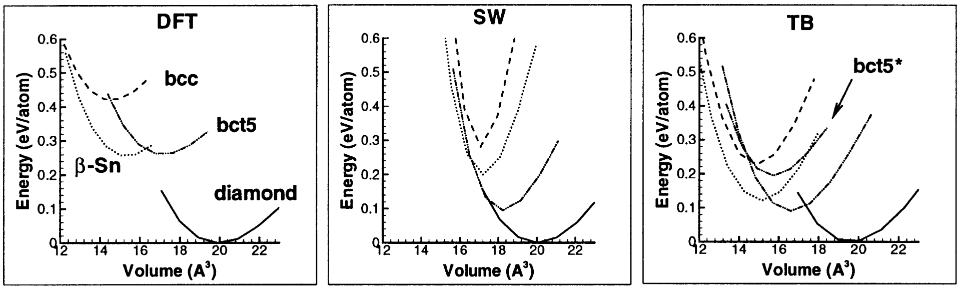


Fig. 5. Comparison of the energy vs. volume curves for selected phases of silicon according to density functional theory (DFT), the Stillinger–Weber (SW) empirical potential, and the tight-binding (TB) Hamiltonian. For each of the energy functionals, the zero of energy is set by the equilibrium diamond structure. The different phases are identified by curves of a given type, as defined in the DFT plot.

cate that the densities of the various phases according to the TB Hamiltonian are more accurate than the densities according to the SW potential. However, the TB force law gives bct5* as a stable phase, while it is unstable according to DFT results.

Despite the differences between the SW and TB force laws in the energy surface, the macroscopic load vs. displacement curve and onset of the phase transformation are similar in the two simulations as indicated in Fig. 2. There are also similarities on the microscopic level. For example, Fig. 3 indicates that the transformed regions have the same qualitative shape. The anisotropy of the crystal affects the transformed region in the same way, favoring transformations in the negative *x* half-space. Also, in both simulations there is a tendency for the transformed phases to form a cluster of like phases. The fact that simulations using these two very different interatomic force laws produce a transformed region whose shape and clustering properties are similar, suggests that these features are robust results of the simulations. Further studies using 3D meshes confirm this.

As a final point, it is interesting to note that according to Fig. 5, the TB Hamiltonian is not obviously more accurate than the SW potential. The SW potential makes no attempt to model the electrons explicitly, but instead relies on an intuitive understanding of the bonding properties of solid silicon, which motivates a simple functional form. The TB Hamiltonian does model the electrons explicitly, which is expected to be more accurate, or at least more transferable. The TB Hamiltonian does give superior results when compared to the SW potential for a number of important properties of silicon, particularly those related to defects and interfaces [27]. However, in the course of a simulation the QCCC method samples a wide swath of the configuration space of the two-atom unit cell, and apparently for this type of sampling neither of the two force laws is obviously superior. In the 2D indentation simulations, the SW potential favors the bct5 phase, but the TB Hamiltonian produces a phase that is not even stable according to DFT. This, coupled with the fact that the SW potential is significantly faster than the TB Hamiltonian, motivates us to use the former force law exclusively for the rest of the present study.

4. SIMULATIONS WITH A 3D MESH

In this section we explore some preliminary results of simulations using a far more realistic and computationally intensive 3D finite element mesh. We discuss typical load vs. displacement curves, and the evolution of the phase transformed region in the course of loading and unloading. Where justified, we make observations about the simulations that may relate to the behavior of the physical system, in an attempt to provide information otherwise difficult to obtain from experiments. Moreover, we systematically explore how mesh spacing and orientation affect the results.

If we make the finite element mesh finer, or rotate it with respect to the underlying crystal, the physics of the system does not change, so we would expect the simulation results to remain the same. However, the constraints and modification of the rotational and translational symmetry of the system introduced by the finite element mesh affect the simulation, and it is crucial that we determine the extent of this effect. Three mesh spacings were used and their results were compared: in the least fine mesh, the dimensions of the smallest finite element are 1/5 the indenter radius; in the medium fine mesh the smallest finite element size is 1/9 the indenter radius (this finite element mesh is shown in Fig. 1(b)); in the very fine mesh the smallest finite element size is 1/13 the indenter radius. For these simulations, the [111] crystallographic direction of silicon points along the *z*-axis. Finally, in most of the simulations, the $[\bar{2}11]$ direction was pointed along the *x*-axis, but in one simulation, the $[\bar{5}41]$ direction was pointed along the *x*-axis to test how orientation of the mesh affects results; results from this orientation of the mesh showed minor and insignificant changes relative to the other orientation, so we will not discuss this aspect further.

