Tribology Letters, Vol. 19, No. 1, May 2005 (© 2005)
DOI: 10.1007/s11249-005-4263-8

Multiscale roughness and modeling of MEMS interfaces

C.K. Bora[a,*], E.E. Flater[a], M.D. Street[a], J.M. Redmond[b], M.J. Starr[b], R.W. Carpick[a] and M.E. Plesha[a]

[a]*Dept. of Engr. Physics, University of Wisconsin–Madison, 1500 Engineering Dr., Madison, Wisconsin 53706*
[b]*Sandia National Laboratories, Structural Dynamics Research, P.O. Box 5800, Albuquerque, New Mexico 87185*

Received 4 July 2004; accepted 2 March 2005

Investigation of contact and friction at multiple length scales is necessary for the design of surfaces in sliding microelectromechanical system (MEMS). A method is developed to investigate the geometry of summits at different length scales. Analysis of density, height, and curvature of summits on atomic force microscopy (AFM) images of actual silicon MEMS surfaces shows that these properties have a power law relationship with the sampling size used to define a summit, and no well-defined value for any is found, even at the smallest experimentally accessible length scale. This behavior and its similarity to results for fractal Weierstrass-Mandelbrot (W-M) function approximations indicate that a multiscale model is required to properly describe these surfaces. A multiscale contact model is developed to describe the behavior of asperities at different discrete length scales using an elastic single asperity contact description. The contact behavior is shown to be independent of the scaling constant when asperity heights and radii are scaled correctly in the model.

KEY WORDS: multiscale, fractal, roughness, surface, summit, asperity, modeling, MEMS, AFM

1. Introduction

The ability to design reliable MEMS devices with sliding surfaces in contact depends on knowledge of contact and friction behavior at multiple length scales. While friction at macroscopic scales can be modeled with Amontons' law, as the dimensions of a structure become smaller, the importance of surface roughness and surface forces (e.g., adhesion) are magnified and frictional behavior can change [1].

Surface roughness plays a crucial role in contact and friction between surfaces. For describing the roughness of a surface, statistical parameters for the surface height distribution function, i.e., root-mean-square (RMS) height, slope and curvature, have been used in several studies. These parameters can be directly related to the density of summits, summit curvature, and standard deviation of the summit height distribution function, which are the key inputs to models of rough contact based on the Greenwood-Williamson approach [2–4]. However, these parameters, if determined experimentally, can vary with sample size and instrument resolution [5,6].

Examination of rough surfaces shows that they often have multiscale features. That is, when a section of a rough surface is magnified, smaller scales of roughness appear. This general characteristic of surfaces was recognized long ago by Archard who described an engineering surface as consisting of "protuberances on protuberances on protuberances" [6,7]. Further, roughness at smaller scales has been shown to be similar to that at larger scales, but usually with a different scaling of length and height [8,9], a property known as self-affinity. The self-affinity of a shape at different length scales is a property displayed by fractal models for surface topography.

A surface is fractal when it is too irregular to be described in traditional geometric language, when it has detail on arbitrarily small scales, and when it has a structure that repeats itself throughout all length scales [10]. While there are no true fractals in nature that range over infinitely small to infinitely large length scales, most natural surfaces show multiscale geometrical characteristics, having roughness over multiple length scales that frequently span many orders of magnitude.

The power spectral density (PSD) describes the frequency content (in this case, spatial frequencies) of a set of data. It is defined as the Fourier transform of the autocorrelation function of a profile [10]. An equivalent definition of PSD is the squared modulus of the Fourier transform of the data itself, scaled by a proper constant term. Figure 1 shows two atomic force microscope (AFM) topographic images from the same region of a polycrystalline silicon MEMS surface at two different magnifications, and the respective averaged PSD of both. The PSDs are seen to correlate well over their shared length scales of measurement.

The fractal dimension of a surface can be extracted from its PSD. The PSD of a fractal surface profile can be related to its fractal dimension by:

*To whom correspondence should be addressed.
E-mail: bora@cae.wisc.edu

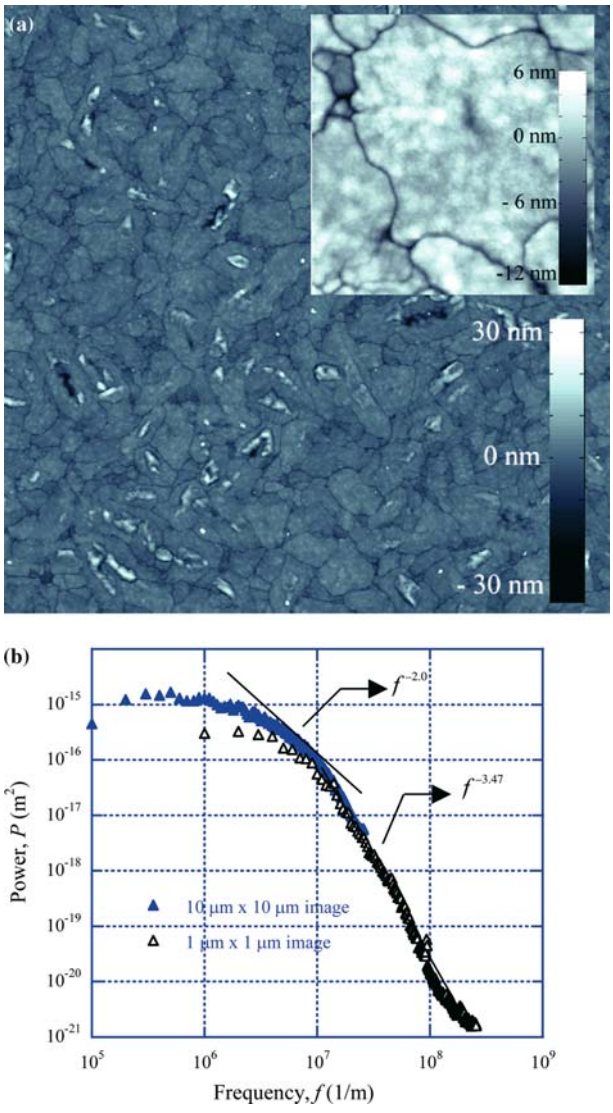

Figure 1. (a) 512 × 512 pixel AFM images of the same region of a polycrystalline silicon surface with RMS roughness ~3 nm, taken at 10 μm and 1 μm (inset) scan sizes. (b) Power spectral density of the two AFM images.

$$P(f) = \frac{C}{f^{(5-2D)}}, \tag{1}$$

where C is a scaling constant and D is the fractal dimension of a profile vertically cut through the surface [8,10,11]. For a physically continuous surface, we will obtain $1 < D < 2$.

The PSDs shown in figure 1(b) are not linear at low frequencies, and over the full range of frequencies they do not give a single value for the dimension of its fractal function representation. This surface could be characterized as a multiple-fractal, where for frequencies less than 1/100 nm^1 (10^7 m^{-1}), the PSD has varying slope. For the location shown in figure 1(b) (10^7 m^{-1}), the slope of −2.0 in the log-log graph gives a fractal dimension of 1.5. In the region where the slope is −3.47, equation (1) gives $D = 0.77$, which is a physically impossible value for any real continuous surface. Because the PSD in figure 1(b) is not linear, except for high frequencies, and because the linear region corresponds to an unobtainable fractal dimension for a real surface (i.e., $D < 1$), it is not possible to characterize this particular surface with a fractal function. Nonetheless, the PSD reveals that the surface has roughness at all length scales sampled, and furthermore it indicates that the effects of the multiple scales of roughness on mechanical contact phenomena should be taken into consideration [8]. A methodology that goes beyond using the fractal representation is needed.

