

Elastic–plastic contact of a rough surface with Weierstrass profile

BY YAN-FEI GAO AND A. F. BOWER*

Division of Engineering, Brown University, Providence, RI 02912, USA

We analyse contact between a rigid cylindrical indenter and an elastic–perfectly plastic solid with a fractal surface roughness, which is idealized as a Weierstrass profile. A hierarchical scheme is used to calculate quantities such as the number of contact spots; their size; the true contact pressure acting on the contact spots, and the total true area of contact. In addition, we calculate critical conditions that will initiate plastic flow in the asperities, and determine the extent of this plastic flow as a function of the applied loading, the material properties of the deforming solid, and the surface roughness. An important conclusion of our study is that a perfectly fractal description of surface roughness appears to lead to unphysical predictions of the true contact size and number of contact spots, for both elastic and elastic–plastic solids. Possible approaches to resolving these difficulties are discussed.

Keywords: self-affine fractal; contact mechanics; surface roughness; plasticity

1. Introduction

When two rough surfaces are pressed into contact, they meet only at the highest points on the two surfaces. To predict the size of the resulting contact spots, the true contact pressure, and the true area of contact is a venerable problem in contact mechanics, and is clearly of considerable practical importance, since phenomena such as friction, wear, and contact fatigue are controlled by the nature of asperity contacts. Nevertheless, a fully satisfactory solution remains elusive.

Difficulties in modelling contact between rough surfaces arise because of the fractal nature of surface roughness. Early models, such as the classic work of Greenwood & Williamson (1966) and subsequent refinements by Whitehouse & Archard (1970), Nayak (1971), Bush *et al.* (1975), Greenwood (1984) and McCool (1986) among others, idealized a rough surface as a collection of asperities with spherical tips of equal radius, but with a Gaussian distribution of heights. These models predict a well-defined number of total contact spots, contact size, contact pressure and contact area, which depend on the curvature of the asperity tips, the asperity height distribution, and the elastic and plastic properties of the two contacting surfaces. The Greenwood–Williamson model has proved difficult to apply to real surfaces, however. Surface profiles are measured by sampling the

* Author for correspondence (allan_bower@brown.edu).

The electronic supplementary material is available at <http://dx.doi.org/10.1098/rspa.2005.1563> or via <http://www.journals.royalsoc.ac.uk>.

surface at discrete intervals. Asperity tips are identified by finding points that are higher than their neighbours. The curvature of each asperity tip is to be determined by fitting a curve (in two-dimensions) or surface (in three-dimensions) through the points surrounding the peak. For a differentiable surface, this procedure converges to identify a well-defined set of asperities as the sampling interval is reduced. When this procedure is applied to experimentally measured surfaces, however, the number of asperities and their curvature appear to increase without limit as the sampling interval is reduced to extremely small values.

This is, of course, a consequence of the fractal nature of surface roughness. There has been some debate as to the most accurate description of realistic surfaces, (Whitehouse & Archard 1970; Sayles & Thomas 1978; Majumdar & Bhushan 1990; Majumdar & Tien 1990, 1991) but all models agree that surfaces have an approximately self-affine fractal character over a wide range of length-scales, and usually continue to exhibit fractal behaviour down to the limits of the resolution of standard surface profilometres. Indeed, Majumdar and Bhushan (1990, 1991) show data that suggests that roughness spectra for hard disk drives remain fractal down to wavelengths of a few hundred nanometres. For a perfectly fractal surface, Greenwood and Williamson's description of the surface topography breaks down completely, since such a surface does not contain well-defined summits.

Many recent studies of rough surface contact have, therefore, abandoned the Greenwood–Williamson description of surface geometry in favour of one that captures in some way the fractal character of surface roughness. Various approaches have been proposed, including procedures for adapting Greenwood and Williamson's calculations to fractal surfaces by introducing generalized definitions of asperities (Majumdar & Bhushan 1990, 1991), direct numerical simulations of experimentally measured profiles (e.g. Sayles 1996; Borri-Brunetto *et al.* 1999; Polonsky & Keer 2000*a,b*; Hyun *et al.* 2004; Pei *et al.* in press), or hybrid schemes which combine numerical simulations with Greenwood and Williamson's approach (see Mihailidis *et al.* 2001 for a review). Attempts to adapt Greenwood–Williamson to a fractal surface appear to suffer from the same difficulties as the original Greenwood–Williamson analysis: there is no unambiguous way to characterize asperities on a fractal surface. Numerical simulations provide considerable insight into the deformation fields associated with rough surface contact, but there is a limit to the spatial resolution of any numerical scheme, which artificially truncates the roughness spectrum. This again introduces problems very similar to those associated with the Greenwood–Williamson model: for a fractal surface, predictions of the contact size, number of contacts, and true contact area will not converge as the resolution of the numerical scheme is refined. Great care must be taken to ensure that this truncation does not lead to spurious predictions (Borri-Brunetto *et al.* 1999).

A significant step towards a satisfactory model of the elastic contact between rough fractal surfaces was made recently by Ciavarella *et al.* (2000) and by recent work of Persson (2001, 2005). The approximate approach of Ciavarella *et al.* (2000) will be followed closely in our work and so will be described in more detail in §§2 and 3 of this paper. Briefly, they chose to approximate surface roughness using the two-dimensional Weierstrass function (given in equation 2.1 below). The Weierstrass function consists of a series of discrete harmonic functions, with progressively decreasing wavelength and amplitude. Ciavarella *et al.* devised a hierarchical scheme to predict the apparent contact pressure and area on each successive scale of

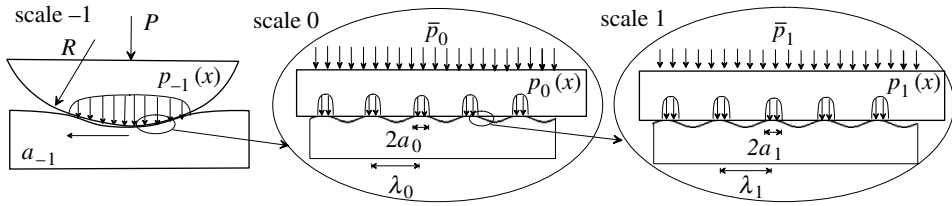


Figure 1. Idealized two-dimensional model of contact between a smooth cylinder and a surface with Weierstrass profile with widely separated scales. Successive scales interact through the condition that the nominal pressure acting on scale n is equal to the true pressure acting on scale $n-1$.

roughness. For the limit of a perfectly fractal surface, prediction that the true contact area consists of an infinite number of contact spots, with zero size, infinite pressure, and vanishing total area. Interestingly, they also showed that the same conclusion can be derived from Archard's (1957) early models of rough surface contact (Ciavarella *et al.* 2000; Ciavarella & Demelio 2001). This is a purely formal mathematical result, of course: in practice phenomena such as plastic flow or adhesion might alter these predictions, and in any event no real surface is perfectly fractal. There are some applications, such as modelling friction, where the true contact size and area are of great interest, and predicting these quantities remains an unresolved problem. For modelling the compliance of the surface, or its electrical contact resistance, and possibly physical phenomena such as plastic flow, wear, and the initiation and growth of contact fatigue cracks, the behaviour of very short-wavelength roughness is less important, since damage events occur at sub-micron length-scales rather than nanometre length-scales. The main advantage of the approach devised by Ciavarella *et al.* is that the behaviour of long-wavelength roughness scales can be predicted in terms of measurable surface parameters without needing to identify real asperities on the surface.

The work of Ciavarella *et al.* (2000) was restricted to elastic solids. In this paper, our goal is to extend their calculations to account for plastic deformation. The problem to be solved is illustrated in figure 1. We consider an isotropic, elastic-perfectly plastic solid with Young's modulus E , Poisson's ratio ν and yield stress σ_Y . The solid has a rough surface, which is characterized by the two-dimensional Weierstrass profile given in equation (2.1). The solid is indented by a smooth cylinder with radius R , which is subjected to a force P per unit length out of the plane. The goal of our analysis is to estimate quantities such as the number of contact spots, their size (or more precisely, their size distribution); the true contact pressure acting on the contact spots, and the total true area of contact. It is of particular interest to calculate critical conditions that will initiate plastic flow in the asperities, and to determine the extent of this plastic flow as a function of the applied loading, the material properties of the deforming solid, and the surface roughness.

The remainder of this paper is organized as follows. In §2, we give a more careful statement of the problem to be solved, the assumptions underlying our calculations, and the results that will be calculated. Section 3 gives a brief review of the indentation response of an elastic-perfectly plastic solid with sinusoidal roughness, as the first step towards analysing the Weierstrass profile. In §4, we review the procedure proposed by Ciavarella *et al.* to analyse contact of the Weierstrass profile, and devise a simplified, approximate procedure that can be

used to derive analytical results for both elastic and plastic contacts. Section 5 presents detailed results and §6 contains conclusions.

2. Rough surface model

Majumdar & Bhushan (1991) suggest that the two-dimensional Weierstrass function

$$h(x) = G_0 \sum_{m=0}^{\infty} \gamma^{(D-2)m} \cos(2\pi x \gamma^m / A_0), \quad (2.1)$$

with $\gamma \approx 1.5$ and $D \approx 1.4$ provides a good qualitative description of many practical surface profiles. The power spectrum for the ideal Weierstrass profile is a series of Dirac delta functions, but Berry & Lewis (1980) argue that the Weierstrass distribution can be used to approximate a continuous power spectrum by ensuring that the integral of the continuous power spectral density matches that of the Weierstrass function. The resulting continuous approximation to the power spectral density is given by

$$S(\omega) = \frac{G_0^2 A_0}{2 \log(\gamma)} (A_0 \omega)^{2D-5}, \quad (2.2)$$

where the circular frequency ω has dimension of L^{-1} . Although, the Weierstrass function captures the multi-scale nature of surface roughness, it is not a completely realistic description of real surfaces. The actual power spectrum for measured surfaces is continuous, not a series of Dirac delta functions. In addition, roughness wavelengths in the range $50 \mu\text{m}$ – 1mm are strongly sensitive to processing and in general are not well characterized by even the continuous approximation to the Weierstrass function. Similarly, the Weierstrass profile becomes unphysical at very short wavelengths: as $m \rightarrow \infty$, the surface becomes needle-like, with roughness amplitude much greater than its wavelength. Any departure from ideal behaviour at long wavelength (low m) will lead to quantitative, but not qualitative changes in our predictions and could readily be incorporated in the computations. The departure from ideal behaviour at short wavelengths (large m) has a critical influence on the true size of the contact spots, and for ideally elastic surfaces, on the total true area of contact and contact pressure. It will not, however, have a significant influence on the behaviour of longer-wavelength scales of roughness.

Analysing contact of a Weierstrass surface with a realistic value of γ is difficult, because roughness scales with successive values of m do not clearly decouple. Ciavarella *et al.* (2000) show that calculations are greatly simplified by approximating a fractal surface with a larger value of $\gamma > 5$. The surface then consists of a series of harmonic profiles, with widely separated wavelength, each of which can be analysed independently. Although, the resulting surface does not resemble any actual surface, it does capture the self-affine fractal character of real surfaces, and so provides an invaluable qualitative picture of the behaviour of two contacting fractal surfaces.

We now consider an elastic–perfectly plastic solid, with Young’s modulus E , Poisson’s ratio ν and yield stress σ_Y , which has a surface roughness defined by equation (2.1). The solid is indented by a smooth cylinder with radius R as illustrated in figure 1. The cylinder is loaded by a total force (per unit out of

plane length) P , to induce a contact pressure distribution $p(x)$. Of course, the two surfaces actually meet only at the tips of asperities. We shall find that for an elastic-perfectly plastic solid with an ideal Weierstrass profile, the true contact consists of an infinite number of contact spots of zero size, which are all subjected to the same, uniform contact pressure. This is a purely formal result, however, and does not characterize the loading or deformation of the solid in a useful manner. We obtain a better picture by regarding the loading and deformation to consist of a series of separate scales, $n = -1, 0, 1, \dots, \infty$, as illustrated in figure 1. The lowest scale, $n = -1$ corresponds to the nominal contact. Successive terms $n = 0, 1, 2, \dots, \infty$ characterize the behaviour of the solid at length-scales comparable to successive wavelengths in the Weierstrass profile. The fields associated with each term interact weakly through the requirement that the pressure at scale n must be statically equivalent to the pressure at scale $n+1$.