Figure 6 shows a load vs. displacement curve from a simulation that uses the least fine mesh, for a deep indentation with the maximum depth about 1/2 the

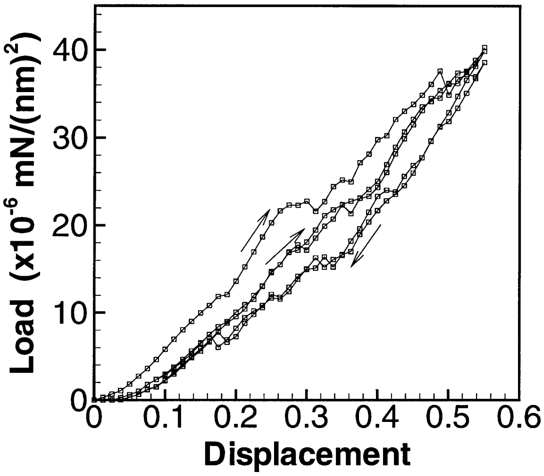


Fig. 6. The load vs. displacement curve from a 3D simulation using the least fine mesh, showing a reproducible hysteretic loop. The displacement is scaled by the indenter radius, R_I, and the load is scaled by R_I^2.

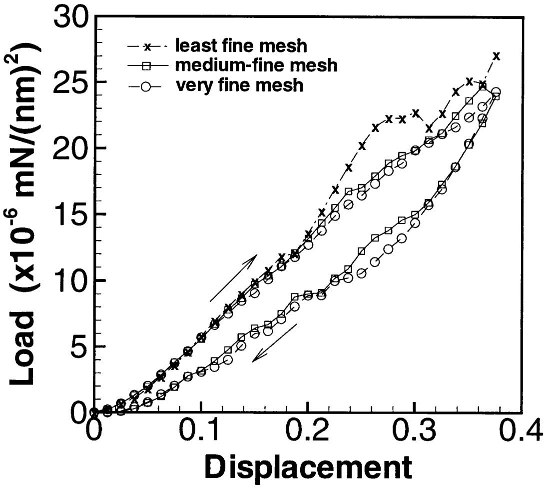


Fig. 7. A comparison of load vs. displacement curves from 3D simulations using several meshes. The displacement is scaled by the indenter radius, R_I, and the load is scaled by R_I^2.

indenter radius. Since there is no absolute length scale in the simulations, the displacement is scaled by the indenter radius R_I and the load by R_I^2. The most striking feature of this curve is the reproducible hysteresis loop, a feature that has been observed in some experiments [5, 6]. The hysteretic behavior in the simulation curve in Fig. 6 is due to the formation and disappearance of high-coordination silicon phases beneath the indenter, the same mechanism presumed to cause hysteresis in experiment [5].

Even using this relatively coarse mesh, our simulations are able to capture reversible phase transformations underneath the indenter and a reproducible hysteresis loop, which was not the case for previous computational models [13, 15–17]. However, some qualitative features of the simulation curve are quite different from experiment. For example, the simulation curve is very jagged, while experimental curves tend to be mostly smooth. One possible explanation has to do with the finite size of the finite elements, which sets the minimum volume over which a transformation can occur; the larger the volume, the larger the perturbation of the system when the transformation occurs. This consideration motivated us to compare load vs. displacement curves for a range of mesh spacings, which are shown in Fig. 7. The curve from the coarsest mesh is slightly different from the others, suggesting that the coarse mesh may not be fully converged with respect to mesh spacing. All subsequent discussion will not include results from simulations using this mesh. The other two load vs. displacement curves are significantly smoother, and both exhibit the same basic features and magnitude. For a sufficiently fine mesh the macroscopic load vs. displacement curve is fairly robust, unaffected by perturbations introduced by modifying the mesh or the force law used.

Without exception, in all the simulations several phases form beneath the indenter. These silicon phases include bct5, β-Sn, simple cubic, and body-centered cubic. The structures of some of these phases are shown in Fig. 4. In addition, several non-physical phases appear that are stable within the SW description of silicon, but are not stable according to DFT results. In the indentation simulations, the only non-physical phase that appears in any appreciable amount is a layered hexagonal structure. Generally, bct5, followed by β-Sn and the non-physical hexagonal structure are the most prevalent phases that form upon loading, while bct5 and simple cubic are the most prevalent upon unloading. All of the non-diamond phases are metallic, according to DFT calculations.

Physically, it is reasonable that many phases, not just a single transformed phase, would form under the indenter. The energy surface of the two-atom silicon unit cell is very complicated, with a large number of local minima corresponding to different phases. The indenter imposes large, rapidly varying strain fields in its vicinity which, in conjunction with the strong elastic coupling between finite elements, spread the finite elements in the system over an extended region of configurational space of the silicon unit cell, sampling a variety of different local minima. It is possible, however, that the formation of cracks or defects, or the presence of an interface energy penalty, would modify this behavior.

Of particular interest are the details of the distribution of various phases. To investigate this, we examine the evolution of the number of finite elements of various phases as a function of indentation depth and we track the size and number of clusters during an indentation cycle. A “cluster” is a group of contiguous finite elements of the same phase, each of which is connected to another element of the group by at least one common finite element face. Fig. 8 compares the amount of the bct5 and β-Sn phases of silicon in the course of a simulation for the medium fine and very fine mesh, for indentation reaching a depth of 3/8 of the indenter radius. The basic shape and scale of the curves, the load at the initial transformation, and the final unloaded values are essentially independent of the mesh spacing. Also consistent are basic features such as the saturation of the number of bct5 elements upon unloading and the steep increase in the number of β-Sn elements upon loading.† The phases that are present upon complete unloading are responsible for the hysteresis and residual deformation evident in the load vs. displacement curves.