In this study, we discuss the scale dependence of the average height, the average radius of curvature, and the density of summits on an actual polycrystalline silicon MEMS surface. Also, the relationship between the scale dependences and the fractal dimension of the surfaces is investigated. In the second part, a straightforward multiscale contact model (comparable to Archard's idea [6,7]) is developed, where the asperity force distributions and the contact area are determined as a function of the length scale using the elastic Hertz contact model at each length scale.

2. Analysis of surfaces

Determination of the heights, locations, and curvatures of summits on contact surfaces is necessary to model contact and friction. While this would seem to be a straightforward task, the multiscale roughness properties of real surfaces make the concept of a "summit" ambiguous and imprecise. Consider a surface profile whose height is determined at a finite number of discrete positions, such as by profilometry. A "peak" or "summit" can be defined as a location where the height is a local maximum. In the case of a two dimensional surface (i.e., a line trace), a sample point is a "peak" if its height exceeds that at each of its two neighboring sample points, while for a three dimensional surface (i.e., an AFM image), a sample point is a "summit" if its height exceeds that at each of its eight neighboring sample points. Ambiguities arise when more sampling points are used, i.e., when the definition of a summit is changed to require that a pixel is higher than a larger region of its surrounding neighborhood. To study this phenomenon, we use AFM topographic imaging like that shown in figure 1(a) to obtain a pixelized height distribution for a Si MEMS surface from Sandia National Laboratories SUMMIT process [12]. The AFM images were acquired in contact mode using a Digital Instruments Multimode AFM with a Nanoscope IV controller with a silicon nitride cantilever having nominal force constant of ~0.05 N/m. The piezo scanner was calibrated using the manufacturer's recommended procedure. The tip shape was tested before and after the measurements using *in-situ* tip imaging samples (Aurora Nanodevices, Edmonton, Canada) to ensure that it started and

remained a sharp, single protrusion of radius <30 nm, so as to minimize the effect of convolution of tip shape. Numerous tips with blunt, multiple, or asymmetric terminations were rejected. Low loads (in the adhesive regime) were used to minimize the contact area and enhance the lateral spatial resolution. The lateral spatial resolution of a contact mode AFM image is approximately determined by the contact diameter, which we estimate to be of the order of ~2 nm, comparable to the size of one pixel in our highest resolution images analyzed.

A MATLAB routine is then used to determine the heights and locations of summits by examining each sample point (pixel) with coordinates (x, y) and comparing its height z to the heights of n neighboring pixels, where n is called the *neighborhood size*. For a given value of n, the region of neighbors surrounding a particular sample point is a square with size d_n by d_n, where the *box size* d_n is given by:

$$d_n = (2n+1)\frac{L}{N}, \tag{2}$$

where L is the physical width of the square AFM image and N is the number of pixels per each side of the image. When a summit is found for a given n, the MATLAB routine determines the least squares best fit elliptic paraboloid to the data around the point, to determine the curvature of the summit in two dimensions. The major axes of the paraboloid are constrained to fall along the x and y axes of the image, and the maximum is constrained to occur at the point (x, y) with height z.

The average height, average radius of curvature and number of summits are calculated as a function of box size d_n for 1 μm × 1 μm and 10 μm × 10 μm ($L = 1$ and 10 μm, respectively) AFM images of a polycrystalline silicon surface with an overall RMS roughness of 3 nm (as measured for a 10 μm × 10 μm region), with the results shown in figure 2. Note that only a subset of pixels within a central square region of the image is considered when searching for summits, such that the same portion of the data is considered for all neighborhood sizes.

In a log–log plot, the radius and number of summits are seen to be almost perfectly linear functions of the box size d_n over almost three orders of magnitude of size, whereas the average height plot is slightly bilinear, changing slope around $d \sim 100$–200 nm, which is close to grain size observed in figue 1(a). Most strikingly, even at the small box sizes, (~10 nm), the distributions have not converged to well-defined values. Scans over smaller regions show this behavior continues for even the smallest box size we were able to consider ($d = 2$ nm), which is reaching the lateral resolution limit of the AFM itself. The power law dependence of the summit geometry and density seen in figure 2 illustrates the scale dependence of these values.

In contrast to a real MEMS surface, figure 3 shows the average summit radius of curvature vs. neighbor-

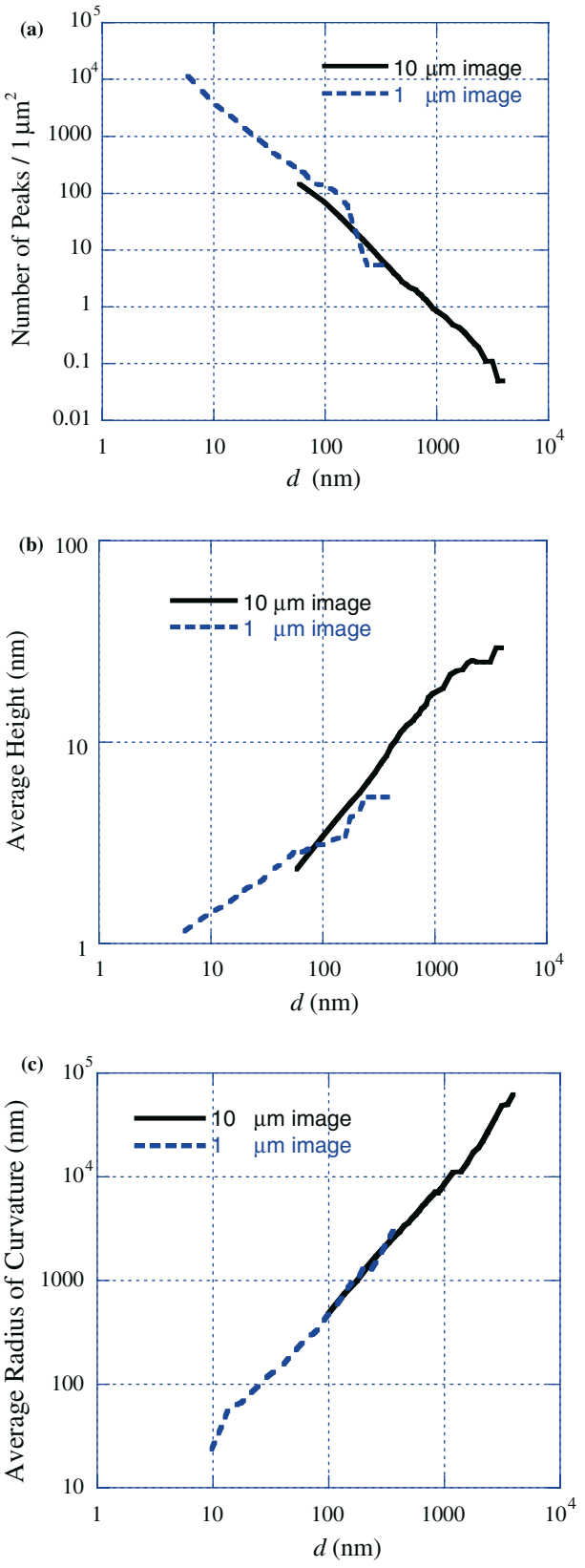

Figure 2. (a) Number of summits per area, (b) average height, and (c) average radius of curvature vs. the box neighborhood size d for an AFM image of a polycrystalline silicon surface. The RMS roughness of the surface was measured to be 3 nm for a 10 × 10 μm AFM image. The smallest data point in the radius plot is omitted due to an algorithmic artifact.