The nominal contact area and pressure (scale $n = -1$) can be computed using results for smooth contacts (see e.g. Johnson 1985). It is helpful to review these results briefly. The severity of the nominal load is conveniently characterized by a dimensionless load parameter $\sqrt{PE^*/\sigma_Y^2 R}$. For $\sqrt{PE^*/\sigma_Y^2 R} < 3.2$, the contact is elastic, and the contact width and nominal contact pressure distribution follow as

$$a_{-1} = \sqrt{\frac{4PR}{\pi E^*}} \quad p_{-1}(x) = \frac{4}{\pi} p_{-1}^{\text{mean}} \sqrt{1 - x^2/a_{-1}^2}, \quad (2.3)$$

where $p_{-1}^{\text{mean}} = P/2a_{-1} \leq 1.4\sigma_Y$ is the average nominal contact pressure. For $\sqrt{PE^*/\sigma_Y^2 R} > 40$, the contact is fully plastic, and the contact width and stress distribution can be approximated using the rigid plastic slip-line field solution, which gives

$$a_{-1} = P/(2H_0\sigma_Y) \quad p_{-1}(x) = H_0\sigma_Y, \quad (2.4)$$

where $H_0 \approx 2.9$ is the ratio of the plane strain hardness of the solid to its tensile yield stress. In the intermediate regime $3.2 < \sqrt{PE^*/\sigma_Y^2 R} < 40$, the contact width and pressure distribution would need to be computed using numerical simulations. This will not be attempted here: we shall focus instead on the limiting cases of perfectly elastic and fully plastic bulk contacts.

It is important to note that the nominal contact geometry introduces a natural and well-defined long-wavelength cut-off for the roughness spectrum. Roughness wavelengths greater than a_{-1} clearly do not influence the contact pressure. The roughness component with wavelength closest to a_{-1} should be regarded as part of the nominal contact geometry (its effect is equivalent to modifying the shape of the indenter). The first component of the profile to influence the contact pressure distribution has index $m = 1 + \log(A_0/a_{-1})/\log \gamma$, and has wavelength and amplitude:

$$\lambda_0 = a_{-1}/\gamma \quad g_0 = G_0(\gamma A_0/a_{-1})^{(D-2)}. \quad (2.5)$$

Henceforth, we will characterize the surface roughness within the area of contact as

$$h(x) = g_0 \sum_{n=0}^{\infty} \gamma^{(D-2)n} \cos(2\pi x \gamma^n / \lambda_0), \quad (2.6)$$

where g_0 and λ_0 represent the amplitude and wavelength of the first roughness component that has wavelength less than the nominal contact width. It will be

important to keep in mind that while γ , D , G_0 and A_0 are properties of the surface geometry only, g_0 and λ_0 depend also on the nominal contact geometry as shown in equation (2.5). We shall return to this issue when presenting our results in §5.

If we now examine a small region of the nominal contact, with length $\lambda_1 \ll dx \ll a_{-1}$ and centred at some point x within the contact area, we observe a sinusoidal surface with wavelength λ_0 in contact with a nominally flat indenter. The indenter is subjected to a uniform nominal contact pressure $\bar{p}_0 = p_{-1}(x)$, which induces a contact width $2a_0$ and pressure $p_0(x)$ on the zeroth roughness scale. We can then examine more closely a region with size $\lambda_2 \ll dx \ll \lambda_0$ within one of these contacts. Here, we observe a sinusoidal surface with wavelength λ_1 , subjected to nominal contact pressure $\bar{p}_1 = p_0(x)$, inducing a contact size a_1 , and contact pressure $p_1(x)$. This process can evidently be continued indefinitely.

The goal of our analysis is to determine, for each roughness scale n , the total number of contact spots N_n , the apparent size of the contact spots a_n , and the apparent contact pressure distribution $p_n(x)$. Calculations are complicated by the fact that the quantities of interest vary with position on the surface. It is possible (but exceedingly cumbersome) to calculate this spatial variation. In practice, it is preferable to quantify the variation by computing distribution functions, which, for example, specify the fraction of asperity contacts with size between a_n and $a_n + da_n$, or the fraction of the nominal contact area that is subjected to pressure between p_n and $p_n + dp_n$, for each scale $n = 0, \dots, \infty$.

The distribution function for the pressure plays a particularly important role in subsequent calculations. For each scale of roughness, we define a dimensionless function $q_n(p/\sigma_Y)$ such that $q_n(p/\sigma_Y)(dp/\sigma_Y)$ is the fraction of the nominal contact (which has total width $2a_{-1}$) that is subjected to a dimensionless pressure between p_n/σ_Y and $(p_n + dp_n)/\sigma_Y$. As an example, the distribution function $q_{-1}(p/\sigma_Y)$ characterizing the nominal contact can easily be computed. The cumulative distribution $Q(p)$ (the proportion of the contact over which the pressure exceeds p) can be computed from the contact pressure distribution as $Q_{-1}(p) = x(p)/a_{-1}$. The distribution function $q_{-1}(p)$ then follows as $q_{-1}(p) = -\partial Q_{-1}/\partial p$. For example, if the nominal contact is elastic, we have

$$Q_{-1}(p) = x(p)/a_{-1} = \sqrt{1 - (\pi p/4p_{-1}^{\text{mean}})^2}. \quad (2.7)$$

The distribution function follows as

$$q_{-1}(\xi) = -\sigma_Y \frac{dQ}{dp} = \frac{\xi(\pi/4\xi_{-1}^{\text{mean}})^2}{\sqrt{1 - (\pi\xi/4\xi_{-1}^{\text{mean}})^2}}, \quad (2.8)$$

where $\xi = p/\sigma_Y$, $\xi_{-1}^{\text{mean}} = p_{-1}^{\text{mean}}/\sigma_Y$. Similarly, for fully plastic contact, the distribution function is evidently

$$q_{-1}(\xi) = \delta(\xi - H_0), \quad (2.9)$$

where δ denotes the Dirac delta distribution and $H_0 \approx 2.9$ is the ratio of the hardness to the tensile yield strength of the solid, as before. Similar distributions will characterize the pressure distribution for each higher roughness scale. Although, it is not possible to write down a simple expression relating the

contact pressure $p_n(x)$ to $q_n(p)$ for $n \geq 0$, Ciavarella *et al.* (2000) show that this is not necessary. Their procedure for computing q_n will be reviewed in detail in §4.

The distribution function $q_n(p)$ provides a convenient description of the contact pressure acting on the n th roughness scale, but is difficult to interpret physically. To provide a more direct measure of the deformation of each roughness scale, we introduce the contact size distribution function $S_n(a)$ defined such that

$$dn = N_n S_n(a) da, \quad (2.10)$$

with N_n the total number of contact spots at the n th scale, gives the number of contacts with size between a and $a+da$. Similarly, as a more direct measure of the loading applied to the contacts in the n th scale, we define the mean contact pressure for an individual contact spot as

$$p^{\text{mean}} = \frac{1}{2a} \int_{-a}^a p(x) dx. \quad (2.11)$$

We then introduce the contact pressure distribution $P_n(p^{\text{mean}})$ such that

$$dn = N_n P_n(p^{\text{mean}}) dp^{\text{mean}} \quad (2.12)$$

gives the number of contacts in the n th scale that are subjected to mean pressure between p^{mean} and $p^{\text{mean}} + dp^{\text{mean}}$. Naturally, since there is a direct relationship between the contact size and mean contact pressure for individual contact spots, there will be a similar direct relationship between $S_n(a)$ and $P_n(p^{\text{mean}})$.

Although, we shall formally compute expressions for $S_n(a)$ and $P_n(p^{\text{mean}})$ for our model surface, we readily concede that in view of the rather crude description of surface topography that forms the basis for our calculations, the exact distribution functions may have little practical significance. In many cases, the *average* values of contact size and pressure for each scale of contact are sufficient. These are formally related to $S_n(a)$ and $P_n(p^{\text{mean}})$ through

$$\langle a_n \rangle = \int_0^\infty S_n(a) a da \quad \langle p_n^{\text{mean}} \rangle = \int_0^\infty P_n(p) p dp. \quad (2.13)$$

In practice, however, it is simpler to compute $\langle a_n \rangle$ and $\langle p_n^{\text{mean}} \rangle$ directly, rather than via the distribution functions. The total contact area and nominal contact pressure:

$$A_n = N_n 2 \langle a_n \rangle, \quad p_n^{\text{nom}} = P/A_n. \quad (2.14)$$

We shall calculate these quantities using the elegant procedure developed by Ciavarella *et al.* (2000) to model contact between elastic solids with a Weierstrass profile. A key assumption in their work is that the behaviour of a particular roughness scale n at a point on the surface can be calculated from the solution to a sinusoidal surface indented by a flat, frictionless punch. Successive roughness scales interact only through the requirement that the nominal pressure \bar{p}_n that acts on scale n is equal to the true pressure $p_{n-1}(x)$ for the preceding scale $n-1$. This assumes that no interaction occurs between the deformation fields associated with successive roughness scales.

For an elastic solid, this assumption can easily be justified using Saint Venant's principle. For elastic-plastic materials, no rigorous proof exists. We speculate, however, that provided the bulk contact (at scale $n-1$) consists of

a single, isolated contact area, with a size much smaller than any characteristic structural dimension, the same decoupling can be applied for elastic–plastic solids. As an extreme example, consider the limiting case of a rigid–perfectly plastic solid. The slip-line field solutions for contact between a semi-infinite solid and a single indenter invariably show a rigid region under the contact area (Hill 1950; also reproduced in Johnson 1985). Consequently, the deformation associated with the first roughness scale in a Weierstrass surface (scale 0) may be embedded within the rigid region associated with the nominal contact. The behaviour of the first scale of roughness is, therefore, equivalent to that of a sinusoidal surface, with rigid boundary conditions at infinity. Our numerical solution to this problem shows that a rigid region forms under each contact area. Consequently, the second roughness scale can similarly be embedded within this region, and this process can be continued for all higher roughness scales. Since, the deformation field constructed in this manner forms a kinematically admissible collapse mechanism, our solution should strictly be regarded as an upper bound to the collapse load, and a lower bound to the contact size and contact area. We present some further support for our calculations in electronic supplementary material, where the analytical calculations are compared with the predictions of full-field finite element computations for up to three roughness scales. It is important to recognize, however, that if the deformation field for the bulk contact (scale -1) is not contained, a rigid region may not form below the bulk contact in the fully plastic limit. In this case, the first (longest-wavelength) scale of roughness is not contained, and so cannot be idealized as a sinusoidal surface with rigid boundary conditions at infinity. Similarly, there is no reason to suppose that rigid regions would develop to contain shorter-wavelength scales. Such unconstrained flow is particularly likely to occur in metal forming operations, and existing slip-line field solutions suggest that asperities are easily flattened in the presence of unconstrained far-field deformation (Sutcliffe 1988).

3. Indentation response of a sinusoidal surface

Clearly, establishing the response of a simple sinusoidal surface to indentation loading is a key step towards solving the contact problem outlined in §1. The indentation response of an elastic–perfectly plastic solid with a harmonic roughness was analysed in detail by Gao *et al.* (in press). Only a few of their results are needed here: these are reviewed briefly below.

