Tribol Lett (2013) 50:331–347
DOI 10.1007/s11249-013-0128-8

ORIGINAL PAPER

A Numerical Contact Model Based on Real Surface Topography

Can K. Bora · Michael E. Plesha · Robert W. Carpick

Received: 6 July 2012 / Accepted: 9 March 2013 / Published online: 31 March 2013

Abstract A numerical finite element contact model is developed to make use of the high precision surface topography data obtained at the nanoscale by atomic force microscopy or other imaging techniques while minimizing computational complexity. The model uses degrees of freedom that are normal to the surface, and uses the Boussinesq solution to relate the normal load to the long-range surface displacement response. The model for contact between two rough surfaces is developed in a step-by-step manner, taking into account the far-field effects of the loads developed at asperities that have come to contact in previous steps. Method accuracy is verified by comparison to simple test cases with well-defined analytical solutions. Agreement was found to be within 1 % for a wide range of practical loads for the high precision models. Applicability of extrapolation from lower precision models is presented. The real contact area estimates for micrometer-size tribology test machine surfaces are calculated and convergence behavior with mesh refinement is investigated.

Keywords Contact mechanics · Finite elements · Boussinesq solution

C. K. Bora (✉) · M. E. Plesha
Department of Engineering Physics, University of Wisconsin-Madison, Madison, WI 53706, USA
e-mail: bora@cae.wisc.edu

R. W. Carpick
Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA 19104, USA

1 Introduction

Most engineering surfaces are rough, regardless of whether the surfaces are naturally created or are processed. When two surfaces come into contact, this roughness causes multi-point contacts such that the actual area of contact is only a small fraction of the available contact area.

Atomic force microscopy (AFM) and other similar imaging techniques enable measurement of surface roughness at near-atomistic length scales. There is potentially substantial revenue in utilizing this high-detail surface topography to model and investigate the connection between micro- and macroscales of contact, adhesion, and friction.

Most conventional methods of modeling rough surfaces in contact replace the actual surface roughness with a distribution of non-interacting hemispheres, using statistical information about surface heights and slopes at a single length scale [1, 2] or multiple length scales [3, 4]. This allows for application of well-known contact models to the individual hemispherical contact points, and allows for investigating the multi-scale geometry of surfaces. However, modeling a surface using a hierarchy of hemispheres implies a loss of information in that the high-detail topography of the original surface is not directly exploited in the analysis.

Furthermore, most conventional methods do not take into consideration the effect of contacting zones on their surrounding areas. When the contact pressure increases over a given macroscopic surface area, an increasing number of asperities, at various distances from each other, come into contact and it becomes crucial to account for interaction between the microcontacts. Furthermore, the elastic deformation due to the compression of a local region tends to persist over significant lateral lengths (a point load only decays as $1/r$, where r is the distance from the point of application of the load) [5]. Therefore, the deformation of one asperity influences the deformation of neighboring asperities, and vice versa [6]. Thus, contact models that are based on single asperity contact behavior become deficient. Polonsky and Keer [7] argue that a numerical solution technique using actual geometry of real surfaces is necessary to accurately account for the interaction between the microcontacts. However, Persson et al. [8] have shown convincingly that, for surfaces with self-affine roughness, the power spectrum can be used for calculating overall contact properties (but not for creating actual maps of contact points).

One method to directly use the actual surface topography and to model inter-asperity interactions is to use conventional three-dimensional (3D) finite element discretization. Such models, while potentially highly accurate, require an enormous number of degrees of freedom (d.o.f.) and, correspondingly, computer power and time. Hyun et al. [5] model a 512 × 512 pixel surface contacting a flat surface using over 911,000 nodes with three d.o.f. each and 568,000 tetrahedral solid elements; a typical finite element model is shown in Fig. 1. To model the approach between the contacting surfaces, a dynamic method is used wherein inertia is included in the equations of motion, which requires small time increments and the introduction of artificial damping. All of these make the solution computationally expensive.

Another numerical modeling approach, which is also employed in this study, is to make use of an analytical solution, such as the Boussinesq solution, to characterize the elastic deformation of a uniform, planar substrate due to normal direction loads [9, 10]. In this method, the displacement effects of multiple points of contact are coupled with each other, and solved in a system of algebraic equations. There are a number of models in the literature applying this method. Webster and Sayles [11] present a semi-analytical contact solution where they subdivide the contact area into rectangular segments, on which they assume a constant pressure. They demonstrate a solution for a 2D problem with a cylinder pressed on a perfectly aligned, directionally rough surface. Poon and Sayles [12] describe a similar method for an idealized 3D situation, where a smooth sphere is brought into contact with a directionally rough surface, with surface heights varying in only one direction. The pressure distribution in the direction with variable roughness is allowed to vary with increasing applied load, while the pressure in the direction with no change in heights is assumed to follow a Hertzian distribution. The authors include plasticity, such that the contact pressure is allowed to increase only until it reaches the hardness of the softer material. Ren and Lee [13] develop a moving grid method to avoid large sizes of the matrices that define the deformation coupling effect between the contact points. Polonsky and Keer [7] use a fast numerical integration technique to calculate the surface deflections and they employ a conjugate gradient method iteration scheme to reach contact distribution convergence. Following this study, Liu et al. [14] develop a 3D model for thermo-mechanical contact between two rough surfaces. All of these models generally start with a prescribed amount of normal approach between the surfaces, predict an overlap region, and then correct this while attempting to obtain convergence to a solution. This is a very useful approach when elastic contact is being considered. However, such an approach is not straightforward to use for contact problems with path and/or history dependencies, which are likely to occur if adhesion, plasticity, or viscoelasticity are of interest.

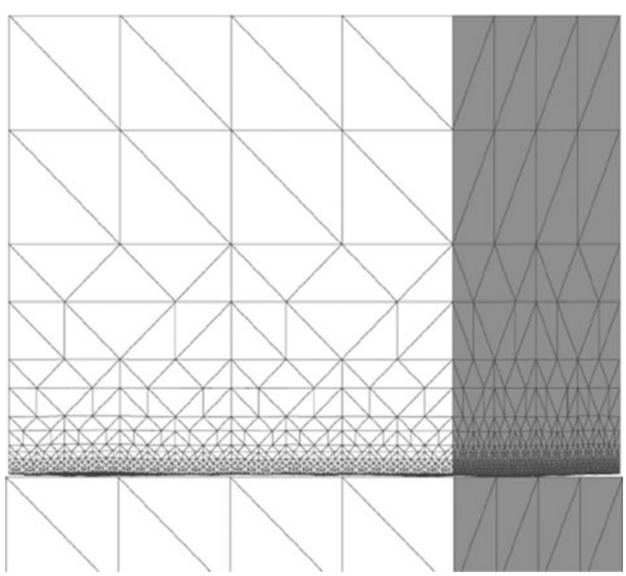

Fig. 1 Side view of a three-dimensional finite element mesh of an elastic body (*top*) with a 512 × 512 pixel resolution rough surface that is pressed onto a flat, rigid substrate. Hyun et al. [5]

Dickrell et al. [15] discuss a simple numerical model that takes into consideration the pixelated data from real surfaces obtained by common profilometry techniques. In their model, the surface asperities are assumed to be rigid-perfectly plastic and supported by a rigid substrate. Thus, there are no elastic deformations, and the softer of the two surfaces is assumed to yield wherever there is contact. The material that is displaced by plastic deformation is allocated to adjacent pixels. This enables the redistribution of plastically deformed material and its effect on contact area to be easily calculated. It is possible to add simple elastic behavior to the individual pixels in this model, but that would not account for the possibly significant long-range coupling effects due to the deformation of the substrate [5]. In this article, we extend the approach of Dickrell et al. using the Boussinesq displacement relations to create an elastic foundation that laterally couples different contact regions together. Here, we consider only elastic deformations, but the approach we describe can be further enhanced to include Dickrell et al.'s method to account for plastic deformation.

Our finite element approach uses a combination of analytically calculated surface behavior of a linear, elastic, homogenous, isotropic material subject to normal loads, i.e., the Boussinesq displacements, to characterize far-field deformations, and a surface layer discretization that directly utilizes AFM data to account for roughness. In essence, the surface roughness is a thin layer that overlies an elastic substrate. To investigate the development of contact, we follow a step-by-step approach, which does not require a convergence consideration. We discuss methods to minimize the size of the matrices and to speed up the detection of contact points. In the following sections, we describe our model, show validations through example cases compared to analytical and other numerical solutions, and discuss accuracy of the method. We then apply the method to investigate contact behavior of surfaces from actual MEMS-based friction experiments.

In this article, we focus on elastic contact situations. While other cases, such as plasticity and adhesion [16], are important phenomena to consider for nanoscale contact, the case of purely elastic contact is important and is of wide interest. For example, interfaces that are brought into contact without sliding, especially if between high-strength materials, can be mostly elastic; the contact properties are important for considering heat transfer, electrical conductivity, stiffness, and leak rates. As well, the steady-state sliding of worn-in surfaces can progress to mostly elastic contact after the tallest and sharpest asperities are worn away. In contrast to the references cited earlier, our approach offers some significant advantages, including that it is carried out incrementally (i.e., in step-by-step fashion), wherein the evolution of asperity interactions is determined. An incremental approach is required for problems that have path and/or history dependencies, such as are likely to occur with adhesive, plastic, or viscoelastic materials. Thus, the approach developed in this article fundamentally provides a general foundation upon which enhancements can be subsequently added.

In addition, instead of using aggregate properties of the surfaces, like the power spectrum [8], the direct calculation of contact properties enables the contact points to be individually determined, enabling visualization, mapping, and further characterization of the entire calculated contact interface.

2 Description of the Model

The topography of a surface, as obtained by AFM imaging, is a set of height data for a rectangular region of a surface area, as shown in Fig. 2. Our model features a one-to-one representation of each of the contacting surfaces, using rectangular prisms of material that protrude from each surface at every pixel, and these prisms of material are called voxels. In other words, voxels are the smallest box-shaped parts of a 3D scan and the name is derived by contracting the words "volume" and "pixel."

The model discretizes each of the two contacting surfaces using two regions. The first region, defined as the *substrate*, is an elastic half-space that is discretized using a set of nodes that lie in a horizontal plane and whose deflections are fully coupled with each other. The second region, the *interface*, is a thin surface layer consisting of individual, uncoupled springs that protrude from the substrate at every pixel, as shown in Fig. 3. The surface topography, such as that obtained from an actual AFM image, is represented in the interface domain. Nonlinear material properties including adhesion and plastic deformations can be assigned to the springs that define the interface. While our model uses an elastic half-space for the substrate region, it is possible to use other substrate domain types, such as a thin or thick plate, etc.