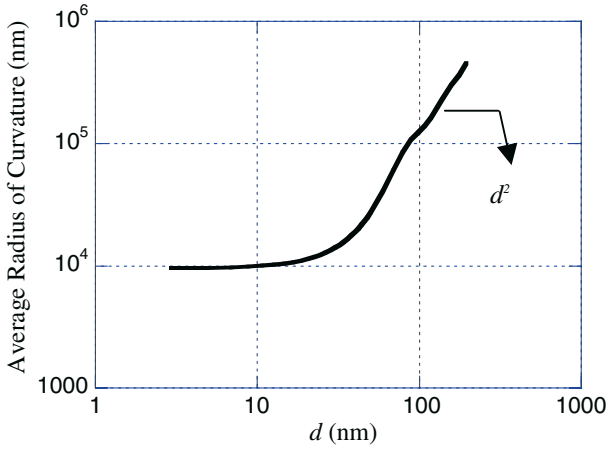

Figure 3. Average summit radius of curvature vs. neighborhood size for a surface with a single scale of roughness (a sinusoidal function).

hood size for a hypothetical surface with sinusoidal shape, which is created by:

$$z(x,y) = \cos(x/10)\cos(y/10)\text{nm}. \tag{3}$$

This surface obviously has a single scale of roughness. As expected, figure 3 shows that the average curvature becomes constant for sufficiently small neighborhood sizes, unlike the MEMS surfaces. For box sizes larger than the period of equation (3) (i.e., $d \geq 20\pi$), the slope of the power law approaches 2.0. Because our method of fitting a paraboloid requires the fit to have the same z coordinate as the summit in the data, this response for large values of d is simply a mathematical artifact.

In other words, fitting a paraboloid to a point at fixed height while the lateral extent of the fit is enlarged will produce a power law of 2 if the surface is nominally flat at large length scales. This can be shown analytically in three dimensions using a least squares fit of a paraboloid to the sinusoidal surface,

$$\frac{d}{dr}\left[\int_{-b}^{b}\int_{-b}^{b}\left(\left(1-\frac{x^2}{2R}-\frac{y^2}{2R}\right)-(\cos(x)\cos(y))\right)^2 dydx\right]=0 \tag{4}$$

where R is the paraboloid's radius of curvature and b is the sampling length (essentially equivalent to d). Solving for R, its relation to b is found to be:

$$R \propto b^2. \tag{5}$$

This analysis helps validate the curvature calculation in our MATLAB routine since the analytical and MATLAB results agree. It also shows that the power-law dependence of the radius of curvature calculation is not by itself an indication of multiscale surface roughness. However, the *lack of convergence* of the radius of curvature to a fixed value at small neighborhood sizes for the MEMS surfaces is indeed an indication of multiscale character, in this case down to a length scale of ~2 nm.

To compare the foregoing results to those for a model fractal mathematical surface, our analysis procedure (i.e., pixelation of a continuous map of heights followed by pixel-by-pixel analysis using our MATLAB program) was applied to three-dimensional fractal surfaces including the following equation:

$$z(x,y) = C\sum_{m=1}^{M}\sum_{n=0}^{n_{\max}}\gamma^{(Ds-3)n}\left\{\cos\Phi_{m,n} - \cos\left[\frac{2\pi\gamma^n(x^2+y^2)^{1/2}}{L}\cos\left(\tan^{-1}\left(\frac{y}{x}\right)-\frac{\pi m}{x}\right)+\Phi_{m,n}\right]\right\}$$

$$C = L\left(\frac{G}{L}\right)^{Ds-2}\left(\frac{\ln\gamma}{M}\right)^{1/2} \tag{6}$$

which is a form of a multivariate Weierstrass-Mandelbrot (W-M) function developed by Ausloos and Berman [13] and later used in a three dimensional elastic-plastic contact model by Yan and Komvopoulos [14].

Equation (6) is constructed by taking a two dimensional fractal profile as a "ridge", and then superposing a number of these ridges at different angles to achieve randomization. Φ is an array of random numbers to generate phase and profile angle randomization, M is the number of ridges, and L is the image size in length units. G is the roughness coefficient which is used to correctly scale the height of the function to fit the modeled surface D_s is the fractal dimension of the surface ($2 < D_s < 3$, as the relation between the fractal dimension of a surface D_s and that of a profile D is $D_s = D + 1$ for an isotropic surface). γ is a parameter that governs the frequency and amplitude ratio of successive cosine shapes ($\gamma > 1$) and thus represents that relative frequency separation of successive terms in the W-M function. $L_{\max}$ is the sample size. Finally, $n_{\max}$ is the number of cosine shapes added for a profile. Note that equation (6) is perfectly fractal only if $n_{\max} \to \infty$. For practical applications, finite values of $n_{\max}$ are used, so that cosine shapes with periods larger than $L_{\max,}$ and smaller than $L_{\min}$ are not needed.

Figure 4 shows two sample W-M surfaces. In figure 4(a), the popular but otherwise unremarkable value of $\gamma = 1.5$ was used, which results in successively added cosine shapes whose periods and amplitudes are moderately spaced apart, providing a surface with a seemingly "random" aesthetic character. When $\gamma = 5$, the difference between the periods and heights of consecutive cosine shapes is greater. This coarse separation of roughness scales leads to the easily discernable scales of "bumpiness" seen in figure 4(b). For the W-M function, an appropriate number of the cosine shapes used for defining the surface ($n_{\max}$) can be selected using:

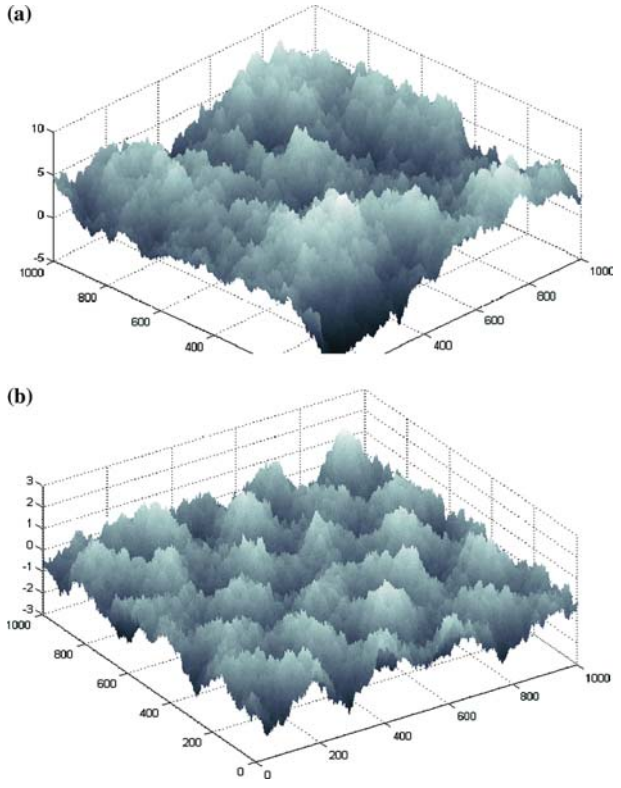

Figure 4. Two W-M surfaces produced using $D_s = 2.4$, $G = 1.36 \times 10^{-2}$ nm, $L_{max} = 1000$ nm, $L_{min} = 1$ nm, $M = 10$, $n_1 = 1$. $\gamma = 1.5$ for (a) and $\gamma = 5$ for (b). All values other than γ are taken from Yan and Komvopoulos 14]). The height scale is in nm.