We also analyzed the distribution of clusters of the bct5, β-Sn and simple cubic phases for the two meshes when the system is maximally loaded, and for when it is subsequently completely unloaded. For all the simulations the majority of the volume of the transformed phases is in a few very large clusters. The number and volume of the larger clusters of the various phases is similar for both meshes at both the fully loaded and fully unloaded state. However, the distribution of the very small clusters is different for the very fine mesh compared to the medium fine mesh. The finer the mesh becomes, the more clusters appear that are one or two finite elements in size, particularly in the fully unloaded state. In the fully unloaded state, these very small clusters make up about 10% of the mass of the transformed volume. The fact that they become numerous as the mesh becomes finer indicates that the microscopic results of the simulation are not converged with respect to mesh spacing. No matter how fine the mesh becomes, microscopic convergence will not be achieved given the nonconvexity of the total energy of the system and the lack of interfacial energy.

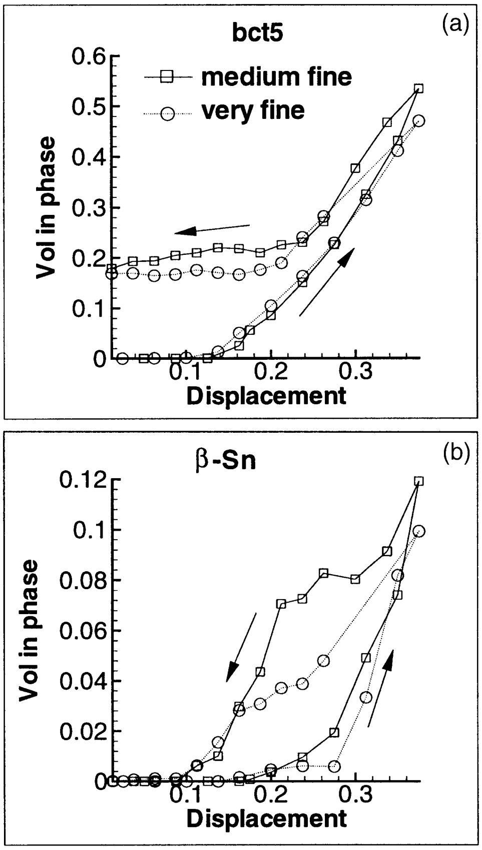


Fig. 8. A comparison of the development of the amount of the bct5 (a) and β-Sn (b) phases during 3D simulations of indentation. The displacement is scaled by the indenter radius, R_I, and the volume of each phase is scaled by R_I^3. Two different meshes are used: a medium fine mesh and very fine mesh (see text for details). Note the different scales of the vertical axis corresponding to the different phases.

Though there are quantitative differences in evolution of the number of finite elements of each phase and their clustering properties, these differences do not significantly affect the macroscopic results. Referring back to Fig. 7, the macroscopic measures for the medium and very fine grids are very similar. In particular, the loads at the maximum indentation depth are very close, even though the exact amount of the various phases and the details of the clustering distribution are different. This implies that the energy surface of the system as a whole is degenerate: at a given indentation depth, there are a number of ways to arrange the various phases underneath the indenter to give similar macroscopic measures. The perturbations introduced by modifying the mesh direct the system toward a different region of the degenerate phase space. This degeneracy is not surprising, as it is related to the complicated nature of the microscopic energy surface and the strong elastic coupling between regions in the sample.

In addition to macroscopic measures and some clustering properties of the transformed region, other features of the simulation appear to be independent of the mesh orientation and spacing. For example, upon loading, a roughly hemispherical transformed region forms, whose geometry is influenced by the three-fold symmetry of the underlying crystal: Fig. 9 shows a cross section of the finite element mesh, a cut in the xy-plane near the indenter toward the bottom of the transformed region. Each dot corresponds to a finite element. The black circles correspond to finite elements that remain in the diamond phase, and the light grey circles correspond to finite elements that have transformed to a high-coordination phase. Also in Fig. 9, the orientation of the crystalline lattice is shown. Clearly, the approximate three-fold symmetry of the transformed region corresponds to the positions of the three tetrahedral atoms. This is not a mesh

---

† The next phase that appears during indentation, the simple cubic structure, occurs in much fewer finite elements (one order of magnitude less than the β-Sn phase at maximum load) and its concentration shows some variation, but is still overall qualitatively the same in the two different meshes.

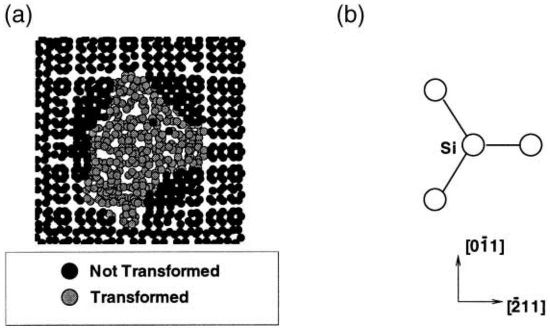


Fig. 9. A slice in the *xy* plane through the 3D finite element mesh near the bottom of the transformed region: each dot corresponds to a finite element, with dark finite elements in the diamond phase, and lighter shaded finite elements in high-coordination metallic phases. (b) The underlying orientation of the crystalline structure before indentation.

effect, because if the mesh is made finer, or if it is rotated about the [111] direction, the effect persists.