In this finite element method, nodes have only normal direction displacements as d.o.f. These displacements are coupled with one another within the substrate using the Boussinesq solution, which provides displacement response for all the surface nodes for a given vertical loading problem [9, 10]. The substrate is thus modeled as a superelement, representing the deformable half-space. Assuming a perfect alignment of the data points on the two contacting surfaces, the possible contact locations are quantized as the square pixels corresponding to the voxels of the two surfaces (i.e., two contacting surfaces imaged with 512 × 512 pixel resolution will have $(512)^2$ possible contact points). While our approach is an example of a finite element method, it is not a traditional finite element approach wherein a volume is discretized using continuum finite elements. Rather, the substrate is a super-element (or an infinite element) and simple spring finite elements are used to discretize the surface layer [17].

2.1 The Substrate

The surface deflection for an elastic half-space subjected to a normal direction point load is described by Boussinesq [9] and Love [10] as

$$u_z(x, y, 0) = -\frac{(1 - \nu)P}{2\pi G\sqrt{x^2 + y^2}}, \tag{1}$$

where x and y are the coordinates of the surface points relative to the point of load application; P is the point load applied at the origin, as shown in Fig. 4a; G and ν are the shear modulus and Poisson's ratio of the elastic half-space, respectively; and u_z is the displacement in the z direction

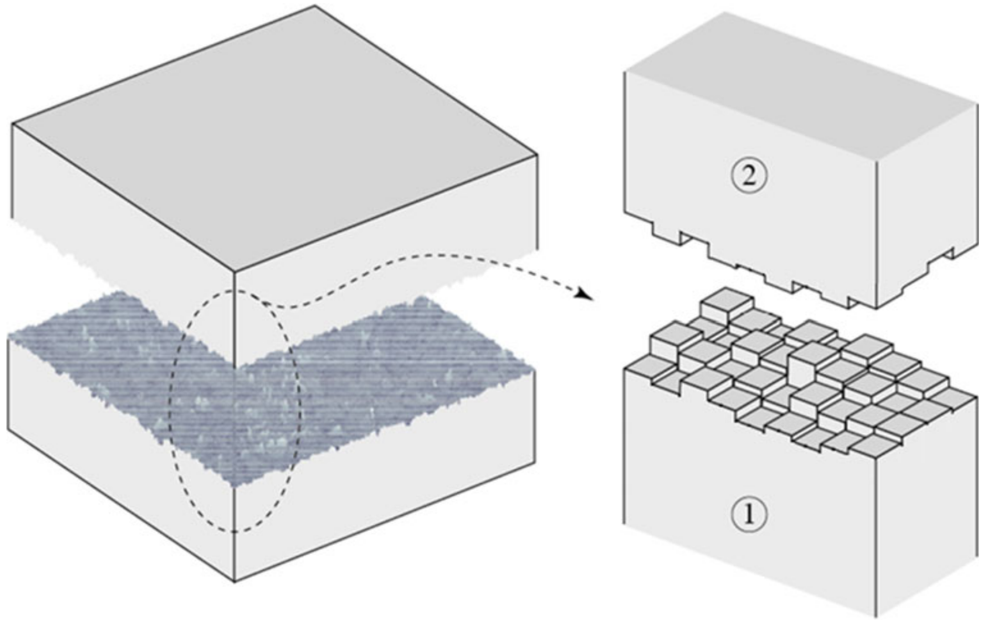

Fig. 2 Surface representation with voxels

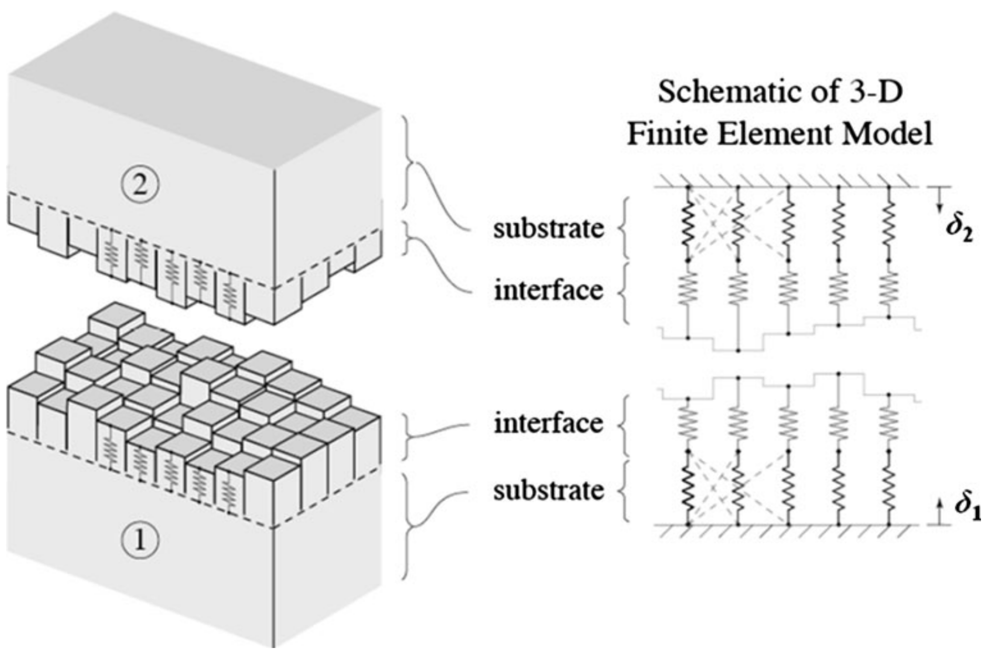

Fig. 3 Surface representation for the finite element model: the elastic horizontal coupling of the voxels is achieved in the substrate domain. The interface domain is represented with individual axial springs protruding from the substrate

which is normal to the interface. This relation is singular at the coordinate system origin, making it impractical to use as a force–deformation calculation. When the point load is relocated to the coordinates (s, t), as shown in Fig. 4b, the surface deflections are

$$u_z(x, y, 0) = -\frac{(1-\nu)P}{2\pi G\sqrt{(x-s)^2+(y-t)^2}} \tag{2}$$

When the loading consists of a normal pressure distribution $p(s, t)$ over an area A, as shown in Fig. 4c, the displacement solution can be obtained using Eq. (2) to integrate the displacement effects due to loading over each infinitesimal area dA, or (ds dt), as

$$u_z(x, y, 0) = -\frac{(1-\nu)}{2\pi G}\iint_A \frac{p(s,t)}{\sqrt{(x-s)^2+(y-t)^2}}\,\mathrm{d}s\,\mathrm{d}t. \tag{3}$$

When the pressure distribution on a square pixel is assumed to be uniform, a surface displacement field can be obtained

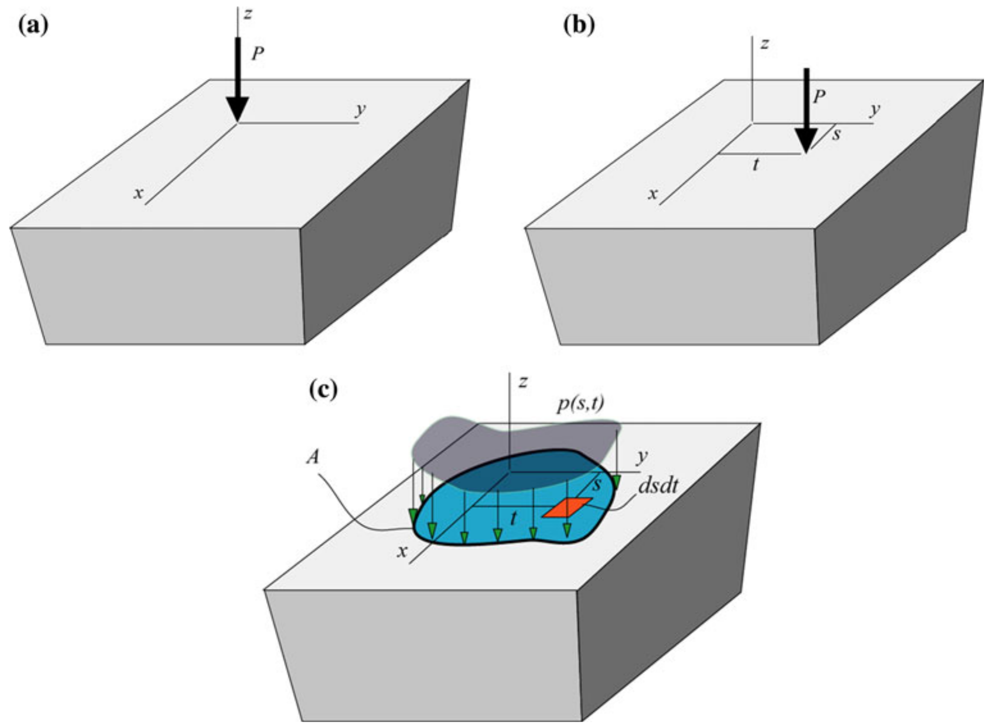

Fig. 4 Depiction of the Boussinesq problem for: **a** a point load at the center of the coordinate system, **b** a point load at coordinates (s, t), **c** a pressurized area A, where the pressure distribution is defined by a function, $p(s, t)$

using the above integral, with a total load of P over a square area of dimension $d \times d$, as shown in Fig. 5, as

$$
\begin{aligned}
u_z(x,y) = & \\
-C\Bigg\{ & (d+2y)\log\left(\frac{d}{2}-x+\frac{1}{2}\sqrt{(d-2x)^2+(d+2y)^2}\right) \\
& +(d-2x)\log\left(\frac{d}{2}+y+\frac{1}{2}\sqrt{(d-2x)^2+(d+2y)^2}\right) \\
& -(d-2y)\log\left(-\frac{d}{2}-x+\frac{1}{2}\sqrt{(d+2x)^2+(d-2y)^2}\right) \\
& -(d+2x)\log\left(-\frac{d}{2}+y+\frac{1}{2}\sqrt{(d+2x)^2+(d-2y)^2}\right) \\
& +(d-2y)\log\left(\frac{d}{2}-x+\frac{1}{2}\sqrt{(d-2x)^2+(d-2y)^2}\right) \\
& -(d-2x)\log\left(-\frac{d}{2}+y+\frac{1}{2}\sqrt{(d-2x)^2+(d-2y)^2}\right) \\
& -(d+2y)\log\left(-\frac{d}{2}-x+\frac{1}{2}\sqrt{(d+2x)^2+(d+2y)^2}\right) \\
& +(d+2x)\log\left(\frac{d}{2}+y+\frac{1}{2}\sqrt{(d+2x)^2+(d+2y)^2}\right)\Bigg\},
\end{aligned} \tag{4}
$$

where $C = \dfrac{(1-\nu)P}{2\pi G d^2}$.