$$n_{max} = \left[\frac{\log L_{max}/L_{min}}{\log \gamma}\right] \quad (7)$$

where the "int" function funcates the value within [ ] (i.e. rounds to the next lowest integer), L_{min} is the period of the smallest cosine shape. Equation (7) assures that the cosine shapes have periods that fully span the length scales from L_{min} to L_{max}. Thus, the function is "fractal" for all practical purposes. When L_{min} and L_{max} are 1 nm and 1 μm, respectively, 18 cosine terms would be needed for $\gamma = 1.5$, and only 5 cosine shapes would be needed for $\gamma = 5$. The frequencies are spaced further apart in the $\gamma = 5$ case. Several W-M surfaces were then created with fractal dimensions varying from 2.01 to 2.99, and with the same two values of γ (1.5 and 5) used to explore the effect of spatial frequency separation. For W-M surfaces with small γ values, the number of summits per area, average height, and average radius of curvature variation with neighborhood size give almost perfectly linear distributions on log-log plots, down to the smallest scales. Examples of the summit density, average height, and radius of curvature as a function of d for both values of γ are shown for the case of $D_s = 2.4$ are shown in figures 5(a), (b), and (c), respectively. For W-M surfaces with high γ values, these plots show deviations from linearity, which is due to the large frequency separation between each consecutive cosine term used in the

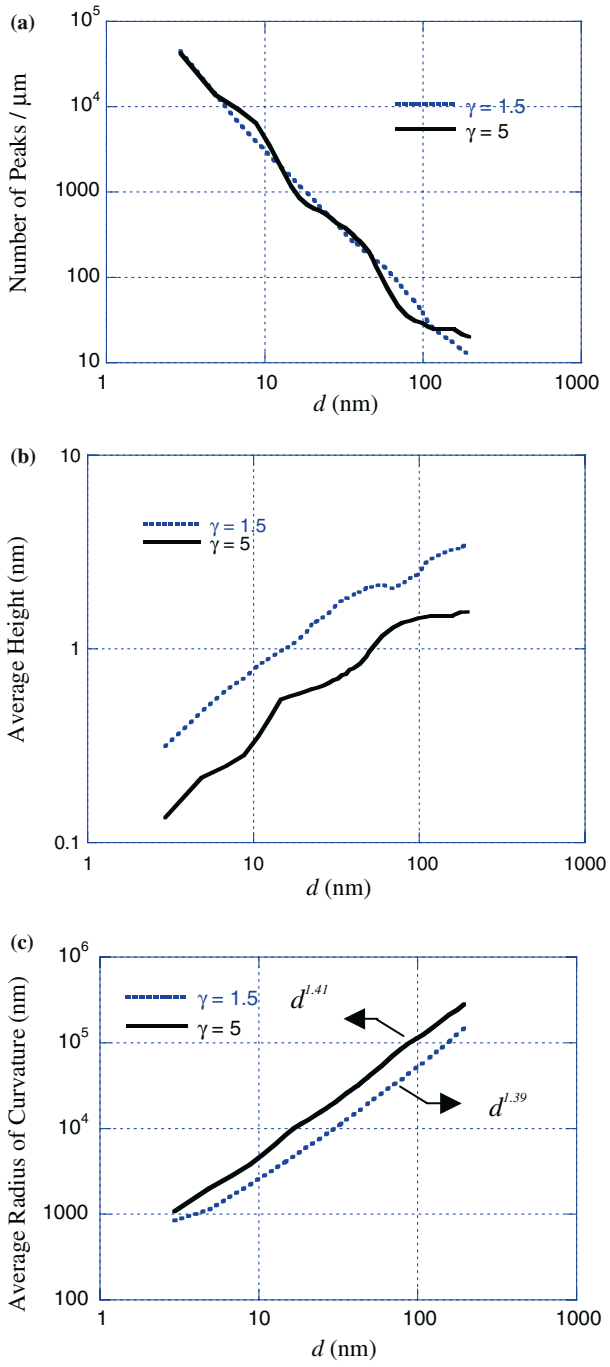

Figure 5. (a) Density of summits, (b) average height, and (c) average radius of curvature vs. box size d for W-M surfaces crated using $\gamma = 1.5$ and $\gamma = 5$, using $D_s = 2.4$ (thus $D = D_s - 1 = 1.4$).

generation of the surface. This is clearly seen in the plot of the number of peaks vs. d (figure 5(a)) and average heights vs. d (figure 5(b)). The average radius of curvature (figure 5(c)) shows little effect of the frequency separation.

The results of our "summit search" method match certain analytical predictions. Majumdar and Bhushan [15] derived the radius of curvature R at the tip of an asperity for the W-M surface as:

$$R = \frac{d^D}{\pi^2 G^{D-1}}. \quad (8)$$

Using this expression with the box size d as the contact length and with a constant value of G, the radius of curvature R changes as d^D, which matches the slopes found in figure 5(c).

Wu [16] argues that the W-M function given in equation (6), developed by Ausloss-Berman, is not exactly a 3D extension of the fractal 2D W-M function, since a vertical cut of the surface is not necessarily a W-M function, and thus the surface is not isotropic. He later shows that, despite this observation, surfaces generated by this function share very similar properties, such as summit curvature, with other fractal functions such as the successive random addition method [16]. Thus, our comparison to the W-M function can be considered as a reasonable way to illustrate the fractal character of the actual MEMS surfaces.

Figure 6 shows the exponents of power law fits (i.e. the slopes on the log-log plots from figure 5) for the number of summits, average height, and average radius of curvature distributions of the W-M surfaces with varying fractal dimensions for the two values of γ. A relation between the fractal dimension and the power law exponents of these fits is seen. At high fractal dimensions, the average radius of curvature changes more rapidly with varying neighborhood size (higher slope value). This effect is reversed for the corresponding behavior of the density and average height of the summits. In other words, a high fractal dimension means smaller variance in the density and average height with changing neighborhood size. The change in γ affects the results somewhat, but further analysis is required to fully understand this effect. For fractal dimensions D between 2 and 2.5, the relation between R and d shown in figure 6(c) follows the behavior described in equation (8) extremely well, while for higher fractal dimensions the slopes are slightly lower than expected.