The geometry of a transformed region also changes dramatically during a load/unload cycle. Upon loading, the roughly hemispherical cluster grows underneath the indenter, reaching a radius approximately equal to R_I for indentation to a depth $0.5R_I$. Upon unloading, the well-connected, roughly hemispherical cluster fragments, and the largest remaining cluster becomes stringy and tenuously connected as shown in Fig. 10. This figure compares two slices in the *xy* plane, one from a fully loaded sample, the other from a fully unloaded sample. Again, a dot corresponds to a finite element. Finite elements that are not part of the largest connected transformed cluster are represented with black circles. The other finite elements are constituents in the largest transformed cluster. Those represented with grey circles are fully connected to the cluster, in the sense that they share all four of their faces with other finite elements in the cluster. Those represented with white circles are not fully connected to the cluster; they share one, two, or three faces with other finite elements in the cluster. It is clear that the cluster at full unloading is much more tenuously connected than that of the fully loaded system. It appears that these tenuous connections are not fully converged with respect to mesh spacing, according to examinations of the cluster geometry for different mesh spacings.

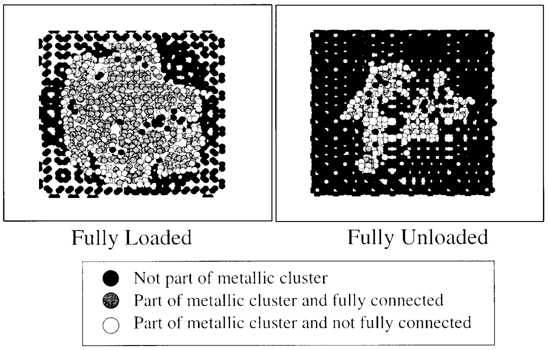


Fig. 10. A comparison of the geometry of the largest metallic cluster in a fully loaded and fully unloaded system. Two slices through the 3D mesh, like the one in Fig. 9, are shown. Finite elements that are fully connected share all four of their faces with other finite elements in the metallic cluster. Finite elements that are not fully connected share between one and three of their faces with other finite elements in the metallic cluster.

5. COMPARISON WITH EXPERIMENT

Direct comparison of our simulation results with available experimental data is a crucial aspect of our study. Fig. 11 compares a simulated load vs. displacement curve with an experiment of Weppelmann *et al.* [8]. To facilitate comparison, the simulation is scaled to the same indenter radius used in the experiment, 8.5 μm. These are indentations with displacements smaller than in the simulations discussed in the previous section (see Fig. 7, for example). Here, in both simulation and experiment, unloading begins soon after transformed phases first appear underneath the indenter.

Analyzing the elastic behavior at low loads is a useful check on the simulation. The solid lines in Fig. 11(a) and (b) are fits to the elastic solution of Hertz, $P = 4(E^*)(R_I)^{1/2}\delta^{3/2}/3$, where P is the load, R_I is the indenter radius, and δ is the indentation depth [23, 24]. The constant E^* can be extracted from the elastic fits in Fig. 11: from the simulation curve $E* = 130$ GPa, while from experiment $E* = 160$ GPa. This deviation is reasonable since the SW elastic constants are about 20% off from the experimental values [25], and since the indenters in the simulation and experiment behave somewhat differently under load.

As an additional check on the simulation the quantity E^* can be expressed in terms of the Young's moduli, E_s, E_I, and the Poisson ratios, ν_s, ν_I of the sample and the indenter [24]:

$$\frac{1}{E*} = \frac{1-\nu_s^2}{E_s} + \frac{1-\nu_I^2}{E_I}. \tag{2}$$

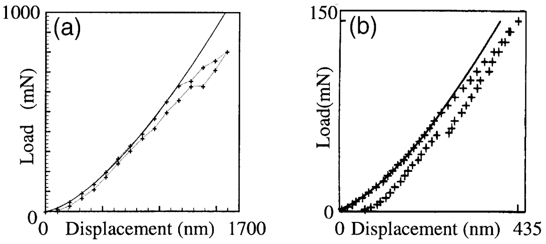


Fig. 11. A comparison of the load vs. displacement curves for indentation on the (100) surface of Si: (a) 3D simulation; (b) experiment [8]. The solid line in each case is a fit to the elastic solution. The simulation is scaled to the same indenter radius used in the experiment, 8.5 μm.

The SW silicon elastic constants [25] can be used to obtain E_s and ν_s in equation (2) [26], but determining the contribution from the indenter is difficult since it is not an elastic body. However, the indenter in the simulation is quite hard compared to silicon, so we can ignore the second term of equation (2) to obtain a value of $E* \approx 150$ GPa, which is in reasonable agreement with the value obtained from the fit.