Figure 6 shows the displacement field produced by a uniform pressure over a unit square pixel located at the origin of the coordinate system; the displacements are shown inverted for better visualization, and the boundaries of the pressure region are marked with thick lines. Equation (4) is defined everywhere except locations where a logarithmic operand is equal to zero, i.e., on line segments

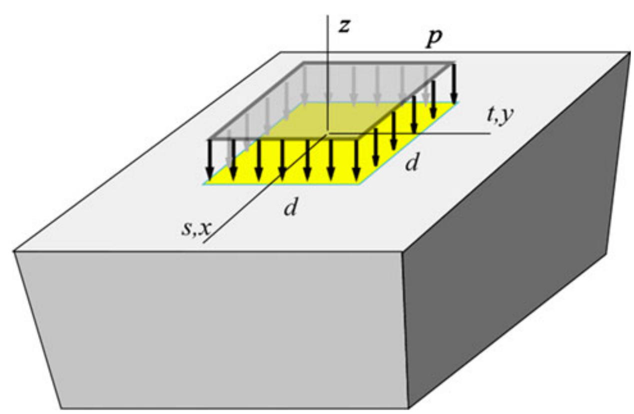

Fig. 5 Boussinesq problem for a square area of dimension $d \times d$ with uniform pressure p

$x = \pm 0.5$, for $-0.5 < y < \infty$, and $y = \pm 0.5$, for $\infty < x < 0.5$. Thus, the function is defined at the centers of all pixels, which are shown with dots in Fig. 6.

When the vertical deflections at the center of each surface pixel are defined as d.o.f, Eq. (4) can be used to obtain a flexibility matrix $\mathbf{S_B}$ that relates forces to displacements, where an entry s_{ij} in the flexibility matrix represents the displacement at the ith d.o.f. due to a unit load at the jth d.o.f. For given material properties and the pixel size d, this value of s_{ij} depends on the difference of the x and y coordinates between the two d.o.f. The flexibility matrix that is obtained is symmetric. By taking the inverse of $\mathbf{S_B}$, the stiffness matrix $\mathbf{K}^*_\mathbf{B}$ can be obtained, where an entry k_{ij} represents the force required at the jth d.o.f to cause a unit

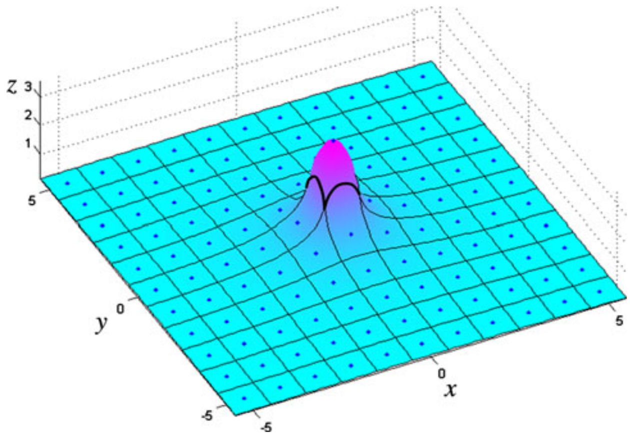

Fig. 6 Displacement field produced by a uniform pressure over a square region of unit area (boundaries of pressurized area indicated by the *heavy lines*) centered at (0,0), when the coefficient C in Eq. (4) is assumed to be 1. The displacements are inverted for better visualization

deflection at the ith d.o.f. Letting n denote the number of d.o.f. per surface, the stiffness matrix $\mathbf{K}_\mathbf{B}^*$ has a size of $n \times n$. This stiffness matrix relates the forces at all d.o.f., $\mathbf{f}$, with the deflections at all d.o.f., $\mathbf{u_z}$

$$\mathbf{K}_\mathbf{B}^*\mathbf{u_z} = \mathbf{f}$$

$$\mathbf{K}_\mathbf{B}^* = \mathbf{S}^{-1} = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{n1} & k_{n2} & \cdots & k_{nn} \end{bmatrix}. \tag{5}$$

Aside from the substrate d.o.f. described earlier, an additional d.o.f. is included in Eq. (5) for each surface to allow for a possible non-zero far-field displacement of the substrate, as denoted in Fig. 3 with δ_1 and δ_2; these are called “handle nodes,” or foundation nodes. The stiffness terms related to the foundation d.o.f. are augmented to $\mathbf{K}_\mathbf{B}^*$ at row and column number $(n + 1)$. The terms in the last row and column of the resulting matrix $\mathbf{K_B}$ are obtained by considering rigid body displacement capability. No forces should be generated when all the surface d.o.f. displace by the same amount as the handle node. Equation (5) can be rewritten with a force vector consisting of zeroes, and a displacement vector consisting of ones, both with size $(n + 1)$

$$\mathbf{K_B}\begin{Bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{Bmatrix} = \left[\begin{array}{ccc|c} & & & k_{1(n+1)} \\ & \mathbf{K}_\mathbf{B}^* & & \vdots \\ & & & k_{n(n+1)} \\ k_{(n+1)1} & \ldots & k_{(n+1)n} & k_{(n+1)(n+1)} \end{array}\right]\begin{Bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{Bmatrix} \tag{6}$$

where $k_{i,n+1} = k_{n+1,i} = -\sum_{j=1}^{n} k_{ij}$

2.2 The Interface

In the interface layer, each voxel element is modeled as an axial spring. To determine the stiffness k of the voxel, we assume the x and y direction strains to be zero throughout the voxel. This assumption is warranted because the voxel is supported from its sides by adjacent voxels, except for possibly a small height difference that might make it extend beyond its surrounding neighbors. Thus, the stress (σ_z)–strain (ε_z) relation for the z direction is

$$\sigma_z = \frac{E(1-v)}{(1+v)(1-2v)}\varepsilon_z, \tag{8}$$

where E and v are the elastic modulus and Poisson’s ratio of the interface material. All pixels have the same area d^2, and hence the stiffness for a voxel with height h above the substrate has the force displacement relation

$$k\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} P_1 \\ P_2 \end{Bmatrix} \tag{9}$$

$$k = \frac{E(1-v)}{(1+v)(1-2v)}\frac{d^2}{h}, \tag{10}$$

where u_1 and u_2 represent the displacements, and P_1 and P_2 represent the external forces at the bottom and top nodes of the voxel, respectively, as shown in Fig. 7. If the interface material is incompressible (i.e., $v = 1/2$), then Eq. (10) provides an infinite value of stiffness k. Difficulties with incompressible elastic media are well known, and the fact that k becomes infinite for incompressibility is not fundamentally problematic and does not present any limitation of our approach.

When the interface deformability is represented in this fashion, the stiffness of the voxel is inversely proportional to the height h, which is a somewhat arbitrary term that needs to be carefully selected. The substrate elasticity represents the exact solution for a flat surface, and any additional finite stiffness due to the interface adds to the overall flexibility. Further discussion about the stiffness of the interface layer can be found in the example problems that are treated later.

2.3 Contact Between Surfaces

When a pixel from one surface makes contact with a pixel from the other surface, we model this contact using an additional very stiff spring element (i.e., a penalty spring). Stiffness of this contact element has the same form as Eq. (9). The topology of the global stiffness matrix for the entire discretization is shown in Fig. 8, where $\mathbf{K_c}$ in this

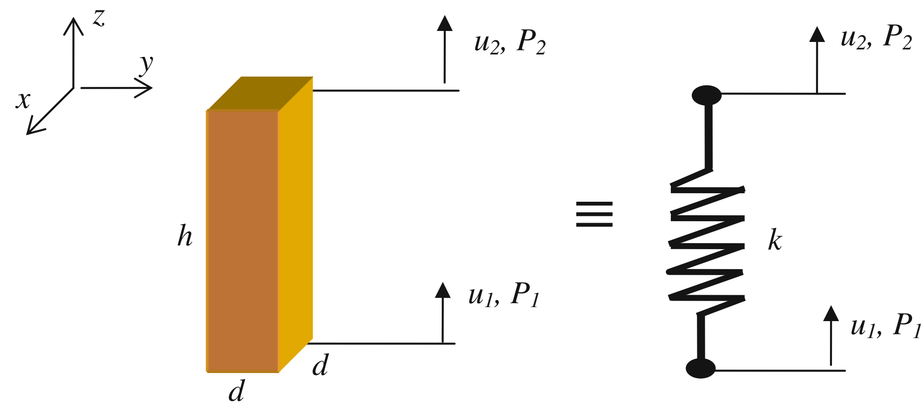

Fig. 7 Geometry of a single voxel with area d^2 and height h

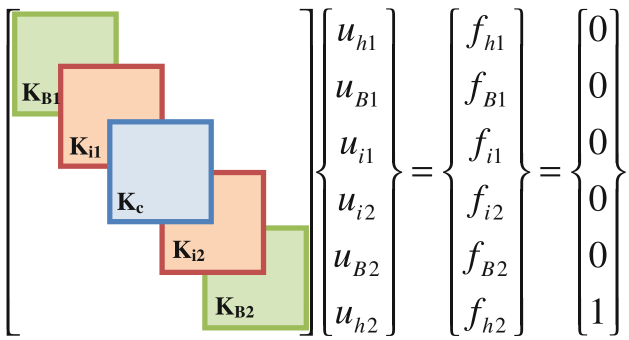

Fig. 8 Topology of the global stiffness equation including stiffness of the Boussinesq elements ($\mathbf{K_B}$), interface spring elements ($\mathbf{K_i}$), and contact elements ($\mathbf{K_c}$). The subscripts of displacements (u) and forces (f) represent the handle (h), substrate (B), and interface (i)

case is the stiffness matrix contributed by the penalty spring elements.

3 Algorithmic Considerations

3.1 Memory and Speed Considerations

Our approach substantially decreases the number of d.o.f. compared to a full three-dimensional finite element analysis. Nonetheless, when all the available pixels are coupled with each other for the Boussinesq half-space, the number of equations to be simultaneously solved becomes large. A grid of fully coupled N_p by N_p d.o.f. results in a Boussinesq superelement stiffness matrix size of $(N_p^2 + 1) \times (N_p^2 + 1)$, where N_p is the number of pixels per side of the surface image. For $N_p = 512$, this amounts to 262,145 equations. The global stiffness matrix including both surfaces with their substrate and interface layers would involve over a million d.o.f. This is less than the 2.7 million d.o.f. in the 3D FEA example in Hyun et al. [5], which discretizes one surface of the contact problem; however, the difference is not satisfactory, necessitating further reduction in our system size.