In figure 6, the range of power law exponents obtained from the "summit search" analysis of several AFM images of rough polycrystalline silicon MEMS surfaces are shown as shaded bands running across the graphs. We see that these values are consistent with the W-M fractal surface properties if we associate the AFM images with low fractal dimensions ($D_s = 2.1$–2.3). The fractal dimensions of these AFM surfaces obtained using PSD analysis varied for different length scales as discussed earlier, with values ranging from non-fractal values (e.g. $D_s = 1.77$ as in figure 1) up to approximately $D_s = 2.5$.

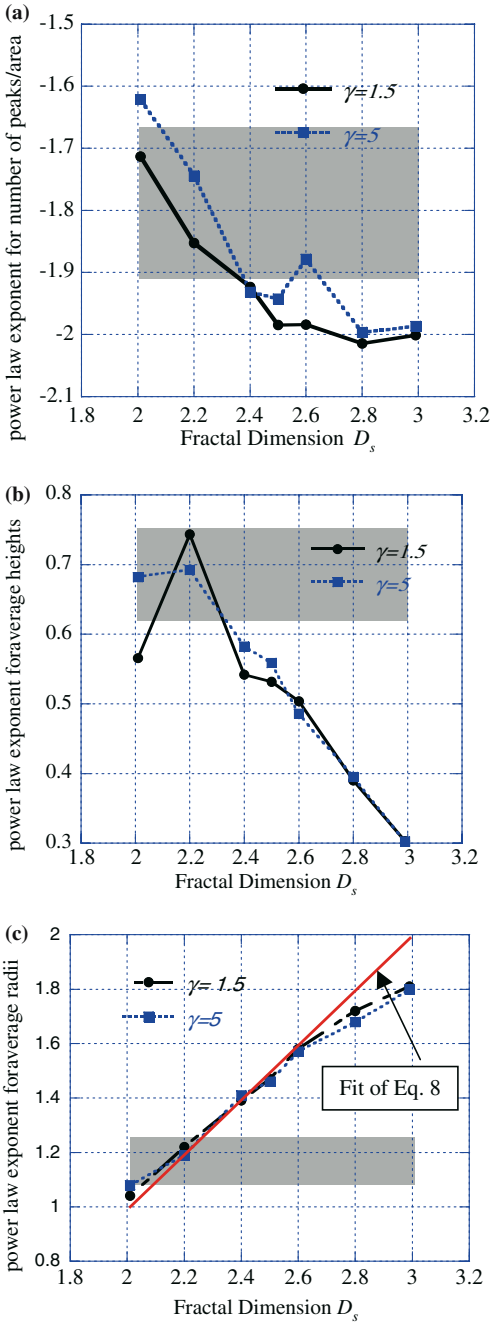

Figure 6. Exponents of the power law fits for the number of summits, average summit height, and average summit radius of curvature vs. neighborhood size for W-M surfaces, plotted for a range of fractal dimensions D. The shaded band running across each figure shows the range of exponents obtained from the analysis of actual AFM images.

3. Contact model

The multiscale nature of the MEMS surfaces revealed by our analysis suggests that any useful contact model must embody this multiscale character. Thus, in our model, surfaces in contact are modeled with roughness at multiple length scales. The asperities at the largest length scale have the largest radii and height variation, and upon these lies a second set of asperities with radii and height variations smaller by a factor $s > 1$ which we call the scaling constant, and so on. Contact between two rough surfaces was approximated by contact be-

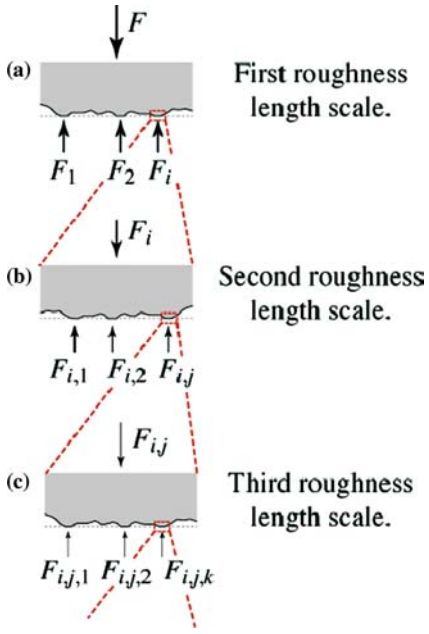

Figure 7. Hierarchy of roughness and load distribution among asperities at different length scales for the contact model we have developed.

tween a smooth rigid surface and a single rough elastic surface. This can be adjusted to represent the behavior of two rough surfaces as desired [4].

Figure 7 shows how the total force on the first set of asperities is divided into forces on asperities at smaller length scales. If only the first scale were considered, then the surface could be thought of as a Greenwood-Williamson [2,3] surface with a given height distribution. In any event, the total force is proportioned among each contacting asperity at that length scale using an appropriate single asperity contact model.

In this algorithm, the rough surface is incrementally advanced into the rigid flat countersurface to determine the asperities at the coarsest length scale that make contact. Then the exact approach distance is found by interpolation via Newton's method. This distance is used to proportion the total force among all of the contacting asperities, using an appropriate single asperity contact model as discussed below. Thus, the total load F is equal to the sum of the forces supported by the contacting asperities at that length scale according to:

$$F = \sum_{i=1}^{n_1} F_i, \tag{9}$$

where n_1 is the number of asperities at the first length scale that are in contact. The actual contact area at this length scale is given by:

$$A_1 = \sum_{i=1}^{n_1} A_i. \tag{10}$$

At the next length scale, each of the loads F_i is proportioned among the second generation asperities that make contact such that the forces and areas are given as:

$$F_i = \sum_{j=1}^{n_{2,i}} F_{i,j} \tag{11}$$

$$A_i = \sum_{i=1}^{n_1} \sum_{j=1}^{n_{2,i}} A_{i,j} \tag{12}$$

where $n_{2,i}$ is the number of second generation asperities in contact that are located on first generation asperity i.

For example, considering this second scale of roughness, the total force supported is obtained by combining equation (9) and equation (11), whereas the total contact area is given by equation (12). The forces and contact areas at subsequent length scales are calculated in the same fashion. This procedure can be carried out for any number of desired scales of roughness.

When no adhesion is included in the model, the Hertz contact model [17] is appropriate. The single asperity contact area-load relation obtained from the Hertz model is:

$$A = \pi \left(\frac{3R}{4E^*} \right)^{2/3} L^{2/3}. \tag{13}$$

In this equation, R is the asperity radius, E^* is the composite elastic modulus of the contact, and L is the total load on the asperity.

There are several significant assumptions in our hierarchical modeling approach. The first is that the asperities are elastic. The second is that, within one length scale, the mechanical response of an asperity is not affected by its neighbors. The third is that the roughness of finer length scales is small enough so that the response of asperities at coarser length scales is unaffected. This imposes limitations on the applicability of our model, but allows us to capture and interpret the effects of multiscale contact within these limitations. In fact, by using sufficiently well-spaced generations of asperities, the use of the Hertz model is justified, whereas such an assumption is questionable in the case of extremely finely-spaced scales of roughness, such as in the W-M function with small γ values.