In Fig. 11(a) and (b) deviation from elastic behavior occurs as indentation proceeds, as phase transformations begin to occur underneath the indenter. Both curves exhibit a step in the unloading part, followed by a smooth decline to an unloaded state with residual plastic deformation. In the simulation, this residual deformation is due to finite elements that remain in a high-coordination phase.

There are some differences in the behavior of the simulated and experimental systems. When the simulation is scaled to the experimental indenter radius the initial transformation occurs at a displacement of 1200 nm, significantly deeper than in experiment. One reason for the difference is that transformations must occur homogeneously in the simulation, as opposed to a nucleation and growth process in the experiments. Another cause for the late transformation in the simulation is that the smallest unit that can transform is a single finite element. Scaled to the experimental length scale, the size of a finite element is roughly 1 μm. In the experiment, transformations can occur on a length scale determined by a balance between volume energy gain and interface energy cost. This length scale is on the order of 10 nm from DFT calculations of interfaces between diamond and β-Sn phases of silicon [21].

Another difference between simulation and experiment is the nature of phases that form underneath the indenter. In experiments the phase deduced to form during loading is β-Sn [8, 9, 12]. During slow unloading, several experimental groups have concluded that the bc8 phase forms [8, 9, 12], while there is evidence that the r8 and hexagonal diamond phases also form [12]. In the simulation the observed phases are bct5 and the simple cubic phase, which forms in small amounts toward the end of the unloading cycle. These discrepancies between simulation and experiment perhaps can be attributed to shortcomings in the SW potential, and to the restriction to a two-atom unit cell in the simulations (the latter restriction can be relaxed by using a larger unit cell). However, it is also possible that bct5 does form during indentation experiments, but has not been detected because an *in situ* probe is not currently available.

It is interesting that despite these differences, the simulation faithfully reproduces the qualitative features of the experimental load vs. displacement curve, including the step upon unloading. In interpreting their experimental results, Weppelmann *et al.* assumed that at this step in the unloading curve one phase transforms completely into another [8, 9]. In contrast, in the simulation the size and composition of the transformed region changes throughout the unloading part of the simulation [30]. Later we will return to this issue and discuss how the two scenarios could be distinguished in experiments.

In addition to load vs. displacement curves, resistance measurements have also been used to gain insight into the behavior of indented silicon. We will focus on two types of resistance experiments that have been performed by Pharr *et al.* [6], shown schematically in Fig. 12. In both, two electrodes are deposited on the surface of the silicon sample, and a voltage difference is applied between them. In one experiment, the indenter is directly on top of one of the electrodes; in the other, the indenter is between the two electrodes, slightly offset towards the negative one. As metallic phases appear underneath the indenter, the resistance between a pair of electrodes on the surface of the sample can change. The connection between the appearance of metallic phases and the change in the macroscopic resistance is very complicated, but we have developed a simple model that captures the basic physics and enables us to connect our simulation results to experimental measurements.

In the experiments of Pharr *et al.*, indentation and the resultant metalization underneath the indenter reduces the resistance from the background (R_b) to the measured (R_m) value. Before indentation, the resistance at each of the electrodes is determined by the Schottky barrier that forms at the metal-semiconductor interface [6], and the spreading resistance [31]. Physically, the spreading resistance is due to the current being forced through a bottleneck just beyond the electrode, before the current can spread through a larger region of the material further away from the electrode. The Schottky barrier and spreading resistance are local to each electrode, so we consider the resistances of the positive and negative electrodes to add in series: $R_b = R_b^{(+)} + R_b^{(-)}$.

As indentation proceeds, clusters of metallic-phase elements form underneath the indenter. When a cluster touches an electrode, current flows through the cluster and changes the measured resistance. We take the resistance of a transformed region, R_t, to be in parallel with the electrode's background resistance. This is a reasonable approximation. Current that completes the circuit via the background network will not be significantly perturbed by the indented region, whose dimensions are extremely small compared to the size of the electrode network (Fig. 2 in Ref. [6] indicates that the electrode network is about two orders of magnitude larger than the indenter); conversely, the resistance experienced by the current that passes through the indentation region will depend primarily on the local geometry of the transformed region. The circuit used to model the resistance measurements is shown in Fig. 13.

The metallic clusters that form underneath the indenter upon loading are roughly hemispherical. Neglecting the influence of the electrode network, by dimensional analysis the spreading resistance of a

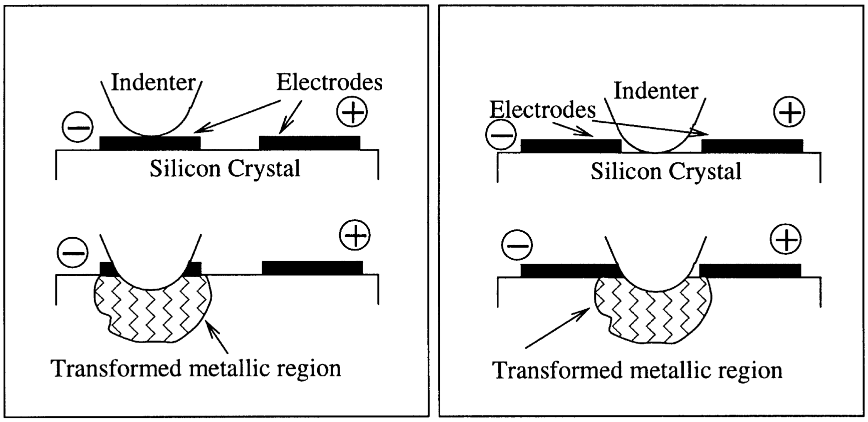


Fig. 12. Schematic of experimental setup for indenting on and between electrodes.