To reduce the size of matrices, two methods were considered. The first method investigated was the use of a coarse substrate mesh with a manageable size and treating intermediary points as “slave” d.o.f. In this case, although the substrate still had the general deflection shape of an elastic half-space, the coarsening of the mesh resulted in reduced precision in the contact area calculation, pressures, and displacements. When a load is applied as a uniform pressure on a larger square area (i.e., a pixel of a coarser mesh), the area of influence of an individual contact point becomes larger, while the maximum deflection and pressure are underestimated. Furthermore, large portions of the stiffness information, namely the equations contributing to the d.o.f. for voxels that are not in actual contact, are not utilized.

The second method, which we discuss below, involves reducing the superelement d.o.f. to only those that are associated with the voxels in contact. Usually, only less than a few percent of the apparent area is in contact, thus the required stiffness matrix for this method has substantially smaller size.

To understand the evolution of the contact area with increased compression, an incremental algorithm is used. One disadvantage of this method is that the stiffness matrix needs to be reformed with the addition of each new contact point. As the size of the stiffness matrix becomes large, this may lead to long calculation times. Several methods were implemented to reduce the program execution time, including: generating nodes only at contact points and updating the node list at each step; numbering the nodes with element-by-element ordering to reduce the populated portion of the stiffness matrix; using the same substrate flexibility matrix for both surfaces and using triple factorization for inverting the matrix; and carefully minimizing the portion of the area where the next contact point is searched.

3.2 The Algorithm

The coding for the finite element analysis was done as an enhancement to the FEMCOD program skeleton [18]. The FEMCOD program has features such as compact column (skyline) storage and an active column equation solver, which are useful for sparse or banded stiffness matrices as shown in Fig. 8.

At the start of the algorithm, the surface heights are entered into the program for each pixel over a square contact region for both contacting surfaces. These values represent the average heights of the voxels, and the d.o.f. are defined at the center point of each voxel. The highest sum of any two of the voxel height pairs is determined as the first contact point. Positioning the surfaces so that they touch at this point without any load allows the gaps between the upper and lower surfaces to be calculated and sorted, to be used in a simplified contact detection scheme.

Starting with the initial configuration described above, the stiffness matrix is created for a single point contact. At this stage, there are six d.o.f., consisting of the d.o.f. for the two handle nodes, the two substrate nodes, and the two surface nodes. A unit load (1 nN) is applied to the top handle node, while keeping the bottom handle node fixed. These linear equations are solved to obtain the displacements at the contacts.

To determine the next pair of contacting voxels, the displacements of the non-contacting voxels are calculated under the unit load for the step. Analysis of the whole surface is cumbersome, and not necessary for this search, as the surfaces are more likely to contact at locations with low gap values. On the other hand, the contact sequence does not simply follow the order of the gap values. To efficiently search for the next contact, a *reduced candidate method* was prepared, where a set of candidate locations is selected starting from locations with the smallest initial gaps. The size of this contact candidate list varies according to the number of existing contacts. Using the force displacement behavior for the load step, the smallest force required to form another contact is found and the associated location is marked as the next contact point. Multiple points that require the same smallest force are all included in the next load step.

New surface and substrate points, interface elements, and contact elements are generated and the connectivity information for the existing Boussinesq elements is updated. The process of solving for displacements and finding new contacts is iterated until the initially selected maximum number of contacts is reached. For each of these steps, a unit load is used to determine the force–displacement behavior. The force increment for every contact point is calculated at each step and added to the previous force value. A final check algorithm is introduced at the end of the simulation to verify that no contacts were missed with the reduced candidate contact detection method.

4 Verification Examples

In this section, we carry out several simulations to help verify the accuracy of the method and to study its convergence properties with mesh refinement. This requires simulations of problems having analytic solutions, and this requires modeling of problems with idealized surface geometry (e.g., perfectly flat, spherical, etc.). Examples using general surface geometries obtained from AFM profilometry are presented in Sect. 5.

For the voxel dimensions considered in the following examples, when the dimension h is chosen such that the entire roughness structure is contained in the interface layer, the layer becomes too soft. In the test cases we considered, it was found that the elastic behavior of the contacting bodies can be modeled solely with the Boussinesq substrates. For this reason, in the examples discussed in this article, the interface elements are given a stiffness value that is five orders of magnitude higher than the substrate layer, making them essentially rigid. In this form, the Boussinesq layer defines the elasticity and the interface layer is retained to model the roughness information and for future introduction of additional phenomena such as adhesion and plasticity.

4.1 Rigid Cylindrical Punch Pressed into an Elastic Half-Space

As a test example, a problem of a rigid circular punch is investigated. The lower surface is modeled as a flat elastic substrate, as shown in Fig. 9, with $E = 200$ GPa and $\nu = 0.25$. The upper surface is modeled as a rigid cylinder protruding from a rigid flat surface, with an elastic modulus that is five orders of magnitude higher than that of the lower surface.

Figure 10a shows the discretization of the circular punch for the first model, which has a 1 nm pixel size. From the discretized circular contact area, an effective radius was obtained and used in the analytical solution for comparison. Keeping the area of the circular punch constant, the mesh was refined twice, generating models at one-third and one-ninth of the initial mesh size, as shown in Fig. 10b, c.

The analytic solution for a rigid cylindrical punch of radius R contacting the surface of a semi-infinite body provides the surface displacement and pressure values as [19]

$$\delta = \frac{P(1 - \nu^2)}{2RE} \tag{11}$$

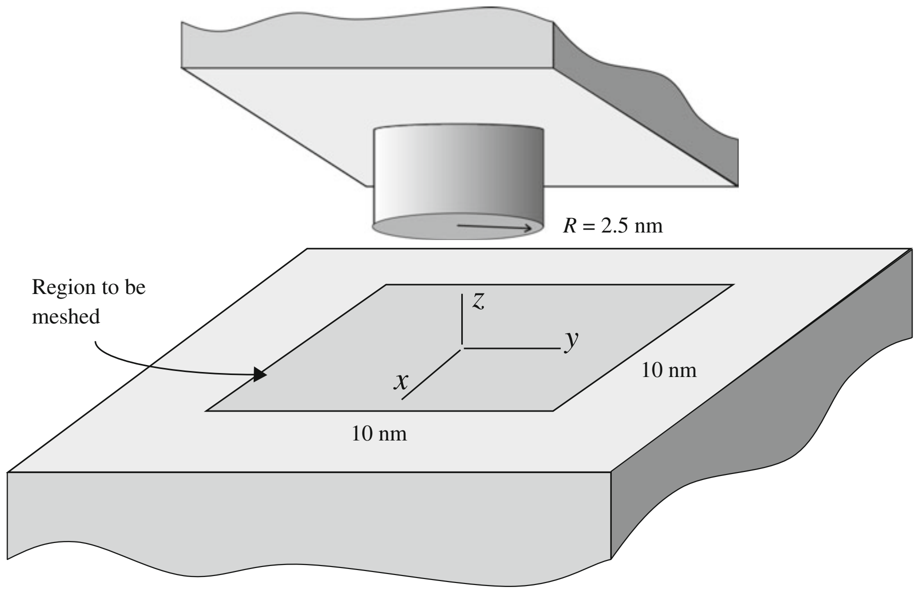

Fig. 9 Rigid cylindrical punch pressed into an elastic half-space

$$\sigma_C(r) = \frac{P}{2\pi R\sqrt{R^2 - r^2}}, \tag{12}$$

where δ is the displacement of the punch, P is the applied load, v is the Poisson's ratio, and E is the Young's modulus of the elastic half-space. The theoretical pressure σ_C at the edge of the punch is infinite. The analytical calculation for the pressure along a radius of the punch is shown with the solid line in Fig. 11, compared with the calculated pressure values for the different mesh sizes. The finest mesh size gives excellent agreement with the analytic solution for pressures and displacement, with errors less than 1 %. Table 1 compares the center node pressure values and the displacements with the analytical solution.

Richardson extrapolation [17] can be used to improve the results obtained using multiple mesh sizes according to the following relation, provided that the meshes have undergone regular refinements:

$$\phi_0 = \frac{\phi_1 h_2^{p_r} - \phi_2 h_1^{p_r}}{h_2^{p_r} - h_1^{p_r}}, \tag{13}$$

where φ_0 is the extrapolated value of the solution, φ_1 and φ_2 are FE approximate solutions obtained from different mesh sizes, i.e., h_1 and h_2, respectively, and p_r is the rate of convergence for the model. For our application of Eq. (13) to this example, the exact solution φ_0 is known, and with the FE results φ_1 and φ_2 obtained for two mesh sizes h_1 and h_2, Eq. (13) contains one unknown, namely the rate of convergence p_r. Using the results for the two coarsest meshes, shown in Table 1, we determine $p_r = 1.15$ for the rate of convergence for displacements, and $p_r = 1.06$ for the rate of convergence for stress. Using these convergence factors, another application of the extrapolation between the two finer mesh models provides estimates of the displacement value with an error of 0.01 % and the center point stress with an error of 0.031 %, when compared to the analytical solution.

With the values of p_r cited above, the rate of convergence is slightly better than linear, which is slow. Because of the differences between our model and the classical finite element methods, the exact nature of the rate of convergence is not immediately apparent. For this example, one factor affecting the rate is that the area that is discretized is not the same for each refinement, as we are approximating a circular edge using piecewise straight lines. Another factor is our attempt to converge to the asymptotic singular behavior of the stresses at the punch edge using square areas with uniform pressure. However, the displacement solution is nodally exact at the d.o.f. for the pressure distribution represented with the discretization, and the results show outstanding accuracy, even for the coarsest mesh sizes.

4.2 Rigid Square Punch Pressed into an Elastic Half-Space

To investigate the effect of approximating a curved boundary using piecewise straight line segments, we study a similar problem with a square punch, as shown in Fig. 12. Beginning with a single contact point solution, the mesh is refined three times, each time dividing the size by three, thus increasing the number of contact points by a factor of 9. Table 2 gives the calculated values for the vertical stress at the center point of the contact area and the displacement of the punch.