A surface that appears to satisfy the third assumption of our model is the Weierstrass-Mandelbrot function, in the form given in equation (6), with sufficiently high γ value ($> \sim 4$) to obtain adequate frequency (asperity size scale) separation. The W-M function with G and D values given in figure 4 and with $\gamma = 10$ was used to create such a surface. Using an appropriate neighborhood size, the locations, heights, and curvatures of the asperities produced by the largest cosine function were found by using the "summit search" algorithm described above.

To investigate the effect of changing the scaling constant, summits obtained from the W-M surface described in the previous section were again used as the

first scale of roughness, and different scaling constants $s > 1$ were used to define the roughness at smaller length scales. For example, for $s = 10$, the asperity heights at the second order of roughness are one-tenth of the height of those at the first order, and have radii of curvature that are one-tenth of those at the first order. In the calculation $L = 10\ \mu$N and $E^* = 200$ GPa was used. This E^* value is in the upper range of reported values for polycrystalline silicon. Figure 8(a) shows the computed total contact area as a function of the number of roughness scales, using different scale constants s. For a given value of n, the lateral length scale of the asperity size depends on s. For illustration purposes, we replot the data in figure 8(a) using a common dimensionless distance axis defined as $1/s^{n-1}$. Thus the largest scale has an asperity dimension of 1.

Figure 9 shows the total area of contact versus the total load applied to the surface, calculated for five roughness scales using a scaling constant of $s = 5$. The behavior is very close to linear, and a power law fit shows that the area is related to the total load by a power of 0.94. This is reminiscent of the well-known result from the G-W model whereby the contact area scales nearly linearly with load for a collection of equally-sized asperities randomly distributed about a mean height. The difference in the two models is that G-W model uses a single scale of roughness [2]. In 1940, Zhuravlev [3], making assumptions similar to G-W model, found a similar dependence, namely $A \propto L^{0.91}$.

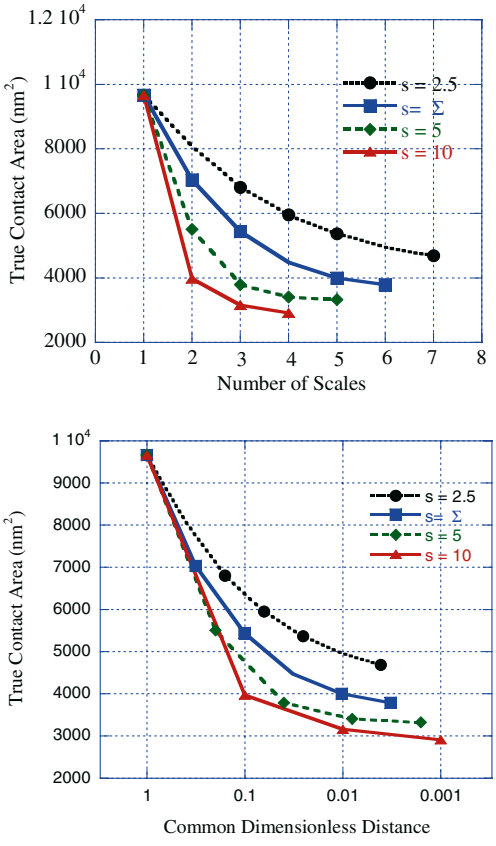

Figure 8. (a) True contact area as a function of the number of roughness scales for non-adhesive (Hertzian) asperities, computed using different scale constants. ($E^* = 200$ GPa, $L = 10\,\mu$N) (b) True contact area plotted vs. the common dimensionless distance associated with the scales.

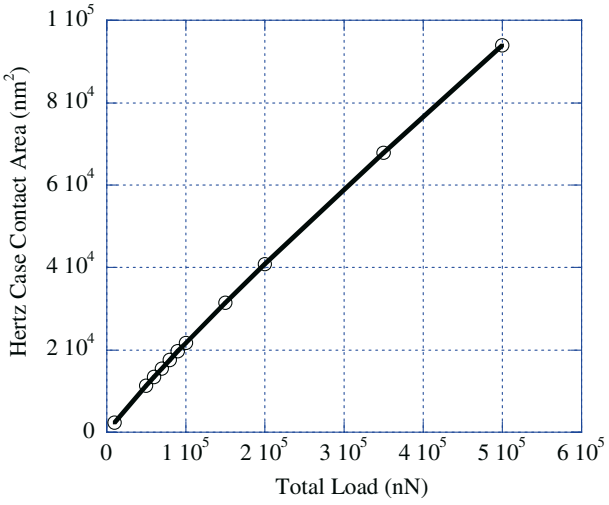

Figure 9. True contact area as a function of total load, using the model with five scales of roughness. The circles are the individual results of the model; the solid line indicates the power law fit.

For all scaling factors tested, the total contact area shown in figure 8 decreases with increasing scales of roughness, and appears to converge to well-defined values, but those values are highly dependent on the particular scaling factor chosen. In Appendix 1, we further illustrate this effect using a simple calculation for a set of Hertzian contacts at the same height. This effect will be further discussed in the next section.

3.1. Constraints on smaller scale roughness features

A constraint that occurs with real surfaces is that the number of smaller asperities that can be present on a contacting asperity (i.e., a host asperity) at the larger length scale is limited. In the example presented in figure 8, the finer scale contacts were assumed to be hexagonally close-packed on the contact area of the host asperity, and a limit was introduced according to the ratio (m) of the large asperity contact area to the average of the small asperity contact areas. The number of small asperity contact spots (N_{max}) that would fit on the larger contact area can be estimated by $N_{max} \propto m^2$, when m is greater than 3. We call this the *close-packed* constraint, and it imposes an upper bound to the possible numbers of asperities at successive scales.

Another limitation of the example presented in figure 8 is that the same scaling constants are used for both the heights of the asperities and radii of curvature for the asperities. In reality, the heights and radii scale differently, as seen in the surface analysis results shown in figure 2. The data of figure 2 suggest a better, more realistic way to model multiscale roughness driven by

the experimental observations of the surface properties, as follows. The trends seen in figure 2 give scaling constants for the heights and radii for the AFM image analyzed. The corresponding power relations are $h \propto d^{0.6}$, $N \propto d^{-1.87}$, and $R \propto d^{1.25}$. With s used as the scaling constant for length, the neighborhood size in figure 2 is scaled by $d_2 = d_1/s$. The asperity heights for the subsequent smaller scale are obtained from the heights of the previous scale by $s_h = s^{0.6}$ so that $h_2 = h_1/s_h$. Similarly, the radii are scaled by $s_R = s^{1.25}$, and the number of asperities per unit area scales by $s_N = s^{-1.87}$. We call this the *asperity-density* constraint, which originates from real experimental analysis.

The information regarding number of asperities per unit area provides a more realistic value for the number of contacts that will be present on the tip —or contact area— of the larger scale host asperity. If we know the contact area at a particular length scale, then we can multiply this with the asperity density at the smaller length scale to obtain a limiting value for the number of asperities that can be present on the host asperity.

We select from the AFM image a set of heights and radii by selecting a neighborhood size which yields a reasonable number of asperities (e.g., $N \propto 100$), and we call this the *roughness template*. Then the scaling constants for h, R, N can be used to calculate the geometry and density of asperities at other scales. In other words, the roughness template is scaled up to provide coarser details of roughness, and is scaled down to obtain finer details of roughness.