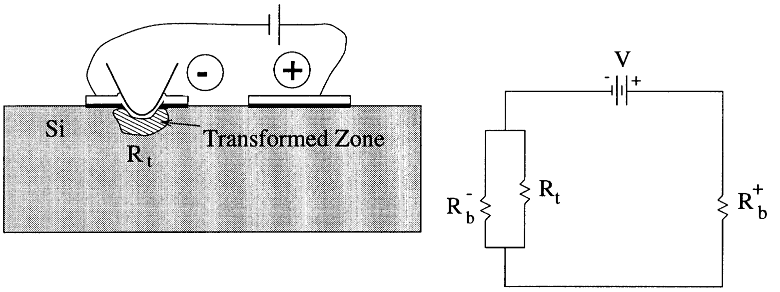


Fig. 13. A schematic of an indentation experiment and the corresponding circuit used to model the resistance measurements.

metal hemisphere of radius a embedded in the surface of a semi-infinite sample of silicon is proportional to the resistivity of silicon, ρ, and inversely proportional to a. By a simple electrostatic argument the constant of proportionality is $1/2\pi$ [31], giving the estimate of $R_t \approx \rho/2\pi a$. With these approximations we arrive at

$$R_m = \frac{\rho R_b^{(-)}}{2\pi a^{(-)} R_b^{(-)} + \rho} + \frac{\rho R_b^{(+)}}{2\pi a^{(+)} R_b^{(+)} + \rho} \tag{3}$$

where $a^{(-)}$ and $a^{(+)}$ are the size of the clusters of metallic-phase elements in contact with the negative and positive electrode, respectively. Occasionally, more than one cluster will be in contact with an electrode, and in those cases we use the largest. As our definition of what constitutes a conducting cluster, we assume that only metallic elements fully surrounded by other metallic elements conduct. If a cluster touches both electrodes simultaneously, then the circuit is shorted and $R_m=0$.

Within this model, at each electrode there is a crossover size for a transformed cluster: $a_c^{(\pm)} = \rho/2\pi R_b^{(\pm)}$. So when there is not a short circuit, we can write

$$R_m = R_b^{(-)}\left[1 + \frac{a^{(-)}}{a_c^{(-)}}\right]^{-1} + R_b^{(+)}\left[1 + \frac{a^{(+)}}{a_c^{(+)}}\right]^{-1}. \tag{4}$$

This equation suggests that in order to optimally track the evolution of cluster size in spherical indenter experiments, the indenter radius R_I should be on the order of $a_c^{(-)}$ or $a_c^{(+)}$, depending on which electrode is being indented. In the following we choose the $(-)$ as the one being indented. Our simulations indicate that once a transformed region begins to form during indentation, its size a is on the order of the indenter radius R_I. If $R_I \ll a_c^{(-)}$, then the cluster size $a^{(-)} \ll a_c^{(-)}$, and in this limit equation (4) indicates that the measured resistance will not change from the background value during indentation. On the other hand, if $R_I \gg a_c^{(-)}$, then after initial formation the cluster can rapidly grow to size $a^{(-)} \gg a_c^{(-)}$, abruptly lowering the resistance to a value much smaller than $R_b^{(-)}$ [30].

Experimental resistance measurements corresponding to the load vs. displacement curve in Fig. 11 are not available. Such measurements could determine if the reverse transformation during unloading occurs only at the step in the load vs. displacement curve, as the experimentalists assumed [8, 9], or throughout the unloading process, as simulations suggest [30].

Pharr *et al.* [6] produced resistance measurements using a Berkovich indenter. Though the simulations use a spherical indenter, the simulated system behaves similarly to the low-load indentations of Pharr *et al.*: both show reproducible hysteretic loops