The analytical solution for this problem is approximate and thus, an exact error analysis cannot be performed. Borodachev [20] offers an approximate solution which is

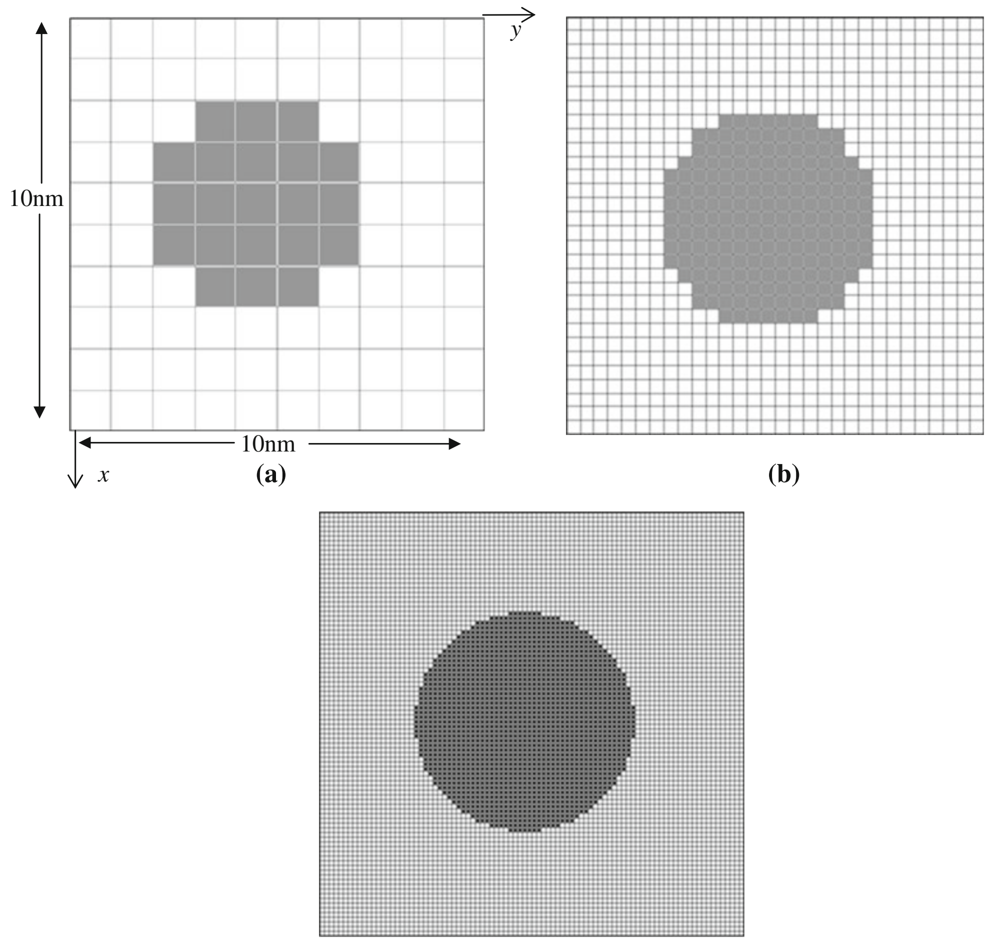

Fig. 10 Top view of the circular punch with contact areas modeled with **a** 21 pixels with 1 × 1 nm size, **b** 189 pixels with 1/3 × 1/3 nm size, and **c** 1,701 pixels with 1/9 × 1/9 nm size

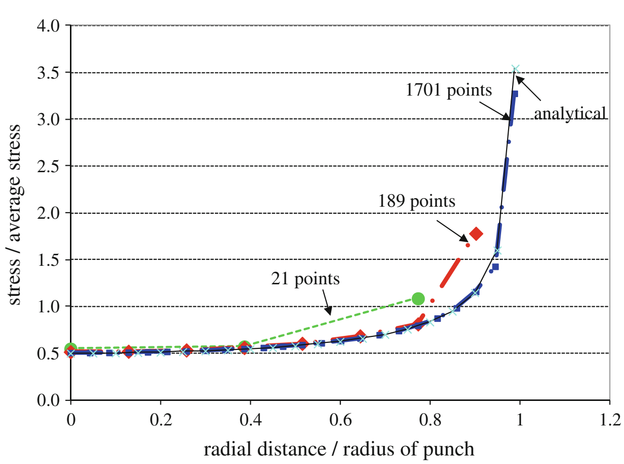

Fig. 11 Radial pressures for the rigid circular punch problem for FE models with different mesh sizes, and the analytical solution

used for determining the displacement and stress values given in Table 2. A rate of convergence p_r can be calculated using displacement results of three mesh sizes, assuming the rate is uniform. The first three mesh sizes yield a rate of convergence of $p_r = 0.874$ and a second calculation using the second, third, and fourth mesh sizes give $p_r = 0.973$. As the first mesh size contributes only to a uniform pressure distribution, the second convergence rate value is deemed to be more reliable. This rate is very close to linear, similar to the rates seen in the circular punch problem. Using $p_r = 0.973$, the displacement estimate can be extrapolated to 0.407×10^{-3} nm and the normalized

Table 1 Center node stress, displacement and calculated errors for the circular rigid punch on an elastic substrate, under a total load of 1 nN

Mesh size (nm)	Displacement (10^{-3} nm)	Displacement error (%)	Pressure at the center (nN/nm^2)	Pressure error (%)
1	0.955	5.37	0.0261	9.54
1/3	0.920	1.51	0.0245	2.97
1/9	0.911	0.485	0.0240	0.934
Analytical	0.907		0.0238	

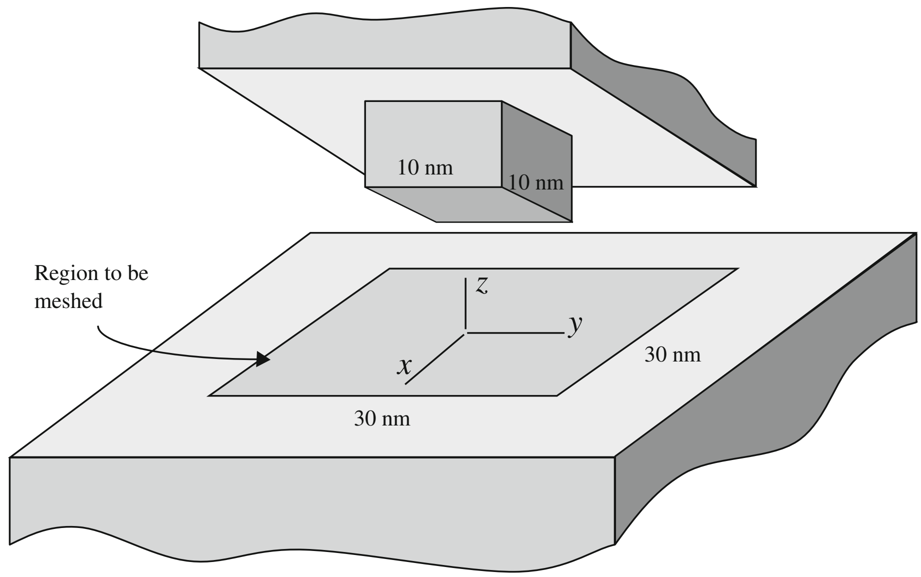

Fig. 12 Rigid square punch pressed into an elastic half-space, with $E = 200$ GPa and $\nu = 0.25$

Table 2 Center node stresses and displacements

Mesh size (nm)	Number of points defining contact	Displacement for 1 nN load (10^{-3} nm)	Normalized pressure at the center
10	1	0.526	1
3.333	9	0.451	0.511
1.111	81	0.422	0.519
0.3704	729	0.412	0.498
Analytical estimate		0.400	0.456

pressure at the center of the punch can be extrapolated to 0.486. The convergence rate for this problem is about the same as that for the cylindrical punch. Even though the contact region is easier to mesh for this example, modeling the pressure distribution at the edges becomes more challenging. Because the convergence rates obtained for the finer mesh for both models are close to 1, a linear rate of convergence is assumed for future examples. As with the circular punch, while this rate is low, the results show outstanding accuracy, even for the coarsest meshes.

The calculated stresses along the x axis (passing through the center point of the punch, parallel to the side of the square) are shown in Fig. 13. Three mesh sizes using the method developed in this article are shown with the square data points. These are compared with solutions from Borodochev's approximate analytical solution [20] and 3D FEM analysis using ANSYS, which are shown with the circular data points. In the two ANSYS models, the contact region is discretized into 36 and 144 elements, respectively, using 20 node quadratic solid elements. Symmetry is employed to simplify the model. Richardson extrapolation was performed using the two mesh sizes.

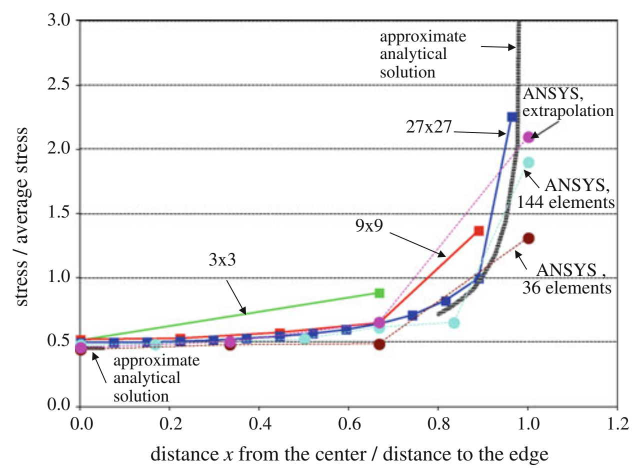

Fig. 13 Comparison of radial pressures for the rigid square punch model to a conventional 3D models and an approximate analytical solution [20]

It is seen in Fig. 13 that our method overestimates the stress along the axis considered, while the 3D ANSYS model underestimates it, and both show better agreement with the approximate analytical result with mesh refinement. The difference between the extrapolated central stress values is less than 1 %. Overall, there is outstanding accuracy of the proposed model even with the coarsest mesh sizes.

4.3 Rigid Spherical Surface Pressed into an Elastic Half-Space

A spherical contact problem, as shown in Fig. 14, was modeled to test the step-by-step contact detection algorithm. In contrast to the previous examples, this example has varying surface heights.