The silicon MEMS surface analyzed in figures 1 and 2 was used as an example. The template scale was found using a summit search box size of ~450 nm, which yielded 103 summits. The summits obtained show a distribution close to an exponential, which is the same distribution used in previous models in literature such as the G-W model [2].

The power law relations cited earlier were used to calculate scaling factors for radii, height and asperity density for different s values. A load of 1 mN and a modulus of $E^* = 200$ GPa was used. Figure 10 shows the prediction of the true contact area as a function of the smallest length scale used in the computation. In figure 10, filled marks represent the results where the number of sub-scale contacts is constrained by a close-packed distribution assumption, and empty marks represent the results where the experimentally obtained asperity density constraint method is used.

The calculations shown in figure 10 are not continued to scales lower than about 1 nm, as this length approaches atomic spacing. Ten times the equilibrium atomic spacing is a reasonable estimate for the limiting value of the elasticity (Hertz) solution [14], and this provides an approximate value where to terminate the calculations (i.e., ~4 nm).

The estimated true contact area at the smallest length scale for the asperity-density approach (~10^4 nm^2) is

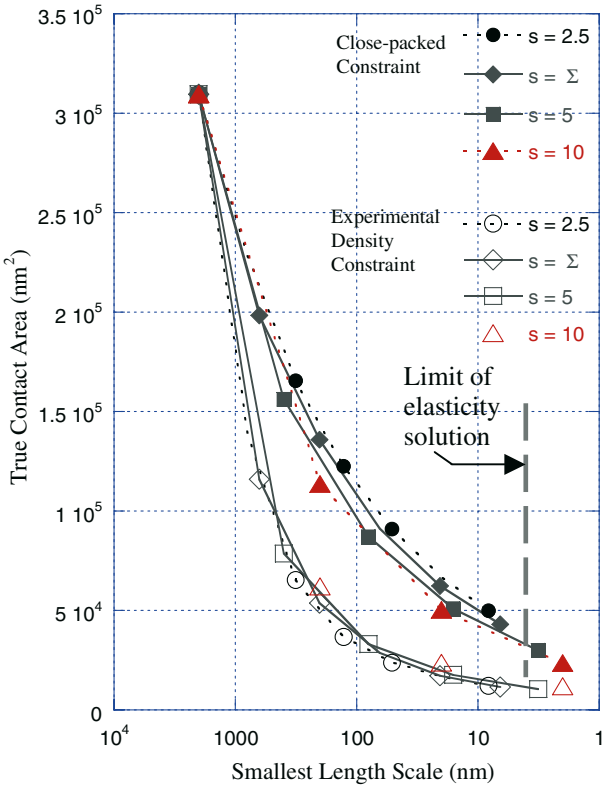

Figure 10. True contact area as a function of the smallest length scale used in the simulation, computed using different scale constants, with $E^* = 200$ GPa, $L = 1$ mN. Filled marks represent the results of the close-packed constraint, and empty marks represent the results from experimental density constraint for the number of sub-asperities.

smaller than the close-packed approach (~3×10^4 nm^2), as expected. It is also seen that the contact area converges to a limit faster in the asperity-density method. It was observed many times in the asperity density approach that there would be only one small asperity present on the tip of the larger host asperity.

When two consecutive scales are considered, the contact behavior of the intermediate size peaks between those two scales is being neglected. The case of the asperity-density constraint is affected more by this, as the number of small asperities that can be added to compensate for the lack of intermediate size asperities is limited. Thus the calculated area ends up being a lower bound estimate, which is imposed by the experimental data. The close-packed distribution constraint results in a theoretical upper-bound for the contact area that can be calculated in the model, as geometrically there can be no more asperities present in contact. So the results of our model can be taken as an estimated range for the real contact area.

In the calculation shown in figure 10, where the scaling constants for radii and heights are more realistic, we see that using different scaling constants ($s = 2.5$–5) does not affect the true contact area as strongly as in figure 8, where the scaling constants for radii and heights were the same. Independence of the contact area

from the scale constant shows that we can get away with a "crude" model where large scaling constants and only a few scales are used to represent the contact behavior. The model surface need not be carrying all the roughness frequencies, thus, for example, a Weierstrass-Mandelbrot surface with a high γ value can be used for simplicity.

The same AFM image is analyzed with the procedure described by McCool [4,18] which uses a Greenwood-Williamson approach, i.e., has one scale of roughness, with $E^* = 200$ GPa, $L = 1$ mN. The analysis gives a nominal asperity radius of 280 nm, summit density of 4.6×10^{-5} nm^{-2} and a summit height standard deviation of 2.55 nm. The true contact area estimate from this method is on the order of 10^5 nm^2 when the apparent contact is 10 μm by 10 μm. This analysis is described in [18] for the same surface.

The nominal asperity radius of 280 nm corresponds to a neighborhood size of ~70 nm in our "asperity search" analysis shown in figure 2. The contact area estimate of 10^5 nm^2 is close to the close-packed area constraint result for this length scale, but exceeds the total contact area we determine (by considering all length scales down to the atomic limit) by a factor of ~5. Another key difference between McCool analysis and our model is that we allow the force to be supported at smaller contact areas, resulting in higher stresses at contact points.

4. Conclusions

Roughness of polycrystalline silicon MEMS surfaces is strongly scale-dependent. Analysis of summits for AFM scans of actual MEMS surfaces shows that the height, density, and geometry of the summits as determined by a "search and fit" routine have a power law relationship with neighborhood search size.

The analysis of a test surface that has roughness limited to a small range of spatial frequencies, and analysis of a surface with a single scale of roughness, shows that the summit search procedure captures the geometry of the smallest summit features when a surface has a well-defined length scale below which there are no additional details of roughness. When the same analysis is performed on AFM data of a polycrystalline silicon MEMS surface, additional details of roughness emerge even at the smallest neighborhood sizes considered. In other words, no convergence to a uniform value for the height, density, and geometry of summits is observed even at the smallest experimentally accessible lateral length scale (~2 nm).

The power law behavior obtained from AFM images is similar to fractal W-M surface results. However, we find that these MEMS surfaces exhibit a range of spectral frequencies over which the surface is not fractal (the slope of the PSD is less than -3), yet the surface is still multiscale in nature (no well-defined summit radius, for example). Although the validity of the method still needs to be explicitly proven, the results indicate that as an alternative to the conventional power spectral density method [8] for determining the fractal dimension of a fractal surface representation, a "summit search" methodology, which is intuitively more straightforward and potentially more versatile, may be used for describing surface geometry. The appropriate method to select other parameters needed for analytical representations of fractal surfaces (such as the W-M function, where γ and G must be determined) have not yet been addressed here.

A contact model is developed using multiple length scales for roughness. The smaller roughness scales are successively modeled as asperities that are superposed on the asperities of the next larger scale. The total contact area predicted with elastic Hertz behavior approaches a limit with increasing number of roughness scales.