We focus attention on a single roughness scale with wavelength and amplitude

$$\lambda_n = \lambda_0/\gamma^n, \quad g_n = g_0\gamma^{(D-2)n}, \quad (3.1)$$

as illustrated in figure 1. The substrate is an elastic–perfectly plastic solid with Young’s modulus E , Poisson’s ratio ν and yield stress σ_Y . Gao *et al.* (in press) show that the strength of the surface is conveniently characterized by a material and geometric parameter given by

$$\psi_n = \frac{g_n E^*}{\lambda_n \sigma_Y}, \quad (3.2)$$

with $E^* = E/(1-\nu^2)$. In the discussion to follow, the value of ψ_n for the zeroth roughness scale ($n=0$) plays a central role in characterizing the resistance of the

rough surface to plastic flow. In terms of this parameter, we have

$$\psi_0 = \frac{g_0 E^*}{\lambda_0 \sigma_Y}, \quad \psi_n = \psi_0 \gamma^{(D-1)n}. \quad (3.3)$$

Note that the resistance to plastic flow varies in inverse proportion to ψ : a rough surface with a low yield stress has a high value of ψ , while a smooth, hard surface would have a low ψ . In addition, note that ψ_n increases with n , so we anticipate that roughness scales with long-wavelength will be elastic, but all scales beyond a critical value of n will deform plastically.

The solid is indented by a flat, rigid surface, under the action of a nominal pressure. The loading induces a contact width $2a_n$, pressure distribution $p_n(x)$ and mean pressure p_n^{mean} given by the functional relationships

$$\left. \begin{aligned} 2a_n/\lambda_n &= A\left(\frac{\bar{p}_n}{\sigma_Y}, \psi_n, \frac{g_n}{\lambda_n}\right), \\ \frac{p_n(x)}{\sigma_Y} &= \Phi\left(\frac{x}{\lambda_n}, \frac{\bar{p}_n}{\sigma_Y}, \psi_n, \frac{g_n}{\lambda_n}\right), \\ \frac{p_n^{\text{mean}}}{\sigma_Y} &= \Phi^{\text{mean}}\left(\frac{\bar{p}_n}{\sigma_Y}, \psi_n, \frac{g_n}{\lambda_n}\right). \end{aligned} \right\} \quad (3.4)$$

Here, we have shown that formally that A , Φ and Φ^{mean} depend on both ψ_n and g_n/λ_n . In fact, if the surface remains elastic, exact calculations show that the solution depends only on ψ_n and is independent of g_n/λ_n . If the surface deforms plastically, numerical simulations show that the functions depend only weakly on g_n/λ_n . It would not be difficult to include this variation in our computations, but for simplicity we have chosen to neglect it.

Finally, we must also compute the pressure distribution function $q_n(p/\sigma_Y)$, which is defined so that $q_n(p/\sigma_Y)(dp/\sigma_Y)$ is the fraction of the sinusoidal surface that is subjected to normalized pressure between p_n/σ_Y and $(p_n + dp_n)/\sigma_Y$. This distribution function can be determined from the contact pressure, following the procedure given for the nominal contact in equations (2.7) and (2.8). The result will be specified by the functional relationship

$$q_n\left(\frac{p}{\sigma_Y}\right) = q\left(\frac{p}{\sigma_Y}, \frac{\bar{p}_n}{\sigma_Y}, \psi_n\right). \quad (3.5)$$

We have calculated approximations to the functions A , Φ , Φ^{mean} and q by fitting curves to the numerical simulations reported in Gao *et al.* (in press), which are reproduced in figure 2. The resulting expressions are somewhat lengthy, and are given in detail in the electronic supplementary material A. Briefly, we observe four general regimes of behaviour, depending on the normalized contact pressure $\xi = \bar{p}/\sigma_Y$.

- (i) For $0 < \xi < \pi\psi \sin^2(\pi/10\psi)$, the contact remains elastic. Expressions for the contact size and pressure distributions in the elastic regime are given in Westergaard (1939) and Ciavarella *et al.* (2000), and are reproduced in electronic supplementary material.

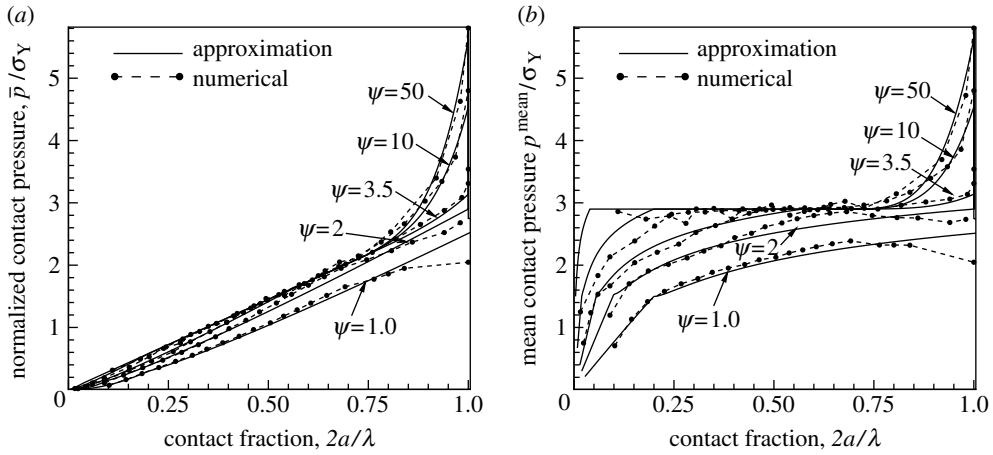


Figure 2. The relationship between: (a) nominal pressure and (b) mean pressure and contact fraction for a sinusoidal surface in contact with a rigid, flat surface. The symbols show the results of finite element computations, while the solid lines show the approximate analytical expressions used in our rough surface model.

- (ii) For $\pi\psi \sin^2(\pi/10\psi) < \xi < 2H_0/\psi$, where $H_0 \approx 2.9$ is the ratio of the (plane strain) hardness to the yield stress of the solid, the contacts are in the elastic–plastic regime. In this regime, we estimate the contact size and pressure using a smooth interpolation between elastic and fully plastic behaviour. Full expressions are given in electronic supplementary material.
- (iii) For $2H_0/\psi < \xi < 2H_0/3$, the contacts behave like isolated (plane strain) hardness indentations, with a uniform contact pressure with magnitude $p = H_0\sigma_Y$.
- (iv) For $\xi > 2H_0/3$, the contact fraction satisfies $A > 2/3$, so that the plastic zones under neighbouring asperities begin to interact. In this regime, the contact pressure under each contact remains approximately uniform, but the magnitude of the pressure increases monotonically as the contact fraction approaches one. The limiting value of the contact pressure as $A \rightarrow 1$ plays an important role in subsequent discussions. The limiting value of is a function of ψ_n , and as $\psi_n \rightarrow \infty$ approaches $p = 5.8\sigma_Y$: approximately twice the material hardness.

Our approximate curves are compared with the results of finite element simulations in figure 2, which shows a parametric plot of the nominal and mean contact pressures as a function of the contact size as the load increases, for various values of ψ .

4. Contact mechanics for a general Weierstrass profile

We proceed to analyse contact of the Weierstrass profile. Our analysis follows exactly the work of Ciavarella *et al.* (2000) and so will be reviewed only briefly. We will first present a general procedure for computing distribution functions for

contact size and pressure for each scale. Subsequently, we suggest a simple approximation that can be used to obtain closed-form estimates for the number of contact spots, their size, and mean pressure for each scale of loading.

(a) *Results for general surface contact*

Although, it is difficult to compute the spatial contact pressure distribution for the Weierstrass surface, it is straightforward to determine the probability distribution function $q_n(p/\sigma_Y)$ defined in §3. The calculation begins by computing the distribution function $q_{-1}(p/\sigma_Y)(dp/\sigma_Y)$, which specifies the proportion of the nominal contact area that is subjected to a pressure between p/σ_Y and $(p+dp)/\sigma_Y$. The result is given in equation (2.8).

The distribution $q_0(p)$ for the zeroth roughness scale may, therefore, be computed as follows. Recall that we introduced the dimensionless function q , such that $q(p/\sigma_Y, \xi, \psi)(dp/\sigma_Y)$ specifies the proportion of the sinusoidal surface that is subjected to pressure between p/σ_Y and $(p+dp)/\sigma_Y$ when loaded by a normalized nominal pressure ξ . The zeroth roughness scale is subjected to a distribution of nominal pressure, as quantified by $q_{-1}(p/\sigma_Y)$. Summing over all possible values of the contact pressure for the nominal contact, we find that

$$q_0(p/\sigma_Y) = \int_0^\infty q(p/\sigma_Y, \xi, \psi_0) q_{-1}(\xi) d\xi. \quad (4.1)$$

The upper limit on the integral here is purely formal, of course: the nominal contact pressure does not exceed $p_{-1} = 2P/\pi a_{-1}$, so $q_{-1}(\xi) = 0$ for $\xi > 2P/\pi a_{-1}\sigma_Y$. This result can be applied recursively, to see that

$$q_n(p/\sigma_Y) = \int_0^\infty q(p/\sigma_Y, \xi, \psi_n) q_{n-1}(\xi) d\xi. \quad (4.2)$$

Once the hierarchy of distribution functions $q_n(p/\sigma_Y)$ has been computed, various quantities of interest can be deduced. For example, to calculate the total (apparent) area of contact for the n th scale, note that the area of contact for one wavelength of the n th scale subjected to pressure \bar{p} is $2a_n = \lambda_n A(\bar{p}/\sigma_Y, \psi_n)$. In addition, there are $2a_{-1}/\lambda_n$ such wavelengths within the nominal area of contact. The total contact area, therefore, follows as

$$A_n = 2a_{-1} \int_0^\infty A(\xi, \psi_n) q_{n-1}(\xi) d\xi. \quad (4.3)$$

The total number of contact spots at the n th scale within the nominal contact is

$$N_n = \frac{A_{n-1}}{\lambda_n}, \quad (4.4)$$

whereupon the average size of the contacts at the n th scale follows as

$$2\langle a_n \rangle = \frac{A_n}{N_n} = \lambda_n \frac{A_n}{A_{n-1}}. \quad (4.5)$$

The average contact pressure can be found by a similar approach, with the result

$$\langle p_n^{\text{mean}} \rangle = \frac{2a_{-1}}{A_{n-1}} \int_0^\infty \Phi^{\text{mean}}(\xi, \psi_n) q_{n-1}(\xi) d\xi. \quad (4.6)$$

Finally, the nominal contact pressure for the n th scale follows as

$$p_n^{\text{mean}} = \frac{P}{A_n} = p_{-1}^{\text{mean}} \left\{ \int_0^\infty A(\xi, \psi_n) q_{n-1}(\xi) d\xi \right\}^{-1}. \quad (4.7)$$

We can also find expressions for the contact size and mean pressure distribution functions $S_n(a)$ and $P_n(p^{\text{mean}})$ defined in equations (2.10) and (2.12). To compute the distribution function for the contact size, denote the pressure required to induce a contact size a is given by the functional relationship $p/\sigma_Y = A^{-1}(2a/\lambda, \psi)$, where A^{-1} denotes the inverse of the function A defined in equation (3.4). Given the distribution function for the contact pressure at $(n-1)$ th scale, we can easily compute the probability distribution function of the contact fraction for scale n as follows. For a single contact, $2a_n/\lambda_n = A(p_{n-1}/\sigma_Y, \psi_n)$, so that let $p_{n-1}/\sigma_Y = A^{-1}(2a_n/\lambda_n, \psi_n)$. The probability of finding a contact fraction in between $2a_n/\lambda_n$ and $2(a_n + da_n)/\lambda_n$ is

$$S_n(2a_n/\lambda_n) = \frac{dA^{-1}}{d(2a_n/\lambda_n)} q_{n-1}(A^{-1}(2a_n/\lambda_n, \psi_n)). \quad (4.8)$$

(b) *Simplified analysis—the uniform contact pressure assumption*

The results listed in the preceding section can be simplified dramatically if the contact pressure distribution for each scale in the contact (including the nominal contact area) is *uniform*. This is an excellent approximation when each scale of contact is fully plastic. At lighter loads, the nominal contact, and some long-wavelength roughness scales, will remain elastic, and for these scales the distribution of contact pressure is not uniform. Surprisingly, however, we have found that averaged quantities such as the total area of contact, the average contact size, and the average contact pressure can be computed remarkably accurately even for elastic surfaces by replacing this non-uniform contact pressure distribution with a statically equivalent, uniform pressure.