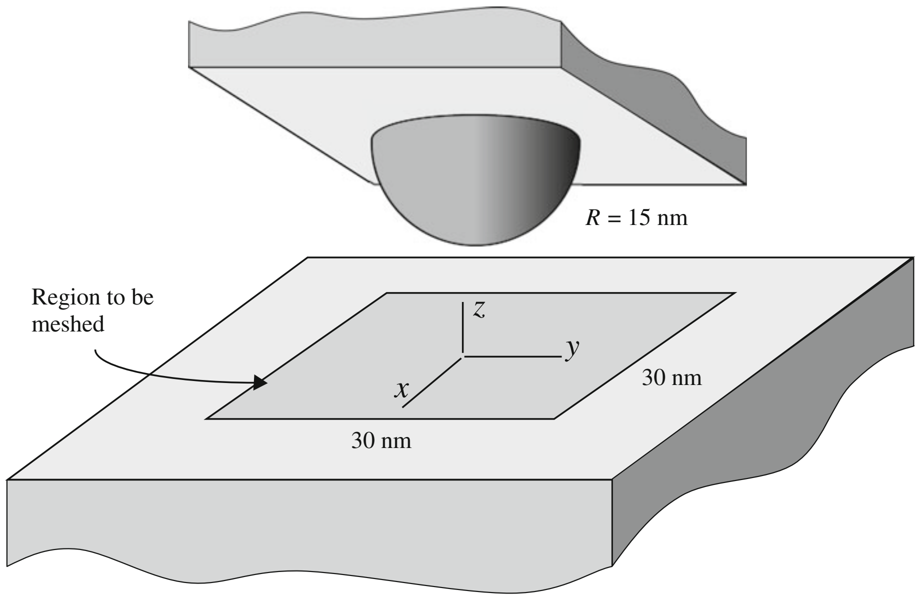

Fig. 14 Rigid spherical surface pressed into an elastic half-space

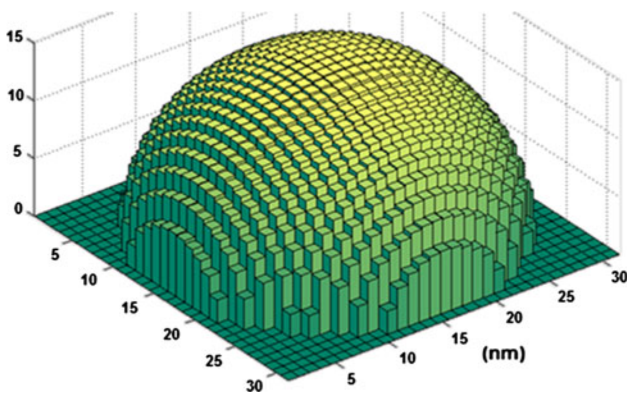

Fig. 15 Spherical punch model used in the example of a rigid sphere contacting a flat elastic surface. The sphere has a radius of 15 nm and the pixel dimension for this model is 1 nm

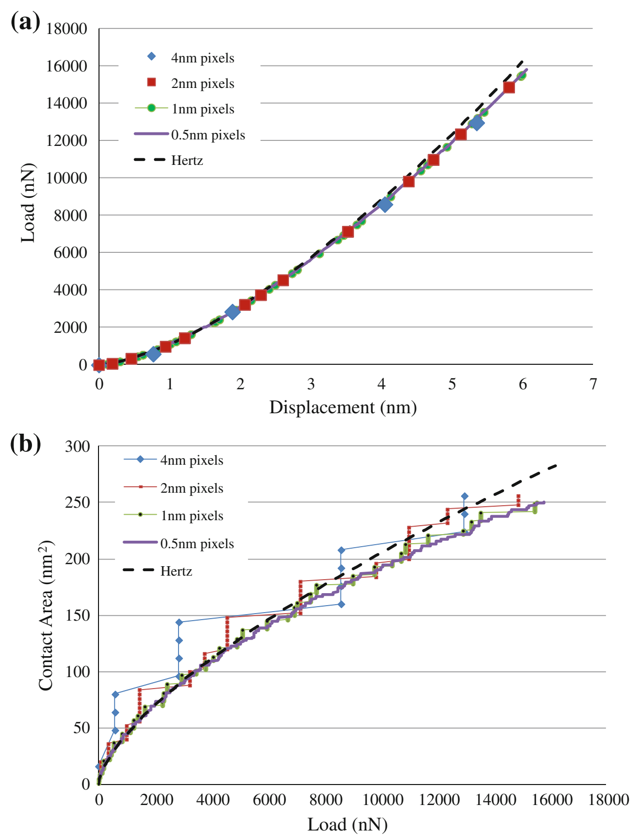

Fig. 16 **a** Load versus displacement and **b** area versus load results for a rigid spherical punch contacting a flat elastic surface, compared to the Hertz solution

For two materials with Young's moduli of E_1 and E_2 and Poisson's ratios of v_1 and v_2, an effective elastic modulus E^* the force–displacement (F–d) relation and the contact area–force (A–F) relation are as follows [19]:

$$\frac{1}{E^*} = \frac{1 - v_1^2}{E_1} + \frac{1 - v_2^2}{E_2} \tag{14}$$

$$F = \frac{4}{3} E^* R^{1/2} d^{3/2} \tag{15}$$

$$A = \pi \left(\frac{3FR}{4E^*} \right)^{2/3}, \tag{16}$$

where R is the radius of the sphere, A is the contact area, and F is the contact load. For this example, $E_1 = 200$ GPa, $v_1 = 0.25$, $R = 15$ nm and the second material is assumed to be rigid.

Pixel sizes of 0.5, 1, 2, and 4 nm are used to investigate the performance of the algorithm. Figure 15 shows the spherical surface modeled with the 1 nm pixel size. The force displacement behavior does not change with changing pixel sizes in our model, as seen in Fig. 16a. The contact area calculation seen in Fig. 16b shows a step-wise increase with increased load, caused by the discretized nature of the surface, but the overall trend between the models with different pixel size is consistent.

In both the figures, excellent agreement with the Hertz model is seen until a contact area of about 100 nm^2. The Hertz solution is not considered to be valid past this region, as it assumes the contact radius to be much smaller than the radius of the spherical surface [19]. A power law fit to our data for the 0.5 nm pixel case in the full range shown in Fig. 16 gives an area–load dependence of $A \propto F^{0.68}$, which is in close agreement with the $A \propto F^{0.667}$ relation for the Hertz solution given in Eq. (16).

5 AFM Surface: Experiments with Resolution

The AFM topography image of a polycrystalline silicon surface-micromachined nanotractor actuator was used as a sample case [21]. The AFM surface was placed at the bottom and its contact with a rigid flat surface was modeled. For the AFM surface, the material properties are $E = 200$ GPa and $\nu = 0.25$, while the rigid surface was modeled with an E value that is larger by five orders of magnitude. The image used is of a 5 × 5 μm area measured at 512 × 512 pixel resolution ($N_p = 512$). To investigate the behavior of elastic contact with varying sampling sizes, the surface resolution was reduced to obtain images of $N_p = 256$, $N_p = 128$, and $N_p = 64$. Let i, $j = 1 \ldots N_p$; to reduce the resolution by half, for example, pixels with odd i and odd j indices can be selected, ignoring the other pixels. This method is consistent with the way an AFM instrument measures surface heights for different resolutions. An alternate method is also investigated, which consisted of averaging the neighboring four pixels to obtain a lower resolution height value. The displacement and force analysis for the first method with varying sets of odd and even i, j and the alternate method give similar results, with errors within ±2 % for 1,000 contacts.

Figure 17 shows the load versus displacement and contact area versus displacement graphs for the different resolution AFM images. The resolution of the image does not have any significant effect on the load versus displacement behavior, while the contact area for a given displacement is strongly dependent on the resolution. For a given displacement, the high resolution image gives a much smaller contact area. According to our model, simply dividing a pixel under uniform pressure into four smaller pixels of equal height does not change the results. However, in the higher resolution image the four pixels are generally not at the same height. When the highest of these pixels come into contact, it delays the contact of the remaining pixels. In the purely elastic case, this effect is exacerbated, whereas in a plastic model, the pixel that

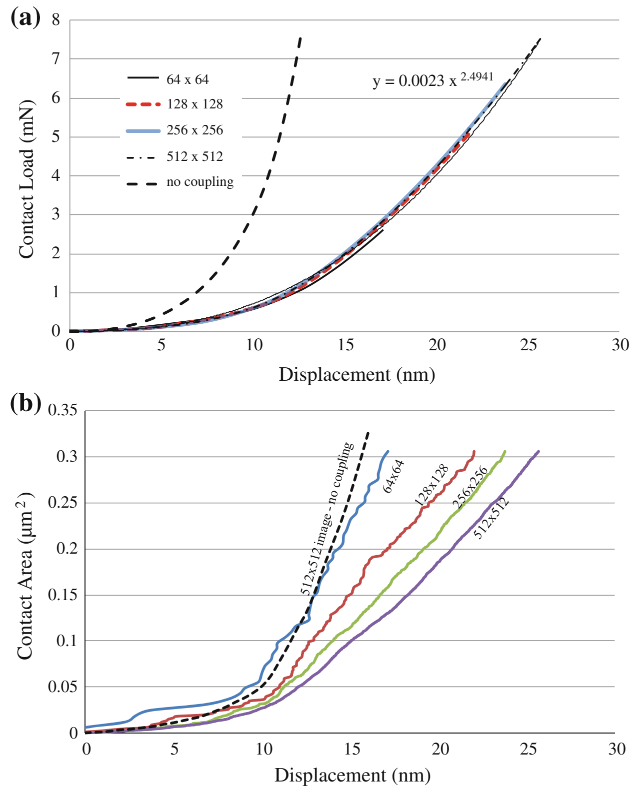

Fig. 17 The *solid lines* represent **a** load versus displacement and **b** area versus displacement results for the polycrystalline silicon surface at different resolutions, modeled with the elastic Boussinesq substrate model pressed against a rigid flat surface. The *dashed line* represents the contact area obtained by an elastic response without any coupling between the contact points (i.e., with a rigid substrate)

comes into contact first would likely yield and the surrounding pixels would more easily come into contact. The differences between the contact areas for the different resolution models will likely be smaller if plasticity is included.

The results of a simple contact area calculation representing no elastic coupling (i.e., the substrate of the rough surface is rigid, and the elastic voxels deform independently from each other) between the contact points is shown in Fig. 17a, b with dashed lines. The stiffness of an uncoupled voxel was obtained from the Boussinesq problem with a single pixel under contact; i.e., using Eq. (4) with $(x, y) = (0, 0)$. For the 512 × 512 pixel image representing a 5 × 5 μm surface with $E = 200$ GPa and $\nu = 0.25$, the center point of a single pixel under uniform pressure would deform with a stiffness value of 5.39×10^{-4} N/m. The area data is obtained by counting the voxels in the 512 × 512 image that are higher than the given displacement. For a given displacement value, the contact area estimated with no elastic coupling is much higher, and the contact load is much lower than the results from our model, as expected.