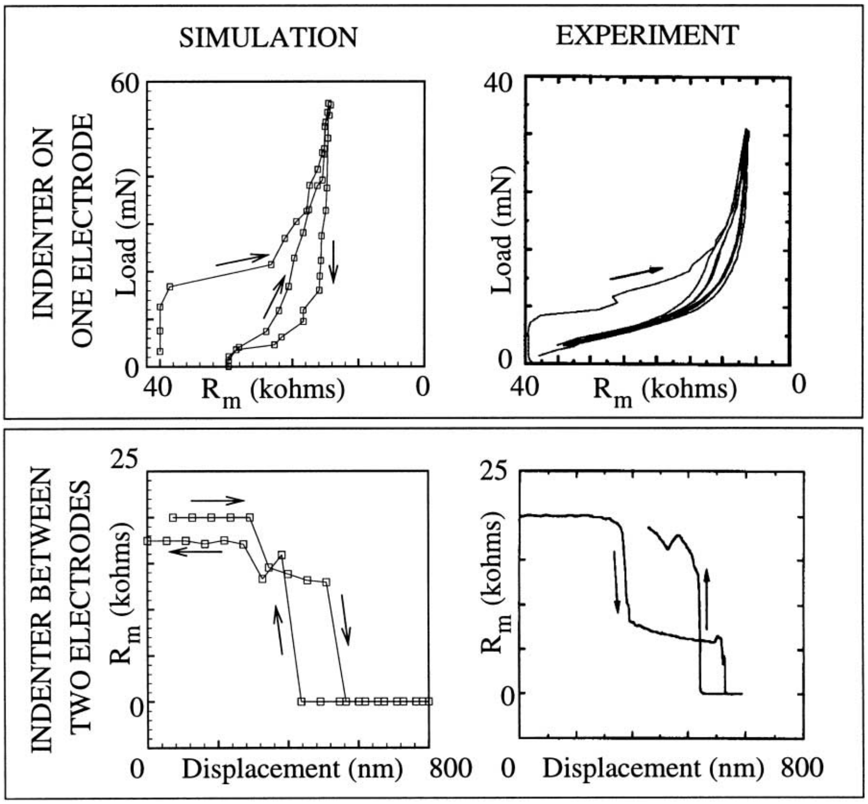


Fig. 14. Comparison of experimental and simulated resistance curves for the two different experimental geometries shown in Fig. 12. The experimental curves are taken from [6].

in the load vs. displacement curve (compare Fig. 5 in [5] to Fig. 6 in this work), and in both cracking is absent. This encourages us to compare our simulated resistance measurements with the experimental data of [6].

In our model we use the experimentally determined values for the constants in equation (4): $\rho \approx 10\Omega$ cm, $R_b^{(+)} \approx 5$ kΩ. The total background resistance R_b varies from one experiment to the next; $R_b^{(-)}$ is chosen to ensure that $R_b = R_b^{(+)} + R_b^{(-)}$ is equal to the experimentally measured total background resistance. With these parameters, $a_c \approx 1$ µm, which is the approximate size of the indentations in the experiments.

To obtain resistance vs. load curves from our simulations, the size and positions of the transformed clusters were tracked as a function of indentation depth in a 3D simulation using the medium fine mesh. At each depth, the largest cluster in contact with each electrode was identified to determine the quantities $a^{(\pm)}$.† This data, along with the empirically determined values for the constants in equation (4), was used to obtain the simulated resistance curves shown in Fig. 14. The simulation data was scaled to a maximum indentation depth that matches that of the corresponding experimental curves, also shown in Fig. 14. The simulated curves reproduce well the qualitative features of the experimental results. Given the shortcomings of the SW potential used in the present calculations, this qualitative agreement suggests that the basic features of the resistance measurements are insensitive to the details of the microscopic energy surface and the indenter geometry, assuming there is no cracking in the sample.

6. CONCLUSIONS

In this work we have used two different interatomic force laws and several finite element meshes within the quasicontinuum model to study indentation of silicon. The simulations yield both macroscopic measures such as the load vs. displacement and load vs. resistance curves, as well as detailed microscopic information on the morphology of the transformed region underneath the indenter. From our analysis of the simulations we conclude that we can trust the macroscopic curves and the overall geometry of the metallic cluster that forms underneath the indenter, but not the specific types of phases that form during indentation. It is important to point out that the computational method presented here is not yet predictive

---

† From experiments it is not clear where precisely the electrodes should be placed on the sample. Our results indicate that the qualitative features of the simulated resistance measurements do not depend on the exact location of the electrodes, as long as the electrodes are not so far from the indenter that no clusters are in contact with them.

and is not intended to replace experimental investigations. Rather the intent of our simulations is to help shed some light on the physical processes underlying the experimental observations in this complicated system by having direct access to microscopic information and making careful comparisons with experiment.

With these limitations in mind, we were able to provide the following insights into the microscopic behavior of silicon during indentation: (1) For deep indentation, many metallic phases form beneath the indenter, rather than one single phase. (2) The threefold symmetry of the silicon unit cell about the (111) crystallographic direction influences the geometry of the transformed region. (3) A roughly hemispherical transformed region forms beneath the indenter during loading which fragments upon unloading. (4) The energy landscape available to the system upon deep indentation appears to be highly degenerate. The extent to which these conclusions are affected by interface energies or physical processes such as defect or crack formation, which are left out of the current simulation, is still open to investigation.

The above observations at the microscopic level, cannot be experimentally verified at this point. To lend support to the results, we compare the macroscopic predictions of the model with experimental evidence. Comparison of the load vs. displacement curves is straightforward and at low loads shows a qualitative agreement: the simulated curve reproduces well the distinctive step upon unloading. For deep indentation, the simulated curve exhibits hysteresis similar to that observed in low-load Berkovich indentation experiments. More work is needed to understand how factors such as indenter shape and cracking in the indented region influence whether macroscopic curves exhibit a step or a reproducible hysteretic loop.