Proceeding along these lines, therefore, we suppose that the entire nominal contact area is subjected to the same contact pressure. For this approximation, the distribution function $q_{-1}(p/\sigma_Y)$, which specifies the fraction of the nominal contact area (scale -1) subjected to pressure between p/σ_Y and $(p+dp)/\sigma_Y$, must, therefore, be given by

$$q_{-1}(\xi) = \delta(\xi - p_{-1}^{\text{mean}}/\sigma_Y), \quad (4.9)$$

where $\delta(x)$ denotes the Dirac delta distribution and $p_{-1}^{\text{mean}} = P/2a_{-1}$. The distribution function for the zeroth scale of roughness follows from equation (4.2) as

$$q_0(p/\sigma_Y) = \int_0^\infty q(p/\sigma_Y, \xi, \psi_0) \delta(\xi - p_{-1}^{\text{mean}}/\sigma_Y) d\xi = q(p/\sigma_Y, p_{-1}^{\text{mean}}/\sigma_Y, \psi_0). \quad (4.10)$$

Recall now that the function $q(p/\sigma_Y, p_{-1}^{\text{mean}}/\sigma_Y, \psi_0)$ specifies the fraction of a sinusoidal surface that is subjected to pressure between p/σ_Y and $(p+dp)/\sigma_Y$, when indented by a nominal contact pressure p_{-1}^{mean} . If the pressure is approximated by a statically equivalent uniform distribution, with magnitude

p^{mean} , it follows that

$$q(p/\sigma_Y, p_{-1}^{\text{mean}}/\sigma_Y, \psi) = A(p_{-1}^{\text{mean}}/\sigma_Y, \psi) \delta(p/\sigma_Y - \Phi^{\text{mean}}(p_{-1}^{\text{mean}}/\sigma_Y, \psi)), \quad (4.11)$$

where the functions A and Φ^{mean} are defined in equations (3.4), and given specific forms in equations (A2)–(A18) of the electronic supplementary material. Repeating this procedure for subsequent roughness scales gives the following result for the distribution functions $q_n(p)$:

$$q_n(p/\sigma_Y) = \left\{ \prod_{j=0}^n A(p_{j-1}^{\text{mean}}/\sigma_Y, \psi_j) \right\} \delta(p/\sigma_Y - \Phi^{\text{mean}}(p_{n-1}^{\text{mean}}/\sigma_Y, \psi_n)). \quad (4.12)$$

Here, the mean contact pressure at each scale can be determined from

$$p_n^{\text{mean}}/\sigma_Y = \Phi^{\text{mean}}(p_{n-1}^{\text{mean}}/\sigma_Y, \psi_n), \quad (4.13)$$

where the sequence is started with p_{-1}^{mean} given in equation (4.9).

Given p_n^{mean} for each scale, and the distribution functions $q_n(p/\sigma_Y)$, we can determine the remaining quantities of interest using equations (4.3)–(4.7). The total area of contact at the n th scale follows as

$$A_n = 2a_{-1} \int_0^\infty A(\xi, \psi_n) q_{n-1}(\xi) d\xi = 2a_{-1} \prod_{j=0}^n A(p_{j-1}^{\text{mean}}/\sigma_Y, \psi_j), \quad (4.14)$$

where the function A is defined in equation (3.4). The number of contacts follows as

$$N_n = 2(a_{-1}/\lambda_n) \prod_{j=0}^{n-1} A(p_{j-1}^{\text{mean}}/\sigma_Y, \psi_j), \quad (4.15)$$

and the average contact size is given by

$$2\langle a_n \rangle = \lambda_n A(p_{n-1}^{\text{mean}}/\sigma_Y, \psi_n). \quad (4.16)$$

Finally, equation (4.7) provides an additional way to compute p_n^{mean} as

$$p_n^{\text{mean}} = p_{-1}^{\text{mean}} \left\{ \prod_{j=0}^n A(p_{j-1}^{\text{mean}}/\sigma_Y, \psi_j) \right\}^{-1}, \quad (4.17)$$

from which we also recover the expected relationship between the mean pressure at successive scales:

$$p_n^{\text{mean}} = p_{n-1}^{\text{mean}} A(p_{n-1}^{\text{mean}}/\sigma_Y, \psi_n). \quad (4.18)$$

(c) *Approximate analytical expressions for average contact size, pressure, number of contacts, and total contact area for elastic contact*

If the contacts remain elastic, the procedure outlined in the preceding section can be used to derive simple analytical expressions for the average contact size, average contact pressure.

Under these conditions, we may approximate

$$\left. \begin{aligned} A(\xi, \psi) &= \frac{2}{\pi} \sin^{-1} \sqrt{\frac{\xi}{\pi\psi}} \approx \frac{2}{\pi} \sqrt{\frac{\xi}{\pi\psi}}, \\ \Phi^{\text{mean}}(\xi, \psi) &= \frac{\pi\xi}{2 \sin^{-1}(\sqrt{\xi/\pi\psi})} \approx \frac{\pi}{2} \sqrt{\pi\xi\psi}, \end{aligned} \right\} \quad (4.19)$$

whereupon equation (4.13) reduces to

$$\frac{p_n^{\text{mean}}}{\sigma_Y} = \frac{\pi}{2} \sqrt{\frac{p_{n-1}^{\text{mean}}}{\sigma_Y} \pi\psi_n} = \frac{\pi}{2} \sqrt{\frac{p_{n-1}^{\text{mean}}}{\sigma_Y} \pi\psi_0 \gamma^{(D-1)n}}. \quad (4.20)$$

This recursion relation is satisfied by

$$\frac{p_n^{\text{mean}}}{\sigma_Y} = \left(\frac{\pi}{2}\right)^{2-2^{-n}} (\gamma^{D-1})^{2^{-n}-1} \pi\psi_0 \gamma^{(D-1)n} \left(\frac{p_{-1}^{\text{mean}}}{\sigma_Y \pi\psi_0}\right)^{2^{-(n+1)}}. \quad (4.21)$$

The average contact size may be estimated using equation (4.16) as

$$2\langle a_n \rangle = \lambda_n \left(\frac{\pi}{2}\right)^{-2^{-n}} (\gamma^{D-1})^{2^{-n}-1} \left(\frac{p_{-1}^{\text{mean}}}{\sigma_Y \pi\psi_0}\right)^{2^{-(n+1)}}. \quad (4.22)$$

Similarly, the total area of contact at the n th scale follows as

$$A_n = 2a_{-1} \left(\frac{\pi}{2}\right)^{2^{-n}-2} \gamma^{-(D-1)(n-1+2^{-n})} \left(\frac{p_{-1}^{\text{mean}}}{\sigma_Y \pi\psi_0}\right)^{1-2^{-(n+1)}}. \quad (4.23)$$

The total number of contacts at the n th scale can be estimated as

$$N_n = A_{n-1} \gamma^{(n+1)} / (2a_{-1}). \quad (4.24)$$

For large n , these results may be approximated by their asymptotes

$$\frac{p_n^{\text{mean}}}{\sigma_Y} = \left(\frac{\pi}{2}\right)^2 \pi\psi_0 \gamma^{(D-1)n}, \quad (4.25)$$

$$2\langle a_n \rangle = \lambda_n \gamma^{(1-D)}, \quad (4.26)$$

$$A_n = 2a_{-1} \left(\frac{\pi}{2}\right)^{-2} \gamma^{-(D-1)n} \left(\frac{p_{-1}^{\text{mean}}}{\sigma_Y \pi\psi_0}\right) = \left(\frac{\pi}{2}\right)^{-2} \frac{P}{\sigma_Y \pi\psi_n}, \quad (4.27)$$

$$N_n = \left(\frac{\pi}{2}\right)^{-2} \gamma^{D-(D-2)n} \left(\frac{p_{-1}^{\text{mean}}}{\sigma_Y \pi\psi_0}\right). \quad (4.28)$$

As expected, equation (4.27) shows that the total contact area at the n th scale (for large n) is proportional to the contact load P and independent of the nominal contact size a_{-1} or the indenter geometry (other than through its effect of setting the wavelength of the zeroth roughness scale). This result also provides an opportunity to check the predictions obtained using the constant pressure

approximation (§4b) with the exact solution. Ciavarella *et al.* (2000) find that $A_n = 0.391P/(\sigma_Y \pi \psi_n)$, giving a coefficient that is within 4% of the approximate value in equation (4.27). A more detailed comparison of the approximate and exact calculations in §5 shows similar agreement under all conditions.

Note also that equation (4.25) shows that the mean contact pressure acting on each roughness scale becomes progressively more severe as n increases. Consequently, for an elastic-plastic solid, the short-wavelength roughness scales will always deform plastically. We also see that the formal limit $n \rightarrow \infty$ for a perfectly elastic solid consists of an infinite number of contact spots, with zero size, which are subjected to infinite contact pressure. The total true contact area vanishes.

(d) *Deformation and loading for the most heavily loaded region at each scale (elastic contact)*

To calculate the conditions necessary to initiate yield at each successive scale of roughness, or to identify conditions that will just begin to bring a roughness scale into complete contact with the indenter, it is of interest to determine the contact size and contact pressure for the most heavily loaded asperities at each successive scale of roughness. We assume that all roughness scales, as well as the nominal contact, remain elastic; and assume also that $\gamma^{(D-1)} > 2$. Equation (2.3) shows that the maximum contact pressure exerted by the nominal contact is

$$\frac{p_{-1}^{\max}}{\sigma_Y} = \sqrt{\frac{PE^*}{\pi R \sigma_Y^2}}. \quad (4.29)$$

The most severely loaded asperities on the zeroth roughness scale are, therefore, subjected to a nominal contact pressure $\bar{p}_0 = p_{-1}^{\max}$. Equation (A3) in electronic supplementary material gives the maximum contact pressure for a sinusoidal surface that is subjected to a normalized nominal pressure $\xi = \bar{p}/\sigma_Y$ as

$$p^{\max}(\xi, \psi)/\sigma_Y = 2\sqrt{\pi\xi\psi}, \quad (4.30)$$

whence the maximum contact pressure on the zeroth roughness scale is $p_0^{\max}/\sigma_Y = 2\sqrt{\pi p_{-1}^{\max} \psi_0}/\sigma_Y$. Similarly, the maximum contact pressure on successive scales of roughness follows as $p_n^{\max}/\sigma_Y = 2\sqrt{\pi p_{n-1}^{\max} \psi_n}/\sigma_Y$. The recursion relation is satisfied by

$$\frac{p_n^{\max}}{\sigma_Y} = (2)^{2-2^n} (\gamma^{D-1})^{2^n-1} \pi \psi_0 \gamma^{(D-1)n} \left(\frac{p_{-1}^{\max}}{\sigma_Y \pi \psi_0} \right)^{2^{-(n+1)}}. \quad (4.31)$$

Other than a numerical factor, this result is identical to equation (4.21), which estimates the average contact pressure on each roughness scale. However, equation (4.31) does not assume isolated Hertz-type contacts and so is valid for all values of ψ_0 , provided that no scale reaches full contact.