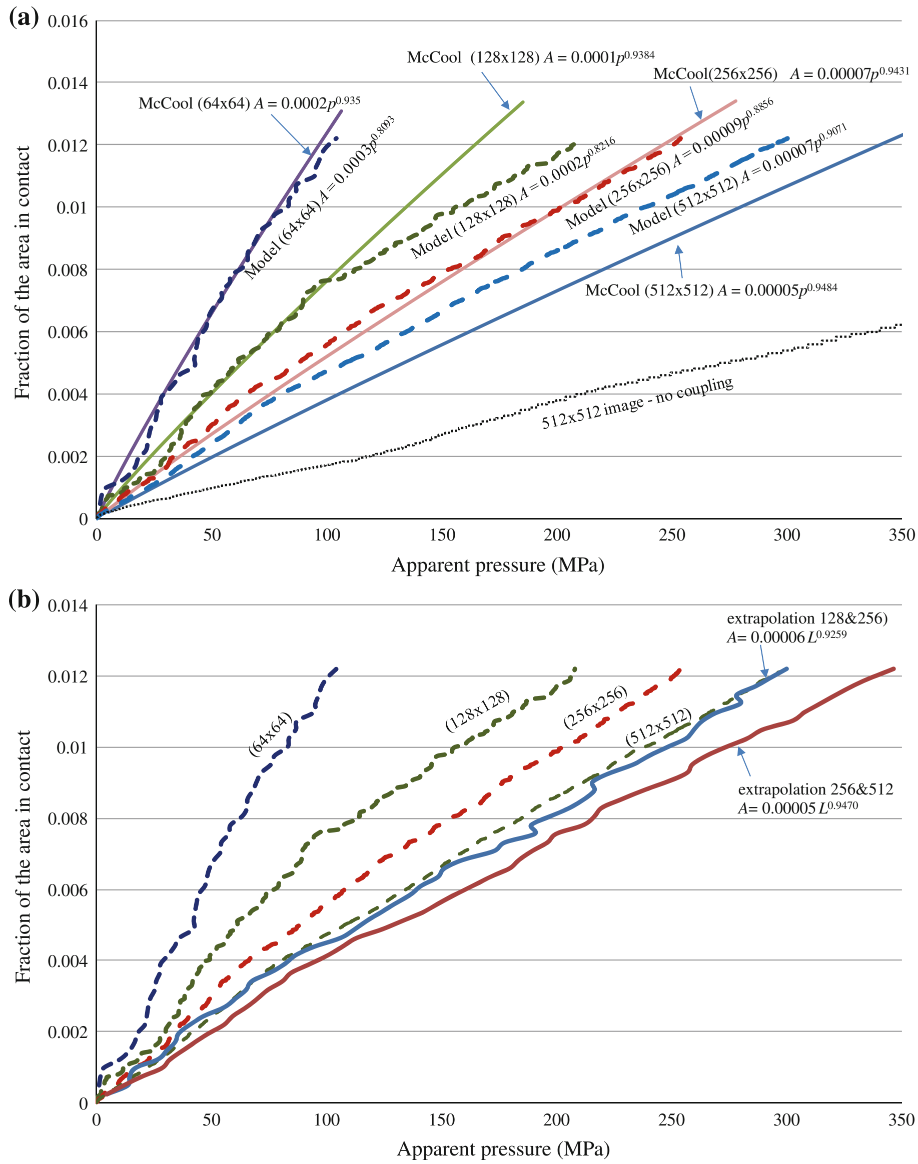

Fig. 18 **a** Contact area fraction versus apparent pressure results for the polycrystalline silicon surface pressed against a rigid flat, modeled at different resolutions. The results from our model are shown by the *dashed lines*, McCool [2] analysis results are shown with *solid lines*. The *dotted line* represents the behavior when the substrate effect is suppressed. **b** Two separate extrapolation calculations (*solid lines*) are shown with the model results

Figure 18a shows the contact area fraction versus apparent pressure graphs for the surfaces using our model, shown in dashed lines, and also using McCool's statistical method [2] based on Greenwood and Williamson's model [1], shown by the solid lines. The trends follow a power law, where the exponent increases as the resolution is increased. Increased resolution also decreases the contact area estimate for a given pressure value. The importance of the sampling resolution is further demonstrated with the results from the McCool analysis, which uses size-dependant RMS heights, slopes, and curvatures as input.

Overall trends for the 64 × 64 and 256 × 256 pixel images look similar between the two methods, with the fraction of the area in contact becoming smaller as the resolution increases. However, the power law predicted by our model has a smaller exponent than McCool's method for every resolution image. Our results for the 128 × 128 pixel image diverge from the statistical estimate at around 0.8 % real contact area, while the 512 × 512 pixel image results are different from the beginning of the contact, with the difference typically larger than 15 %.

The 512 × 512 image provides an area–load dependence relation of $A \propto L^{0.907}$. Figure 18b shows two sets of Richardson extrapolation results for the different resolution images, using a linear convergence rate ($p_r = 1$). When the data from 128 × 128 to 256 × 256 resolution images were used, the extrapolation gives a trend similar to that of the 512 × 512 image. The power law trend of the extrapolation is $A \propto L^{0.926}$. An extrapolation between the 256 × 256 and the 512 × 512 images gives an area–load relation of $A \propto L^{0.947}$, which is similar to the McCool method results from the highest resolution image. It should

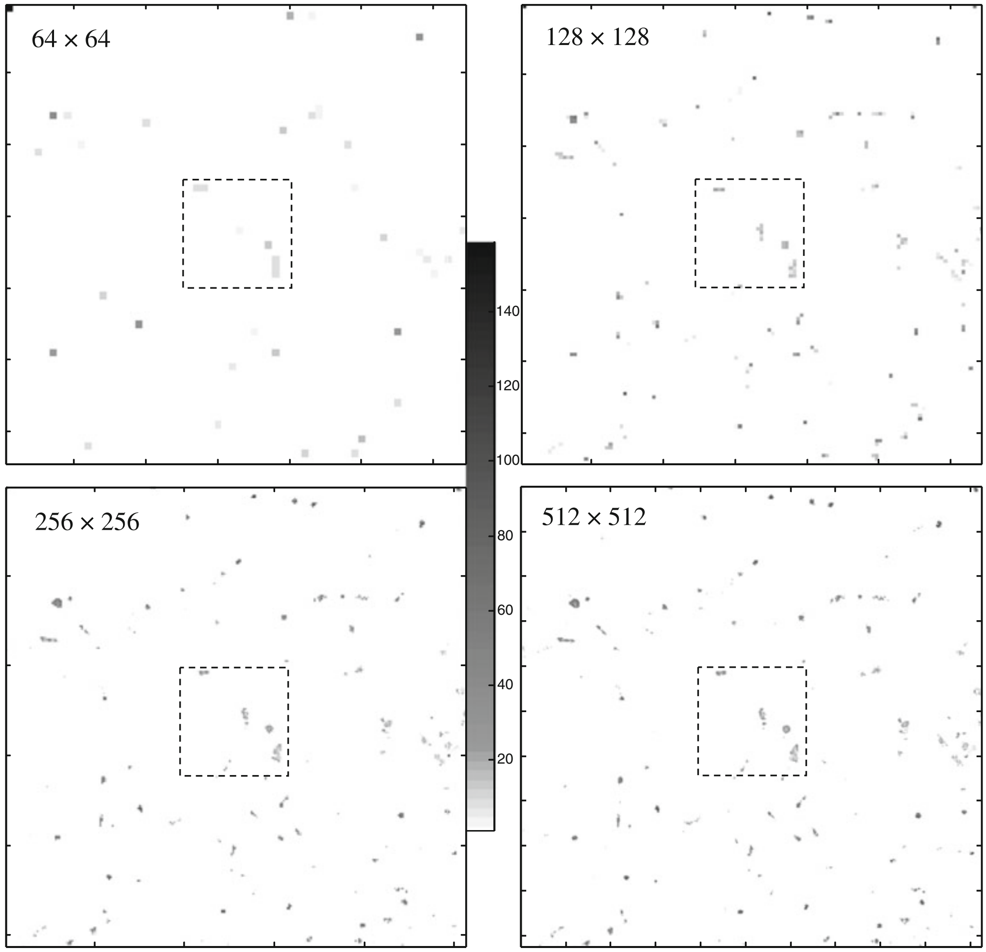

Fig. 19 Pressure maps of the polycrystalline silicon surface placed against a rigid flat surface at different resolutions, when 1.2 % of the area is in contact. Values are given in GPa. See Fig. 20 for zoomed-in images with finer detail of the regions bordered with *dashed lines*

be noted that there are other results in the literature for this relation; one of these is Zhuravlev's model [22], which predicts an exponent of 0.91. For this purely elastic case, the exponent becomes larger with the increased image resolution and extrapolation. The exponent values are within ranges estimated by the statistical models.

The contact area versus load behavior of a simple model with no elastic coupling between contacts is shown with a dotted line in Fig. 18a. For a given contact load, a much smaller contact area is estimated when substrate coupling effects are neglected. This shows the importance of including those effects in the calculation.

Figures 19 and 20 show the actual pressure distribution of the contact spots for the four different image sizes. At 1.2 % of the total area in contact, the calculated maximum elastic contact pressures are close to 150 GPa. Although this value is well past the hardness of the material, the solution was extended to these pressure levels to study the behavior and make comparisons to the statistical models, and also to study the effects of the resolution on the pressure distribution. While the calculations on the smallest image size give only a crude estimate of the contact locations, for the image at 128 × 128 pixels, it is actually possible to identify the contact shapes, pressure distributions, and intensities within individual contact points. The estimated pressures become higher and the shapes become smoother with further increase in resolution.

For the AFM surface example, a comparison of the elastic stresses at each pixel to material hardness (H) shows the number of contact points where the stress exceeds the yield stress, and the fraction of the contact where stress exceeds the hardness for each step (Fig. 21). A hardness value of 11 GPa was used for the polycrystalline silicon material [21]. According to this comparison, the plastic region occupies 60–70 % of the real contact area through all stages of contact development, even at the smallest loads. With a proper plastic response model, the contact area estimate would be higher than what is observed in our results.