In order to make contact with load vs. resistance measurements, we developed a simple model for the electrical resistance of indented silicon based on the size of the transformed region. When used in conjunction with the simulation data we find that the model qualitatively reproduces the available experimental data. In the discussion of the model we point out how spherical indenter experiments that measure electrical resistance could be used to illuminate the nature of the reverse transformation that occurs during unloading.

Acknowledgements—It is a pleasure to acknowledge useful discussions with Michael J. Aziz, Michael Ortiz, Jim Sethna, Frans A. Spaepen, and Umesh V. Waghmare. This work was supported by Harvard's Materials Research Science and Engineering Center, which is funded by the NSF.

REFERENCES

1. Xiadong, L. and Bhushan, B., *Thin Solid Films*, 1999, **340**, 210.
2. Oliver, W. C. and McHargue, C. J., *Thin Solid Films*, 1988, **161**, 117.
3. Gridneva, I. V., Milman, Y. V. and Trefilov, V. I., *Phys. Stat. Sol. A*, 1982, **14**, 177.
4. Clarke, D. R., Kroll, M. C., Kircher, P. D. and Cook, R. F., *Phys. Rev. Lett.*, 1988, **60**, 2156.
5. Pharr, G. M., Oliver, W. C. and Clarke, D. R., *J. Elec. Mater.*, 1990, **19**, 881.
6. Pharr, G. M., Oliver, W. C., Cook, R. F., Kircher, P. D., Kroll, M. C., Dinger, T. R. and Clarke, D. R., *J. Mater. Res.*, 1992, **7**, 961.
7. Pharr, G. M., Oliver, W. C. and Harding, D. S., *J. Mater. Res.*, 1991, **6**, 1129.
8. Weppelmann, E. R., Field, J. S. and Swain, M. V., *J. Mater. Res.*, 1993, **8**, 830.
9. Weppelmann, E. R., Field, J. S. and Swain, M. V., *J. Mater. Sci.*, 1995, **30**, 2455.
10. Williams, J. S., Chen, Y., Wong-Lueng, J., Kerr, A. and Swain, M. V., *J. Mater. Res.*, 1999, **14**, 2338.
11. Mann, A. B., van Heerden, D., Pethica, J. B. and Weihs, T. P., *J. Mater. Res.*, 2000, **15**, 1754.
12. Kailer, A., Gogotsi, Y. G. and Nickel, K. G., *J. Appl. Phys.*, 1997, **81**, 3057.
13. Kallmann, J. S., Hoover, W. G., Hoover, C. G., De Groot, A. J., Lee, S. M. and Wooten, F., *Phys. Rev. B*, 1993, **47**, 7705.
14. Stillinger, F. H. and Weber, T. A., *Phys. Rev. B*, 1985, **31**, 5262.
15. Pérez, R., Payne, M. C. and Simpson, A. D., *Phys. Rev. Lett.*, 1995, **75**, 4748.
16. Bhattacharya, A. K. and Nix, W. D., *Int. J. Solids Struct.*, 1988, **24**, 881.
17. Knapp, J. A., Follstaedt, D. M., Myers, S. M., Barbour, J. C. and Friedmann, T. A., *J. Appl. Phys.*, 1999, **85**, 1460.
18. Tadmor, E. B., Ortiz, M. and Phillips, R., *Phil. Mag. A*, 1996, **73**, 1529.
19. Shenoy, V. B., Miller, R., Tadmor, E. B., Rodney, D., Phillips, R. and Ortiz, M., *J. Mech. Phys. Sol.*, 1999, **47**, 611.
20. Tadmor, E. B., Smith, G. S., Bernstein, N. and Kaxiras, E., *Phys. Rev. B*, 1999, **59**, 235.
21. Tadmor, E. B. and Waghmare, U. V., unpublished.
22. Kelcher, C. L., Plimption, S. J. and Hamilton, J. C., *Phys. Rev. B*, 1998, **58**, 11085.
23. Hertz, H., *J. Reinen Angew. Math.*, 1882, **92**, 156.
24. Johnson, K. L., *Contact Mechanics*. Cambridge University Press, Cambridge, 1985.
25. Balamane, H., Halicioglu, T. and Tiller, W. A., *Phys. Rev. B*, 1992, **46**, 2250.
26. Hirth, J. P. and Lothe, J., *Theory of Dislocations*. Wiley, New York, 1982.
27. Bernstein and Kaxiras, E., *Phys. Rev. B*, 1997, **56**, 10488.
28. Jones, R. O. and Gunnarsson, O., *Rev. Mod. Phys.*, 1989, **61**, 689.
29. Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. and Joannapoulos, J. D., *Rev. Mod. Phys.*, 1992, **64**, 1045.
30. Smith, G. S., Tadmor, E. B. and Kaxiras, E., *Phys. Rev. Lett.*, 2000, **84**, 1260.
31. Holm, R., *Electric Contacts: Theory and Application*. Springer, New York, 1967.