(e) *A graphical approach to calculating contact size, pressure and area for roughness scales in the fully plastic regime*

If the bulk contact is fully plastic (which requires $\sqrt{PE^*/\sigma_Y^2 R} > 40$), and the surface roughness is such that $\psi_0 > 16$, then the procedure outlined in §4b gives

exact results, since all roughness scales are in the fully plastic regime, so the contact pressure distribution at each scale is uniform. In addition, under these conditions, the method can be illustrated using a simple graphical procedure, originally suggested by Childs (1973, 1977) in a discussion of the persistence of asperities under contact loading.

We begin by simplifying the behaviour of the sinusoidal surface for the fully plastic regime. Consider a single scale of roughness with wavelength λ , which is indented by a rigid platen. The surface is subjected to a nominal contact pressure (force per wavelength) \bar{p} , which induces a contact size a . Each contact spot is subjected to a uniform contact pressure with magnitude p^{mean} . For values of $\psi_0 > 16$, the relationship between the mean normalized contact pressure and contact size for the fully plastic regimes given in equations (A14)–(A18) in electronic supplementary material can be approximated by

$$p^{\text{mean}}/\sigma_Y = \Phi^{\text{mean}} = \begin{cases} H_0, & A < 2/3, \\ H_0 + (H_1(\psi \rightarrow \infty) - H_0)(3A - 2)^3, & A > 2/3, \end{cases} \quad (4.32)$$

where $A = 2a/\lambda$, $H_0 \approx 2.9$ and $H_1(\psi \rightarrow \infty) = 5.8$. In addition, the normalized nominal contact pressure follows as $\xi = \bar{p}/\sigma_Y = A\Phi^{\text{mean}}$. Both these functions are plotted in figure 3.

Next, we apply the iterative scheme outlined in §4*b* for this simplified situation. Since the nominal contact (scale -1) is in the fully plastic regime, the nominal contact pressure is uniform, and has magnitude $p_{-1}/\sigma_Y = 2.9$. This implies that the nominal pressure exerted on asperities in the zeroth roughness scale is $\bar{p}_0/\sigma_Y = 2.9$. Figure 3 can be used to deduce the normalized contact size $2a_0/\lambda_0$ corresponding to this nominal contact pressure; whence the true contact pressure $p_0^{\text{mean}}/\sigma_Y$ can also be found. The same procedure may be used to progress to higher scales of roughness: the condition $p_n^{\text{mean}}/\sigma_Y = \bar{p}_{n+1}/\sigma_Y$ determines the normalized contact size a_{n+1}/λ_{n+1} at each successive roughness scale, whereupon $p_{n+1}^{\text{mean}}/\sigma_Y$ can be deduced. In addition, the apparent total contact area at each scale follows as

$$A_n = 2a_{-1} \prod_{j=0}^n \frac{2a_j}{\lambda_j}. \quad (4.33)$$

In particular, the total contact area in the limit $n \rightarrow \infty$ is

$$\lim_{n \rightarrow \infty} A_n = 2a_{-1} \frac{H_0}{H_1(\psi \rightarrow \infty)} = \frac{P}{5.8\sigma_Y}. \quad (4.34)$$

Thus, we find that the contact area is again proportional to the normal load, but instead of converging to zero as $n \rightarrow \infty$, the total contact area converges to a finite value.

A similar conclusion was reached by Willner (2004) using a more sophisticated series of numerical simulations. His analyses did not account for interactions between neighbouring contact spots, however, and consequently gave a limiting area of $\lim_{n \rightarrow \infty} A_n = P/H_0\sigma_Y$. More recently, Pei *et al.* (in press) have conducted an extensive study of the indentation response of an elastic–plastic solid with a nominally fractal surface roughness using three-dimensional finite element computations. They found that the true area of contact is a function of the

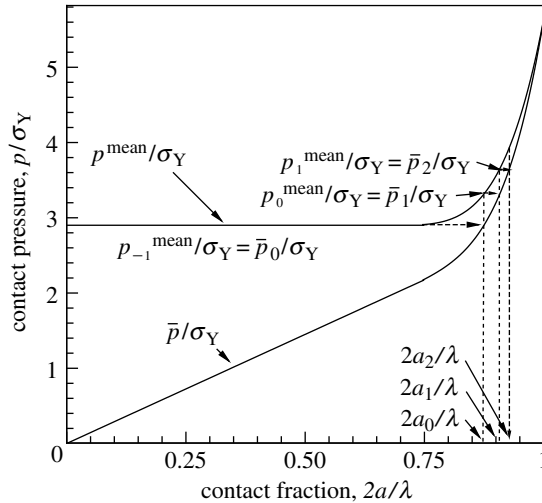


Figure 3. A graphical technique for calculating the contact size and contact pressure for successive scales of a rigid-perfectly plastic surface with Weierstrass profile. The solid lines show the mean pressure p^{mean}/σ_Y and nominal pressure \bar{p}/σ_Y for a rigid plastic sinusoidal surface as a function of contact fraction. Successive scales of roughness are related by the condition that $\bar{p}_n/\sigma_Y = p_{n-1}^{\text{mean}}/\sigma_Y$.

ratio of yield stress to Young’s modulus for the solid: this behaviour is most likely due to an artificial truncation of the roughness spectrum resulting from the resolution of the finite element mesh. For sufficiently low values of σ_Y/E , however, they found that the true contact area was related to the load by $A_{\text{true}} = P/6\sigma_Y$, in excellent agreement with our approximate estimate.

The convergence of contact size and contact area to their limiting values are illustrated in figure 4. Since the normalized contact size $2a_n/\lambda_n \rightarrow 1$ as $n \rightarrow \infty$, while $\lim_{n \rightarrow \infty} A_n/2a_{n-1} = H_0/H_1(\psi \rightarrow \infty)$, we have plotted $2a_n/\lambda_n - 1$ and $A_n/(2a_{n-1}) - H_0/H_1(\psi \rightarrow \infty)$ as functions of n . Evidently, both contact size and total contact area converge to their asymptotic values exponentially: indeed, it is straightforward to show that for sufficiently large n , the contact size and total contact area may be approximated by

$$\left. \begin{aligned} 2a_n/\lambda_n &\approx 1 - 0.1 e^{-0.2n}, \\ A_n/(2a_{n-1}) &\approx H_0(1 - 0.45 e^{-0.2n})/H_1(\psi \rightarrow \infty). \end{aligned} \right\} \quad (4.35)$$

These results can also be used to deduce the asymptotic behaviour of fine-scale roughness on an elastic-plastic solid. We saw in §4c that even if the bulk contact and the first few roughness scales remain below yield, the contact pressure on an elastic surface increases for progressively higher scales of roughness. Consequently, there will always be a critical scale of roughness that reaches yield. By the same argument, there will always be a critical roughness scale which reaches the fully plastic regime. For higher (shorter-wavelength) scales of roughness the graphical method illustrated in figure 3 can again be used to deduce the contact pressure and contact fraction. For sufficiently large n , the true contact area and contact fraction will again approach their asymptotes as shown in figure 4.

It is interesting to note the important role played by interactions between neighbouring asperities in determining the asymptotic behaviour of fine scale

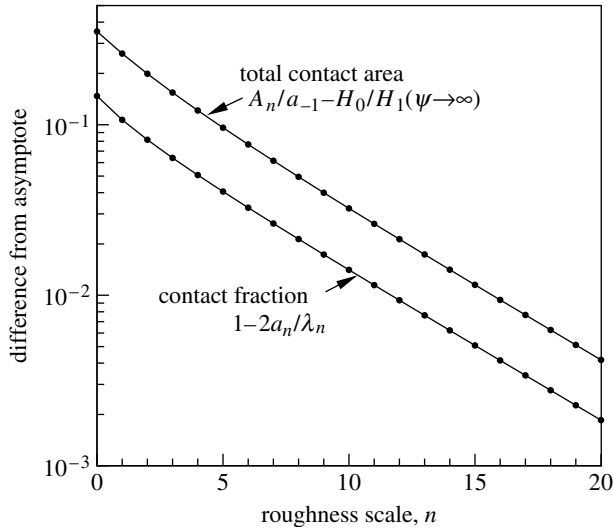


Figure 4. Illustrating the convergence of total contact area and contact fraction towards their rigid plastic asymptotes with roughness scale n .

roughness. Without this interaction, the contact pressure on each scale of roughness is bounded by the material hardness (i.e. the limiting pressure under an isolated indenter). For a rigid plastic contact, therefore, all roughness scales (including $n=0$) must reach full contact $a_n/\lambda_n = 1$ to support the load. For a more general elastic–plastic contact, the total true contact area and the true contact spot size would be determined by the first roughness scale to reach the fully plastic limit, since all shorter-wavelength scales of roughness must be at full contact to support the contact pressure. The true contact spot size could then be computed unambiguously. This, in fact, is the simple picture of plastic contact between two rough surfaces proposed by Tabor (1951), although his focus was on computing the total area of contact, and did not address the question of the contact spot size.

5. Results

We now discuss in detail the predictions of the calculations outlined in the preceding sections. We consider a surface with Weierstrass profile defined in equation (2.1), which is indented by a cylinder of radius R subjected to a load P per unit length out of plane, as illustrated in figure 1. It is convenient to characterize the applied load, surface geometry and material properties using three dimensionless groups:

$$\Sigma = \sqrt{\frac{PE^*}{R\sigma_Y^2}}, \quad \Psi = \left(\frac{G_0 E^*}{A_0 \sigma_Y}\right), \quad K = \left(\frac{\gamma A_0^2}{G_0 R}\right). \tag{5.1}$$

Here, Σ is a dimensionless load parameter, and is proportional to the mean nominal contact pressure. The parameters Ψ and K together characterize the surface roughness and its resistance to plastic flow.

As a preliminary step, we express several parameters of interest in terms of Σ , Ψ and K . If the nominal contact is elastic ($\Sigma < 3.2$) the contact size is $a_{-1}/A_0 = (2/\sqrt{\pi})\gamma\Sigma/(\Psi K)$, while for a fully plastic nominal contact ($\Sigma > 40$) $a_{-1}/A_0 = \gamma\Sigma^2/(2.8\Psi K)$. The wavelength of the zeroth roughness scale (see figure 1) follows as $\lambda_0 = a_{-1}/\gamma$, so we find that the plasticity index ψ_0 (defined in equation (3.3)) for the zeroth roughness scale is given by

$$\psi_0 = \begin{cases} \Psi^D(\sqrt{\pi}K/2\Sigma)^{D-1}, & \text{elastic nominal contact } \Sigma < 3.2, \\ \Psi^D(2.8K/\Sigma^2)^{D-1}, & \text{plastic nominal contact } \Sigma > 40. \end{cases} \quad (5.2)$$

(a) *Conditions for inducing yield*

Controlling plastic flow under contact loading is a central issue in designing failure resistant surfaces for tribological applications. With this motivation, we begin by estimating the conditions necessary to initiate yield in our idealized model of a surface with fractal roughness. With reference to figure 1, we focus attention on each successive scale of roughness under the contact in turn. Strictly, our calculations show that short-wavelength roughness scales will *always* deform plastically, so the surface does not have an elastic limit. Instead, our goal is to calculate the number of roughness scales that remain elastic, as a function of the load (parameterized by Σ) and surface roughness and material properties, parameterized by Ψ and K .

We assume for simplicity that the nominal contact remains elastic, which requires $\Sigma \leq 3.2$ (Johnson 1985). The most heavily loaded region of the zeroth scale of roughness occurs at the centre of the nominal contact, and is subjected to a normalized contact pressure $\bar{p}_0/\sigma_Y = p_{-1}^{\max}/\sigma_Y = \Sigma/\sqrt{\pi}$. The elastic solution for indentation of a sinusoidal surface (§3) shows that the zeroth roughness scale will behave in one of two ways, depending on the value of ψ_0 . For $\psi_0 < 0.2$, the zeroth roughness scale will remain elastic up to full contact, which occurs at a critical load $\bar{p}_0/\sigma_Y = \Sigma/\sqrt{\pi} = \pi\psi_0$. For $\psi_0 > 0.2$, the zeroth scale of roughness will reach yield before full contact. The critical nominal contact pressure at yield is given by $\bar{p}_0/\sigma_Y < \pi\psi_0 \sin^2(\pi/10\psi_0)$.