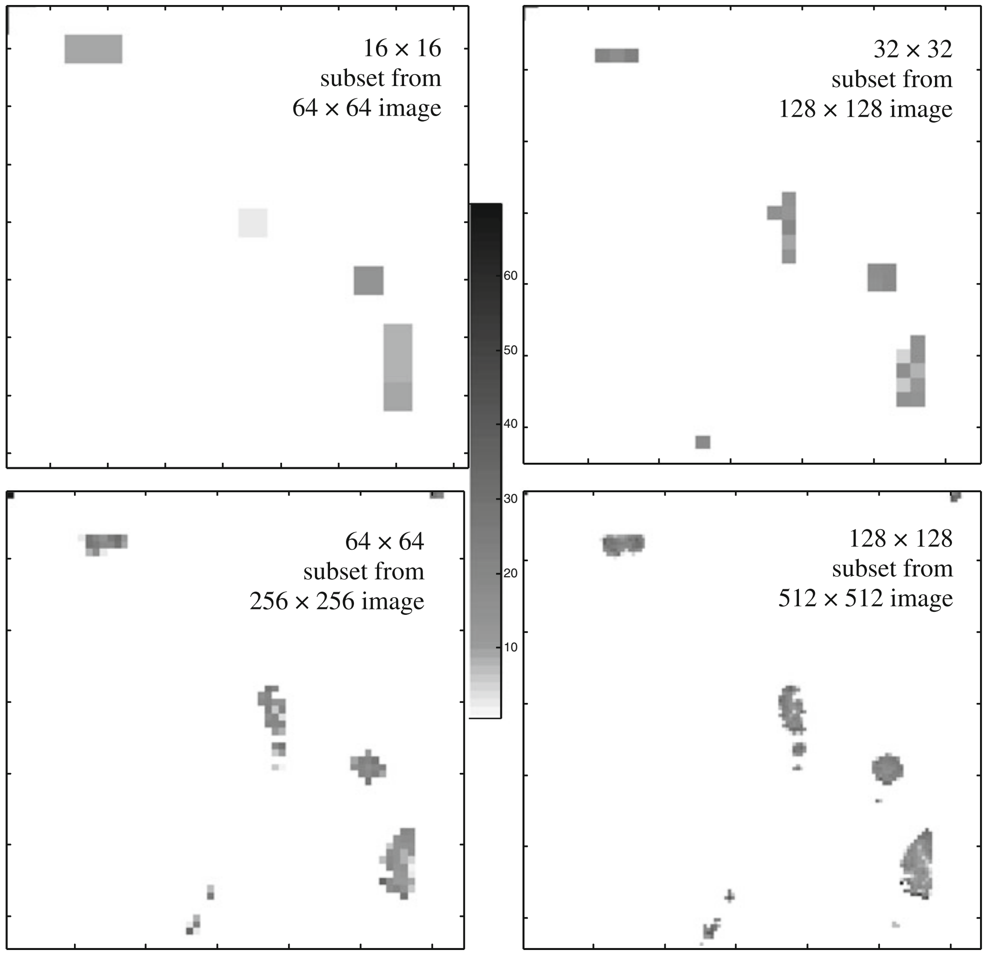

Fig. 20 Zoomed-in pressure maps of the polycrystalline silicon surface placed against a rigid flat surface at different resolutions, when 1.2 % of the area is in contact. Values are given in GPa. The respective images correspond to the areas bordered with *dashed lines* in Fig. 19

6 Conclusion

This model was developed to preserve and fully utilize the high-detail surface topography data obtained from AFM or other profilometry methods. It makes use of analytical solutions to simplify the treatment of the elastic foundations, using d.o.f. only in the normal direction, and suppressing the need to discretize the substrate material in full 3D detail. This enables the important effect of elastic coupling between nearby contact points to be accounted for. The validity of the approach was verified with examples having idealized surface geometries so that analytic solutions were available. For the examples of rigid punches with different geometries pressed into an elastic half-space, we show that our method yields results that are in excellent agreement with analytical and 3D finite element solutions, even using coarse mesh sizes. The strength of the method is that the solution of the surface displacement is nodally exact for the employed pressure distribution. Furthermore, these test cases were used to gain insight into the accuracy and convergence of the method related to resolution (i.e., refinement). Examples of Richardson extrapolation are demonstrated for the trial cases and the real AFM surface models. Despite the linear convergence rate, the method can be reliably utilized to obtain high-accuracy estimates of the contact area–pressure relations using lower resolution image results.

In the tests using AFM surfaces, where the case of general 3D contact is demonstrated, the resolution of the image strongly affects the contact area estimate. The solutions presented in this article are completely elastic, and the differences in the responses for the different image size estimates are expected to decrease with the addition of plasticity. According to our estimate, a large portion (60–70 %) of the real contact area will undergo plastic deformation starting at the smallest loads, and continuing through all stages of contact development.

The method presented can be used to investigate effects of using different materials, surface roughening and texturing methods, and differences between unworn and worn

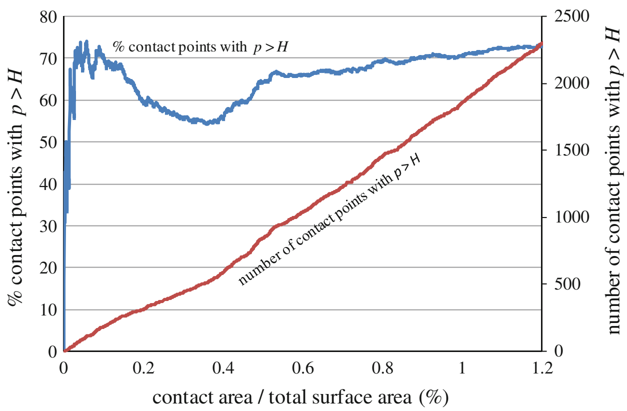

Fig. 21 The number and percentage of contact points that are estimated to be experiencing pressure values above the material hardness. For the 512 × 512 pixel example, when the contact area is 1 % of the total surface area, 1,854 of the 2,562 pixels experience pressure values, $p > H$

surfaces. By presenting here the most straightforward case of non-adhesive elastic interactions, we provide a basis for future extensions. The algorithm is carried out incrementally (i.e., in step-by-step fashion) wherein the evolution of asperity interactions is determined. This makes the algorithm amenable for enhancements that would require path- and/or history-dependent solutions such as adhesion, plasticity, or viscoelasticity. Other possible enhancements that could be considered include anisotropy, heterogeneity, and modification to include solutions for a surface shear distribution using a Boussinesq–Cerruti solution [10].

Acknowledgments We acknowledge Graham Wabiszewksi (University of Pennsylvania) for the MEMS surface images, the Microelectronics Development Laboratory at Sandia National Laboratories for the samples, Matthew A. Hamilton (Exactech, Inc), and W. Gregory Sawyer (University of Florida) for useful discussions. This work was partially supported by the National Science Foundation, grant CMMI 1200019, and by the US Department of Energy, BES-Materials Sciences, under Contract DE-FG02-02ER46016.

References

1. Greenwood, J.A., Williamson, J.B.P.: Contact of nominally flat surfaces. Proc. Roy. Soc. Lond. A **295**, 300–319 (1966)
2. McCool, J.I.: Comparison of models for the contact of rough surfaces. Wear **107**, 37–60 (1986)
3. Yan, W., Komvopoulos, K.: Contact analysis of elastic–plastic fractal surfaces. J. Appl. Phys. **84**, 3617–3624 (1998)
4. Majumdar, A., Bhushan, B.: Fractal model of elastic–plastic contact between rough surfaces. ASME J. Tribol. **113**, 1–11 (1991)
5. Hyun, S., Pei, L., Molinari, J.F., Robbins, M.O.: Finite-element analysis of contact between elastic self-affine surfaces. Phys. Rev. E **70**, 26117 (2004)
6. Persson, B.N.J., Bucher, F., Chiaia, B.: Elastic contact between randomly rough surfaces: comparison of theory with numerical results. Phys. Rev. B **65**, 184106 (2002)
7. Polonsky, I.A., Keer, L.M.: A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. Wear **231**, 206–219 (1999)
8. Persson, B.N.J., Albohr, O., Tartaglino, U., Volokitin, A.I., Tosatti, E.: On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. J. Phys. Condens. Matter **17**, 1–62 (2005)
9. Boussinesq, J.: Application des potentiels à l'étude de l'équilibre et du mouvement des solides élastiques (Application of potentials to the study of equilibrium and motion of elastic solids.). Gauthier Villars, Paris (1885)
10. Love, A.E.H.: The stress produced in a semi-infinite solid by pressure on part of the boundary. Proc. Roy. Soc. Lond. A **228**, 377–420 (1929)
11. Webster, M.N., Sayles, R.S.: A numerical model for the elastic frictionless contact of real rough surfaces. ASME J. Tribol. **108**, 314–320 (1986)
12. Poon, C.Y., Sayles, R.S.: Numerical contact model of a smooth ball on an anisotropic rough surface. ASME J. Tribol. **116**, 194–201 (1994)
13. Ren, N., Lee, S.C.: Contact simulation of three-dimensional rough surfaces using moving grid method. ASME J. Tribol. **115**, 597–601 (1993)
14. Liu, G., Wang, Q., Liu, S.: A three-dimensional thermal-mechanical asperity contact model for two nominally flat surfaces in contact. ASME J. Tribol. **123**, 595–602 (2001)
15. Dickrell, D.J., Dugger, M.T., Hamilton, M.A., Sawyer, W.G.: Direct contact-area computation for MEMS using real topographic surface data. J. Microelectromech. Syst. **16**, 1263–1268 (2007)
16. Delrio, F.W., De Boer, M.P., Knapp, J.A., Reedy, E.D., Clews, P.J., Dunn, M.L.: The role of van der Waals forces in adhesion of micromachined surfaces. Nat. Mater. **4**, 629–634 (2005)
17. Cook, R.D., Malkus, D.S., Plesha, M.E., Witt, R.J.: Concepts and Applications of Finite Elements Analysis, 4th edn. Wiley, New York (2001)
18. Plesha, M.E., Cook, R.D., Malkus, D.S.: FEMCOD—Program Description and User Guide. University of Wisconsin-Madison, Madison (1988)
19. Young, W.C.: Roark's Formulas for Stress & Strain, 6th edn. McGraw-Hill, New York (1989)
20. Borodachev, N.M.: Impression of a punch with a flat square base into an elastic half-space. Int. Appl. Mech. **35**, 989–994 (1999)
21. De Boer, M.P., Luck, D.L., Ashurst, W.R., Maboudian, R., Corwin, A.D., Walraven, J.A., Redmond, J.M.: High-performance surface-micromachined inchworm actuator. J. Microelectromech. Syst. **13**, 63–74 (2004)
22. Zhuravlev, V.A.: On question of theoretical justification of the Amontons–Coulomb law for friction of unlubricated surfaces. Zh. Tekh. Fiz. **10**, 1447–1452 (1940)