The calculated area of contact is dependent on the scale constant s that is used when the radii and the heights of the asperities are scaled with the same constant. When the correct scale dependence of the heights and radii are used, as obtained from the analysis of summits from AFM images, the contact area calculated does not depend on the scaling constant s. This is important as it shows that a simpler surface representation with large scale constants and fewer scales is still valid. This suggests as well that a large γ value in Weierstrass-Mandelbrot function can be used to generate a fractal surface model for simplicity. Separation of length scales renders the use of Hertzian contact mechanics across these length scales more readily believable.

The number of small scale contacts within a large contact area is constrained using two different methods: The experimental asperity density constraint and the close-packed distribution assumption. The latter gives a higher area estimate from the former. Together these two values can be thought as the upper and lower bound estimates for the contact area.

The next step in this modeling approach would be to include adhesion, i.e., JKR [19] and DMT [20] single asperity contact models. Preliminary work regarding these models shows that a technique must be used that accounts for the adhesion associated with the large asperities when considering a small scale of roughness. The adhesion models mentioned account for adhesion only at the contact area [19] or around it [20]. When looking at a roughness scale, the surface forces on the material between two asperities are disregarded. If this is not considered then a force imbalance occurs.

Plasticity also needs to be added to the model. First of all, the small asperities will likely experience yielding, and secondly the material properties, specifically the yield strength, may vary with different length scales.

The "summit search" method and the contact model presented constitute an intuitive approach to understand the multiscale nature of surfaces, making use of real images of MEMS surfaces, and numerical computation.

Acknowledgments

We acknowledge the staff at the Microelectronics Development Laboratory at Sandia National Laboratories for fabricating the MEMS samples, and Maarten P. de Boer and Alex D. Corwin for useful discussions and feedback. This work was supported by the US Department of Energy, BES-Materials Sciences, under Contract DE-FG02-02ER46016 and by Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US Department of Energy under contract DE-AC04-94AL85000.

5. Appendix

5.1. Decrease of contact area with increasing number of length scales

To further investigate the effect reported in Section 3, whereby the contact area decreases with an increasing number of length scales, we use a Hertzian contact model for the following two simple cases: (1) a single asperity interface consisting of one asperity with radius R contacting a rigid flat surface under load L, and (2) a multi-asperity contact where n identical asperities all at the same height and radius $r \leq R$ contact a rigid flat surface, again under load L. These are illustrated in figure A1. In the second case, the area-load relation in equation (13) becomes

$$A_n = n\pi\left(\frac{3r}{4E^*}\right)^{2/3}\left(\frac{L}{n}\right)^{2/3}. \qquad \text{(A1)}$$

Figure A2 shows the area ratio A_n/A from combining equation (13) and equation (A1), as a function of the radius ratio which is analogous to the scale factor s for the multiscale model. The number of asperities used in the calculation was $n = 100$. For a high scaling constant, i.e., $r \ll R$, the ratio for the n-asperity model is small, that is, the total contact area for 100 asperities is much smaller than for a single asperity at the same load. However, this area increases as r approaches R. In other words, for a large scale factor, dividing a single asperity into multiple asperities, with the same total load, decreases the total contact area significantly.

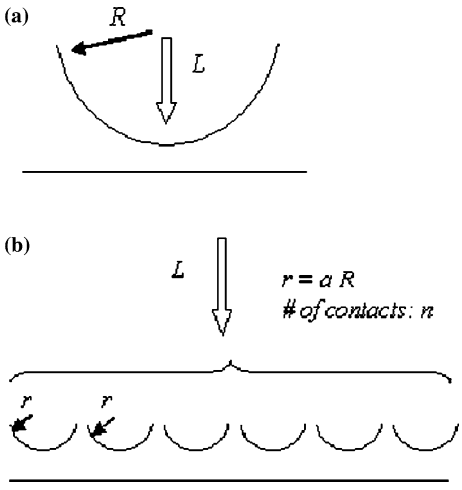

Figure A1. Two cases of contact: (a) a single asperity surface under load L, (b) multi-asperity contact with n identical asperities with same radii under L.

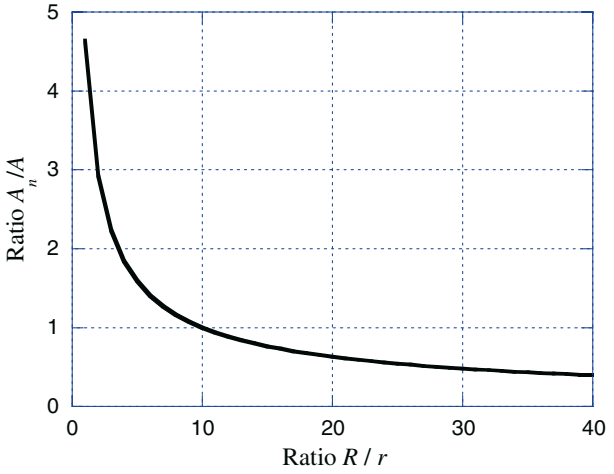

Figure A2. Area ratio A_n/A for the Hertz solution for multiple vs. a single asperity contact, plotted as a function of radius ratio, using $n = 100$ asperities at the same height.

It must be noted that if s (or, for this example, the radius ratio) becomes too small, the assumption of asperities at one scale not affecting asperities at the larger scale breaks down.

References

[1] R.W. Carpick and M. Salmeron, Chem. Rev. 97 (1997) 1163.
[2] J.A. Greenwood and J.B.P. Williamson, Proc. Roy. Soc. London A 495 (1966) 300.
[3] V.A. Zhuravlev Zh. Tekh. Fiz. J. Technical Phys.– translated from Russian by F.M. Borodich 10(17) (1940) 1447.
[4] J.I. McCool, Wear 107 (1986) 37.
[5] R.S. Sayles and T.R. Thomas, Nature 271 (1978) 431.
[6] J.F. Archard, Proc. Roy. Soc. London A 243 (1957) 190.
[7] J.A. Greenwood and J.J. Wu , Meccanica 36(6) (2001) 617.

[8] A. Majumdar and B. Bhushan, J. Tribol. 112 (1990) 205.
[9] B.B. Mandelbrot, *The Fractal Geometry of Nature* (W.H. Freeman, New York, 1982).
[10] K.J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications* (Wiley, New York, 1990).
[11] M.V. Berry and Z.V. Lewis, Proc. Roy. Soc. London A 370 (1980) 459.
[12] E. Garcia and J. Sniegowski, Sens. Actuators A 48 (1995) 203.
[13] M. Ausloos and D.H. Berman, Proc. Roy. Soc. London A 400 (1985) 331.
[14] W. Yan and K. Komvopoulos, J. Appl. Phys. 84(7)(1998) 3617.
[15] A. Majumdar and B. Bhushan, J. Tribol. 113 (1991) 1.
[16] J.J. Wu, Wear 239 (2000) 36.
[17] H. Hertz and J. Reine, Angew. Math. 92 (1881) 156.
[18] R.W. Carpick, E.E. Flater, J.R. VanLangendon, M.P. de Boer, Proc. of the SEM VIII International Congress and Exposition on Experimental and Applied Mechanics, (2002) 282.
[19] K.L. Johnson, K. Kendall and A.D. Roberts, Proc. Roy. Soc. London A 324 (1971) 301.
[20] B.V. Derjaguin, V.M. Muller, Y.P. Toporov and J. Colloid, Interface Sci. 53 (1975) 314.