Similar arguments can be applied to each successive scale of roughness $n = 1, 2, 3, \dots, \infty$. The maximum nominal contact pressure acting on each roughness scale can be computed from equation (4.31); the criteria $p_{n-1}^{\max}/\sigma_Y = \pi\psi_n$ and $p_{n-1}^{\max}/\sigma_Y < \pi\psi_n \sin^2(\pi/10\psi_n)$ give the conditions for full contact and yielding on the n th scale of roughness, respectively.

The results of these calculations are summarized in figure 5, which shows the critical normalized load Σ required to: (i) just cause yield in each roughness scale (shown as solid lines) and (ii) just bring each roughness scale to full contact (shown as dashed lines). The results depend (weakly) on D and γ , and are shown for the particular case $D = 1.5$, $\gamma = 5$. To understand the significance of figure 5, it is helpful to visualize the behaviour of a contact under progressively increasing load. At extremely low values of applied load, the nominal contact width is small, and so all roughness scales under the contact have short wavelength. Since, short wavelengths in the Weierstrass series have a high amplitude/wavelength ratio, these roughness scales easily deform plastically, and so only the zeroth scale of roughness (with longest wavelength) under the contact remains elastic. As the

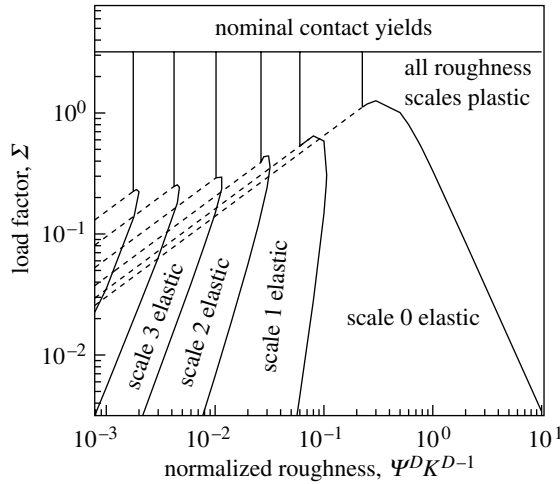


Figure 5. A map showing the behaviour of successive roughness scales during indentation by a smooth cylindrical indenter. Dashed lines show conditions necessary to bring each scale to full contact; solid lines mark transitions from elastic to plastic behaviour.

load is increased, the nominal contact width increases, and so additional roughness scales are brought within the contact. These roughness scales have a longer wavelength, and hence a lower amplitude/wavelength ratio, making them more resistant to plastic flow. Consequently, for a sufficiently large load, and sufficiently smooth surface, the new roughness scales may remain elastic. For example, when a surface with $\Psi^D K^{D-1} = 0.01$ is subjected to a load $\Sigma = 0.1$, scales $n=0, 1, 2$ are all elastic, while all shorter-wavelength roughness scales exceed yield. If the load is increased still further, the roughness scales with longest wavelength under the contact may reach full contact. For the surface with $\Psi^D K^{D-1} = 0.01$, scale $n=0$ reaches full contact at $\Sigma \approx 0.12$, scales $n=0, 1$ reach full contact shortly thereafter. If a roughness scale reaches full contact without yielding, it remains elastic until the plastic zone associated with the nominal contact reaches the surface.

The parameter $\Psi^D K^{D-1}$ evidently plays the role of a plasticity index for the surface: a contact with a low value of $\Psi^D K^{D-1}$ has a large number of roughness scales that deform elastically up to the load where the nominal contact yields, while a contact with high $\Psi^D K^{D-1}$ will experience extensive plastic flow even at modest loads.

For practical purposes, it is of particular interest to estimate the depth of the plastically deformed region under the contact. As a rough guide, the depth of the plastically deformed region can be taken to be one half the wavelength of the first roughness scale to reach yield, since longer wavelengths remain below yield, while all shorter wavelengths will exceed yield. This approximation gives $d_{\text{yield}}/A_0 \approx a_{-1}/(2A_0\gamma^{k+1}) = \Sigma/(\gamma^k\Psi K\sqrt{\pi})$, where k is the number of roughness scales that remain in the elastic regime, which may be estimated from figure 5. The normalized depth of the plastically deformed region as a function of the load parameter Σ is shown in figure 6, for various values of the plasticity index $\Psi^D K^{D-1}$, for $D=1.5$, $\gamma=5$. We observe three general regimes of behaviour: for

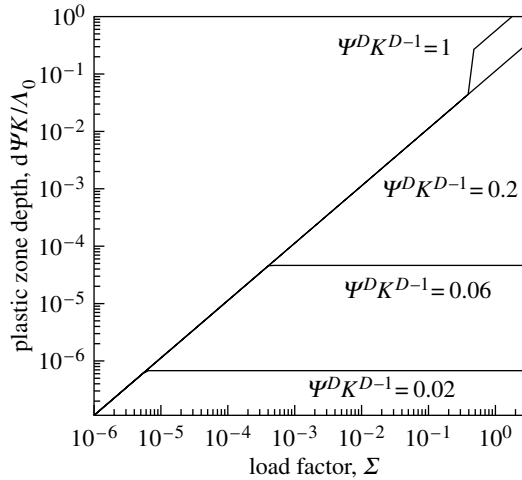


Figure 6. The depth of the plastically deformed layer induced by indenting a Weierstrass surface with a rigid cylinder.

large values of $\Psi^D K^{D-1}$, only the zeroth roughness scale remains elastic at light loads, so the depth of the deformed layer is proportional to Σ . For $\Psi^D K^{D-1} > 0.25$, the zeroth roughness scale yields at large loads, which results in a sudden increase in the depth of the deformed layer. For $0.08 < \Psi^D K^{D-1} < 0.15$, the zeroth scale remains elastic up to full contact, while all remaining scales yield—in this case there depth of the deformed layer is proportional to Σ over the full range. Finally, for $\Psi^D K^{D-1} < 0.08$, the one or more roughness scales transition to elastic behaviour at sufficiently large loads. Under these conditions the depth of the deformed layer is proportional to the load parameter Σ at light loads, but is independent of Σ above a critical load.

(b) Contact pressure distribution functions

We proceed next to examine the contact pressure distributions in more detail. The pressure acting on each scale of contact is completely characterized by the distribution functions $q_n(\xi)$ specifying the fraction of the contact. For an elastic surface, $q_n(\xi)$ are smooth functions (see e.g. figure 5 of Ciavarella *et al.* 2000). For plastic contacts, $q_n(\xi)$ contains singularities (this is a consequence of the fact that contact pressure under fully plastic contacts is uniform—see electronic supplementary material for a more detailed discussion). Consequently, it is more convenient to plot the cumulative contact pressure distribution functions $Q_n(p/\sigma_Y)$, which specify the proportion of each scale of contact over which the pressure exceeds p/σ_Y . The cumulative contact pressure distributions can be computed from the definition $q_n = -\sigma_Y \partial Q_n / \partial p$. Results are shown in figure 7 for a dimensionless nominal pressure $\Sigma = 0.001$ and two values of surface roughness, with $\gamma = 5$ and $D = 1.5$. figure 7a shows results for $\psi_0 = 0.0002$, in which case a large number of scales under the contact are elastic. Under these conditions, the pressure distributions for roughness scales $n = 0, \dots, 9$, which deform elastically, tend towards a self-similar distribution. (Ciavarella *et al.* 2000). The most heavily loaded region of the 10th roughness scale begins to deform plastically,

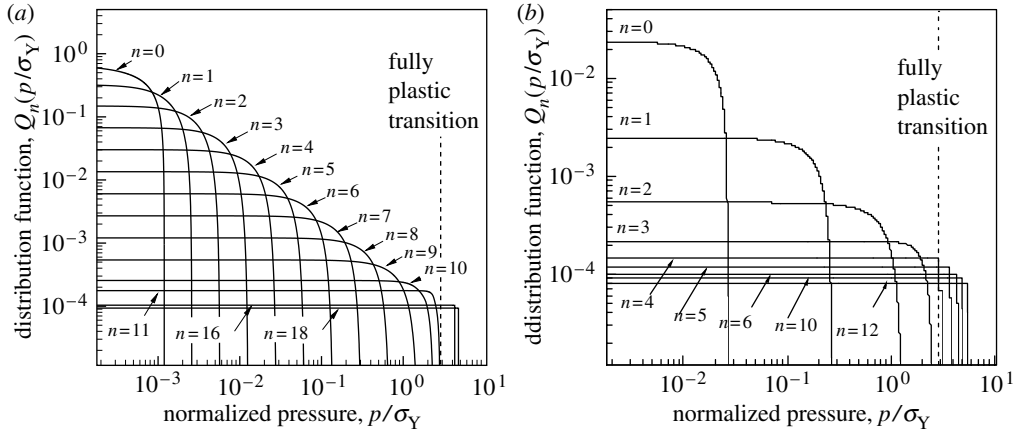


Figure 7. Cumulative contact pressure distributions $Q_n(p/\sigma_Y)$ which specify the proportion of each scale of roughness over which the pressure exceeds p/σ_Y . Results are shown for a cylindrical indenter subjected to nominal load $\Sigma=0.0001$ and for surfaces with (a) $\psi_0=0.0002$ and (b) $\psi_0=0.1$.

while the 11th reaches the fully plastic limit (which occurs when the contact pressure reaches $2.9\sigma_Y$), and begins to transition towards a uniform distribution. The pressure acting on subsequent scales becomes increasingly uniform, and gradually increases towards the limiting value $5.8\sigma_Y$. A contrasting result for a rougher surface with $\psi_0=0.1$ is shown in figure 7b. In this case, only scales $n=0, \dots, 1$ are elastic; scale $n=2$ exceeds yield, and a small fraction of scale $n=3$ reaches the fully plastic limit. Scale $n=4$ is dominated by contacts in the fully plastic regime. The pressure acting on subsequent scales of roughness again becomes progressively more uniform, and increases towards the limit $5.8\sigma_Y$. Pei *et al.* (in press) have noted the tendency of plasticity to drive contact size and pressure distributions to uniform distributions in numerical simulations (they refer to plasticity as a ‘grand equalizer’). Our model suggests an explanation for this phenomenon. The behaviour of a sinusoidal surface is shown in figure 2, and is approximated by equations (A2)–(A18) in the electronic supplementary material. Evidently, nominal pressures in the range $2H_0/\psi < p/\sigma_Y < 2H_0/3$ acting on the sinusoid all induce the same true contact pressure. The first roughness scale to reach fully plastic behaviour, therefore, exerts a nearly uniform pressure on the next scale of roughness.

(c) Average contact size, contact pressure, total contact area, and number of contacts for each scale

The overall behaviour of the contacts at each scale may be computed using the distribution functions given in §5b, and may also be estimated using the approximate procedure outlined in §4b. Figures 8 and 9 show the results of these calculations. Figures 8a and 9a show the contact fraction $2\langle a_n \rangle / \lambda_n$ and the ratio of true to nominal contact area $A_n/2a_{-1}$ for successive roughness scales n under a contact; while figures 8b and 9b show the total number of contacts on each successive scale, normalized by the number of contacts in the zeroth roughness scale, N_n/N_0 , and the dimensionless nominal contact pressure $P/\sigma_Y A_n$ on each scale. All variables have been plotted as a function of $\psi_n = \psi_0 \gamma^{(D-1)n}$ to illustrate

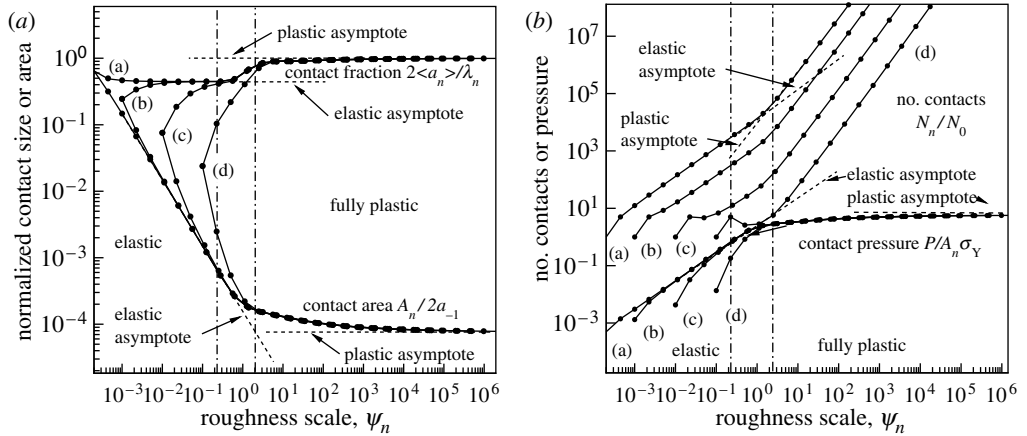


Figure 8. The variations of (a) normalized average contact size and total contact area and (b) the number of contact spots and normalized nominal contact pressure, for successive roughness scales under a cylindrical indenter. The chain lines mark the transitions from elastic to elastic-plastic behaviour, and from elastic-plastic behaviour to fully plastic behaviour. Results are shown for nominal pressure $\Sigma=0.001$ and roughness values (a) $\psi_0=0.0002$, (b) $\psi_0=0.001$, (c) $\psi_0=0.01$ and (d) $\psi_0=0.1$.

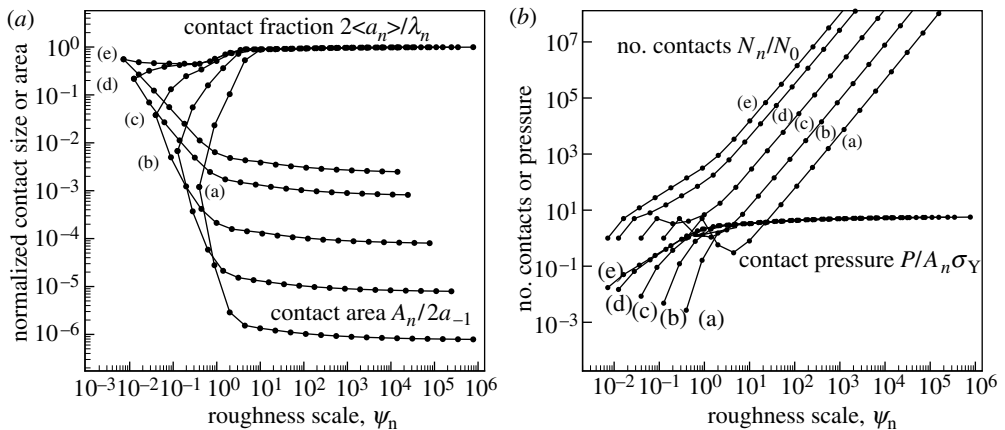


Figure 9. The variations of (a) normalized average contact size and total contact area, and (b) the number of contact spots and normalized nominal contact pressure, for successive roughness scales under a cylindrical indenter. Results are shown for constant normalized roughness $\Psi^D K^{D-1} = 10^{-3}$ and progressively increasing nominal pressure (a) $\Sigma = 10^{-5}$, (b) $\Sigma = 10^{-4}$, (c) $\Sigma = 10^{-3}$, (d) $\Sigma = 10^{-2}$ and (e) $\Sigma = 0.03$.

their asymptotic behaviour clearly. Results are shown in figure 8 for a nominal contact pressure $\Sigma=0.001$ and for several values of surface roughness (parameterized by $\psi_0 = \Psi^D (\sqrt{\pi} K / 2\Sigma)^{D-1}$); while figure 9 shows results for a fixed dimensionless surface roughness $\Psi^D K^{D-1} = 10^{-3}$ and progressively increasing load Σ .

To interpret the figures, note that each curve labelled (a)–(d) corresponds to a different surface. The circles on each curve show the values of variables on

successive scales of roughness $n=0, 1, 2, \dots$, obtained using the procedure outlined in §4*b*, while the solid lines show the exact procedure described in §4*a* and electronic supplementary material. As a representative example, consider case (a) in figure 8. This is a very smooth surface, so the zeroth roughness scale (the left-most point on the curve) has a relatively high contact fraction $2 < a_0 > / \lambda_0 \approx 0.5$. Note also that the contact area and number of contact spots for this scale must always satisfy $A_0/2a_{-1} = 2 < a_0 > / \lambda_0$ and $N/N_0=1$. Finally, since the surface is relatively smooth, and the value $\Sigma=0.001$ corresponds to a light load, the nominal contact pressure on the zeroth scale for case (a) has a low value $P/A_0\sigma_Y \approx 8 \times 10^{-4}$. Successive circles on the curve marked (a) show the behaviour of subsequent roughness scales in the same way. For this surface, a large number of roughness scales (10) remain elastic, so the contact fraction and area for these scales approach the elastic asymptotic limits $2\langle a_n \rangle / \lambda_n \rightarrow \gamma^{(1-D)}$ and $A_n/2a_{-1} \rightarrow \Sigma/\pi^{5/2}\psi_n$, respectively, while the number of contact spots and nominal pressure approach $N_n/N_0 \rightarrow \Sigma\gamma^{(2-D)n}/\pi^{5/2}\psi_0$ and $P/A_n\sigma_Y \rightarrow \pi^3\psi_n/4$. The transition from elastic to elastic–plastic behaviour occurs at $\psi_n \approx 0.1\gamma^{(D-1)}$, marked as a chain line in figure 8 (this approximation is strictly valid only if the contact size reaches its elastic asymptote). For case (a) the 11th and 12th roughness scales exceed yield, and are loaded in the elastic–plastic regime. Finally, the transition to fully plastic behaviour occurs near $\psi_n \approx 2$, and for case (a) scales 13 and higher are in the fully plastic regime. For these scales, the contact size and contact area approach the plastic asymptotes: $2\langle a_n \rangle / \lambda_n \rightarrow 1$; $A_n/2a_{-1} \rightarrow \sqrt{\pi}\Sigma/4H_1(\infty) \approx 0.076\Sigma$, while the number of contact spots and the normalized nominal pressure approach $N_n/N_0 \rightarrow \gamma^n$ and $P/A_n\sigma_Y \rightarrow H_1(\infty) = 5.8$. Contrasting behaviour for a rough surface under the same load is shown in curve (d) in figure 8. In this case, only a scale of roughness ($n=0$) is elastic; scale $n=1$ reaches yield, so the elastic asymptote is not reached. Scales $n=4$ and higher exhibit fully plastic behaviour.

The behaviour of roughness scales under a surface with fixed roughness, which is indented by a cylinder under progressively increasing load is illustrated in figure 9. Results are shown for roughness $\Psi^D K^{D-1} = 10^{-3}$ and five different loads. At the lightest load ($\Sigma=10^{-5}$) only the zeroth scale of roughness is in the elastic regime; there are 2–3 scales in the elastic–plastic regime, but the majority of roughness scales lie in the fully plastic regime and the contact size, area, pressure and the number of contact spots quickly approach their fully plastic asymptotes. As load is increased, additional roughness scales with a lower ratio of amplitude to wavelength are brought within the contact, so the number of scales in the elastic regime increases with load. At $\Sigma=0.03$, there are six roughness scales in the elastic regime: a sufficient number to allow the contact size, area, pressure and the number of contact spots approach their elastic asymptotes. Shorter wavelengths again transition to the fully plastic regime.

6. Discussion

Ciavarella *et al.* (2000) showed that the contact between two elastic, perfectly fractal surfaces consists of an infinite number of contact spots, with zero size, subjected to infinite contact pressure. This paper was motivated by the hope that accounting for plastic flow would resolve these difficulties. We have found,

however, that the total contact area and contact pressure between elastic–plastic solids are well defined, but it remains impossible to predict the actual number of contacts or their size. The ambiguous nature of the true contact could be resolved in either of two ways: (i) it is possible that some physical process, which was not included in our contact model, can regularize the ill posed nature of contact even for perfectly fractal surfaces. For example, Persson *et al.* (2005) shows that accounting for adhesion has this effect, for surfaces with $D < 1.5$. (ii) No real surface can be perfectly fractal, and provided that the surface geometry can be measured, and characterized, at extremely short wavelengths where the fractal description breaks down, standard techniques—including perhaps the Greenwood–Williamson method—can be used to model the contact. Available experimental studies (Majumdar & Bhushan 1990; Bora *et al.* 2004) provide strong evidence that surfaces cease to be fractal at nanometre length scales.

Although, it may ultimately be necessary to abandon perfectly fractal descriptions of surface roughness, it is clear that a fractal description of a surface has several advantages over the conventional approach; and that some features of fractal behaviour must be retained in a complete analysis. In particular, the fractal description captures the important fact that asperity contacts are spatially correlated. In the Greenwood–Williamson description, the focus is entirely on the smallest scale of contacts, which are assumed to be uniformly distributed spatially, and to exert a nominally uniform pressure. In contrast, the fractal description, together with the Ciavarella model, allows one to calculate the collective effect of the contact clusters, without needing to describe the contacts themselves in detail. There are also some applications where the behaviour of the surfaces is determined by long-range surface roughness, and is insensitive to the detailed nature of the contact spots themselves. For example, contact compliance, and contact resistance, are determined by long-range roughness, and may converge even for elastic surfaces with fractal roughness (Ciavarella *et al.* 2004; Manners & Gholami 2005)

Our analytical calculations also help to elucidate the behaviour that one would expect to observe in a numerical simulation of contact between rough, ideally fractal, surfaces. The calculations of Ciavarella *et al.* (2000), as well as those reported here, raise some questions regarding the interpretation of such simulations. It appears to be an inescapable conclusion that the true contact area for elastic contact between two perfectly fractal surfaces consists of an infinite number of contact spots of zero size, which are subjected to infinite contact pressure, and have a vanishing total area of contact. Numerical simulations that predict contact sizes, contact pressures, or the true area of contact between two elastic solids are, therefore, valid only if: (i) the actual surface geometry ceases to be fractal at some short wavelength (e.g. the power-spectral density could be truncated at some short wavelength; or else decay sufficiently rapidly to give a differentiable autocorrelation function at the origin); and (ii) the resolutions of both the surface profilometre used to measure the three-dimensional surface, as well as the numerical simulations, are sufficient to detect and to characterize surface geometry at these short wavelengths. Numerical simulations of elastic–plastic contact are somewhat less sensitive to artificial truncation of the roughness spectrum, but the predicted total number of contacts, and their size, would not converge for a perfectly fractal surface.

The qualitative effects of introducing an artificial short-wavelength cut-off into numerical simulations of contact between fractal surfaces can be estimated using our simple model. As a representative example, we consider a Weierstrass surface (equation (2.6)) containing only terms $n=0, 1, 2, \dots, 5$, which is indented by a perfectly flat, rigid indenter under a uniform nominal pressure p_{-1} . We restrict attention to modest nominal pressures, so that the true area of contact is much less than the nominal contact area. Under these conditions, the true contact consists of a well-defined set of contact spots. The average size of these spots, and the true contact pressure, are essentially independent of nominal pressure, while the number of contact spots and the total true contact area increase in proportion to the nominal pressure. The contact size and mean contact pressure, as well as the relationship between true contact area and the number of contact spots, is determined by the elastic and plastic properties of the surface and the surface roughness, which may be characterized together by ψ_0 . We find that for $\psi_0 < 0.006$, the surface deforms elastically: under these conditions the behaviour of the contacts can be estimated using the expressions given in §4c, with $n=5$. For $\psi_0 > 0.006$, the contact spots exceed yield. The effect of plastic flow is to increase the size and number of contacts, to increase the true area of contact, and reduce the mean pressure. As ψ_0 increases, the true contact progressively transitions to the limiting conditions for a rigid plastic surface: the contact fraction for the fifth scale approaches $a_5/\lambda \rightarrow 1$, the true contact pressure approaches $\langle p_5 \rangle = 5.8\sigma_Y$ and the ratio of true to nominal contact area approaches $p_{-1}/5.8\sigma_Y$. The transition with ψ_0 is gradual, however. To illustrate this transition from elastic to plastic behaviour, we have plotted the variation of the true contact pressure $P/A_n\sigma_Y\psi_0 = P\lambda_0/A_nE^*y_0$ as a function of $1/\psi_0 = \sigma_Y\lambda_0/E^*y_0$ in figure 10 for a surface with $D=1.5$ and $\gamma=5$. Results are shown for surfaces truncated at scales $n=5, 8$ and 11 . We have chosen to plot the results in this form to compare our predictions with the detailed finite element simulations of indentation of a nominally fractal surface reported by Pei *et al.* (in press). In particular, their fig. 4 shows the predicted true contact area as a function of the ratio of yield stress to elastic modulus σ_Y/E . The numerical results are in excellent qualitative agreement with our estimate. They observe a clear transition from elastic behaviour at high σ_Y/E to plastic behaviour at low σ_Y/E ; moreover, they predict $P/A_n \rightarrow 6\sigma_Y$ in the limit $\sigma_Y/E \rightarrow 0$, in excellent agreement with our estimate of $P/A_n \rightarrow 5.8\sigma_Y$. In our calculations, the elastic-plastic transition is a direct consequence of truncating the Weierstrass series, however. As the series is truncated at progressively shorter wavelengths, the elastic regime disappears. These results illustrate the challenges associated with numerical simulations of contact between fractal surfaces.

The simple model we have developed here could be extended in several ways. For example, it would be straightforward to extend the calculations described here to three dimensions. In addition, it would not be difficult to extend our calculations to account for the scale dependent nature of plastic flow: due to indentation size effects (Nix & Gao 1998), we might expect short-wavelength roughness to return to elastic behaviour. It would be of great interest to determine where this transition occurs, because then it would be possible to design surfaces that remain elastic. Finally, the general procedure employed here is clearly not restricted to fractal surfaces, and could be applied to any discrete

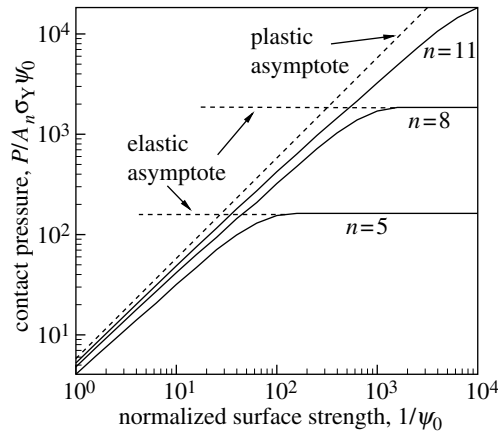


Figure 10. The dimensionless true contact pressure for a truncated Weierstrass surface, as a function of the dimensionless strength parameter $1/\psi_0$.

approximation to a power spectral density, provided that the discrete scales remain widely separated.

A more important issue, however, is the need to adopt a more realistic model of the surface roughness. The Weierstrass profile gives a convenient way to characterize the multi-scale nature of surface roughness, but is clearly not a realistic model of surface roughness. Furthermore, the scale decoupling introduced by Ciavarella *et al.* (2000), which forms the basis of our analysis, strictly holds only in the limit $\gamma \rightarrow \infty$. A number of questions, therefore, need to be addressed. Firstly, it is of interest to determine the range of γ for which the scale decoupling can be applied. For elastic surfaces, this is straightforward. The scales will decouple as long as there are sufficient contacts of scale $n+1$ within each contact of scale n . Recall that (for sufficiently large n), the contact spot size for the n th scale can be estimated as $2 < a_n > = \lambda_n \gamma^{(1-D)}$. The number of contact spots of scale $n+1$ within each contact of scale n follows as $m = 2 < a_n > / \lambda_{n+1} \approx \gamma^{(2-D)}$. Consequently, the critical value of γ to ensure m contact spots is $\gamma \geq m^{1/(2-D)} \approx m^2$. Taking $m=2$ as an optimistic estimate requires $\gamma \geq 4$; in practice, an estimate of $\gamma \geq 10$ is more reasonable. It is difficult to obtain similar analytical estimates for plastic contacts, so instead, we have investigated this issue using finite element computations with a finite number of roughness scales, some of which are described in electronic supplementary material. Simulations with two and three scales of roughness suggest that scales of roughness completely decouple for $\gamma \geq 10$. For smaller γ , we find that a few asperities in the n th scale that are located near the peaks of scale $n-1$ may coalesce, in contrast to the large γ approximation. This is because at light loads, the n th scale contains only a few contacts, and insufficient lateral constraint develops to increase the contact pressure for the finest scale. However, asperities on scale n that are located in the valleys of scale $n-1$ are strongly constrained, and so exhibit behaviour that is in good agreement with the behaviour of a surface with purely sinusoidal roughness. Consequently, the predictions of the scale decoupling analysis continue to predict qualitatively the *average* behaviour of each roughness scale even for low values of γ . Nevertheless, further work is required to properly characterize the behaviour of random surfaces, for both elastic and plastic solids.

7. Conclusions

We have outlined an approximate model that is intended to estimate the behaviour of a rough elastic–perfectly plastic solid with (two-dimensional) self-affine fractal surface roughness subjected to contact loading. The surface was idealized by a Weierstrass profile (equation (2.1)) with widely separated scales of roughness ($\gamma \gg 1$). Under these conditions, each scale of roughness can be regarded as a sinusoidal surface, indented by a flat surface. Successive scales are coupled through the condition that the nominal pressure acting on scale n is equal to the true pressure acting on scale $n-1$. On this basis, a recursive expression can be found for the contact pressure distribution, the contact size, and the total area of contact for each scale of roughness, in terms of the response of a sinusoidal surface (Ciavarella *et al.* 2000). In our calculations, the response of an elastic–plastic solid with sinusoidal roughness was approximated by fitting curves to finite element computations, as outlined in §3. We found that simple but highly accurate estimates for the average contact fraction, pressure and area of contact on successive roughness scales can be obtained by replacing the pressure distribution acting on each scale by a statically equivalent uniform distribution.

We have used this approach to analyse the behaviour of an elastic–perfectly plastic Weierstrass surface when indented by a rigid cylinder. We found that the behaviour of the solid is governed by three dimensionless functions, defined in equation (5.1). The parameter Σ quantifies the magnitude of the applied load; while Ψ and K together characterizes the surface roughness and its resistance to plastic flow. The conditions necessary to induce plastic flow in successive roughness scales under the cylinder are illustrated in figure 5. For light loads, the nominal contact area is small, so roughness scales under the indenter have a large amplitude compared with their wavelength, and consequently a low resistance to plastic flow. As a result, at light loads only the zeroth scale of roughness deforms elastically; all other roughness scales are fully plastic. As the applied load is increased, the nominal contact area spreads, and new roughness scales that come into the contact area may remain elastic, and for sufficiently high loads will reach full contact. The parameter $\Psi^D K^{D-1}$ was identified as a plasticity index for a surface with fractal roughness. Surfaces with low values of $\Psi^D K^{D-1}$ have a large number of scales of roughness in the elastic regime, and consequently plastic flow is restricted to a thin layer at the surface, figure 6. As $\Psi^D K^{D-1}$ increases, the number of plastic roughness scales, and the corresponding depth of the plastically deformed layer, increase.

We also calculated the average contact fraction, total contact area, number of contacts, and average contact pressure for successive scales of roughness. The results are illustrated in figures 8 and 9. The behaviour of the surface depends primarily on the dimensionless parameter ψ_0 defined in equation (3.3), which characterizes the resistance to plastic flow of the longest-wavelength scale of roughness under the nominal contact. This parameter can be computed in terms of Ψ and K using equations (5.2). For low values of ψ_0 , a large number of long-wavelength roughness scales under the contact remain elastic: their behaviour can be calculated using the closed-form expressions given in §4c. Short-wavelength roughness scales always deform plastically, and for large n approach the asymptotes given in §4e. For an elastic–plastic solid with a

perfectly fractal surface, the true area of contact consists of an infinite number of contact spots of zero size, with a true contact pressure $\lim_{n \rightarrow \infty} P/A_n = 5.8\sigma_Y$. The true contact area is bounded, and proportional to the load.

The implications of these findings were discussed in detail in §6. Our calculations suggest that idealizing surface roughness as a perfect fractal leads to unphysical predictions for the contact size and number of contact spots. It is likely that these difficulties can only be resolved by taking into account the departure of actual surfaces from perfectly fractal behaviour at very short wavelengths.

This work was performed as part of the General Motors Collaborative Research Laboratory on Computational Materials Research at Brown University. The authors are grateful to M. Ciavarella, R. W. Carpick and J. F. Molinari for helpful discussions and for providing pre-prints of their recent publications.

References

- Archard, J. F. 1957 Elastic deformation and the laws of friction. *Proc. R. Soc. A* **243**, 190–205.
- Berry, M. V. & Lewis, Z. V. 1980 On the Weierstrass–Mandelbrot fractal function. *Proc. R. Soc. A* **370**, 459–484.
- Bora, C. K., Plesha, M. E., Flater, E. E., Street, M. D., Carpick, R. W. & Redmond, J. M. 2004 Multiscale roughness of MEMS surfaces. In *Proceeding of 2004 ASME/STLE Joint International Tribology Conference*, ref. no. TRIB2004-64370. New York: ASME.
- Borri-Brunetto, M., Carpinteri, A. & Chiaia, B. 1999 Scaling phenomena due to fractal contact in concrete and rock fractures. *Int. J. Fract.* **95**, 221–238. (doi:10.1023/A:1018656403170)
- Bush, A. W., Gibson, R. D. & Thomas, T. R. 1975 The elastic contact of a rough surface. *Wear* **35**, 87–111. (doi:10.1016/0043-1648(75)90145-3)
- Childs, T. H. C. 1973 The persistence of asperities in indentation experiments. *Wear* **25**, 3–16. (doi:10.1016/0043-1648(73)90115-4)
- Childs, T. H. C. 1977 The persistence of roughness between surfaces in static contact. *Proc. R. Soc. A* **353**, 35–53.
- Ciavarella, M. & Demelio, G. 2001 Elastic multiscale contact of rough surfaces: Archard’s model revisited and comparisons with modern fractal models. *ASME J. Appl. Mech.* **68**, 496–498.
- Ciavarella, M., Demelio, G., Barber, J. R. & Jang, Y. H. 2000 Linear elastic contact of the Weierstrass profile. *Proc. R. Soc. A* **456**, 387–405. (doi:10.1098/rspa.2000.0522)
- Ciavarella, M., Murolo, G., Demelio, G. & Barber, J. R. 2004 Elastic contact stiffness and contact resistance for the Weierstrass profile. *J. Mech. Phys. Solids* **52**, 1247–1265. (doi:10.1016/j.jmps.2003.12.002)
- Gao, Y. F., Bower, A. F., Kim, K.-S., Lev, L. & Cheng, Y. T. In press. The behavior of an elastic–perfectly plastic sinusoidal surface under contact loading. *Wear*.
- Greenwood, J. A. 1984 A unified theory of surface roughness. *Proc. R. Soc. A* **393**, 133–157.
- Greenwood, J. A. & Williamson, J. B. P. 1966 Contact of nominally flat surfaces. *Proc. R. Soc. A* **295**, 300–319.
- Hill, R. 1950 *The mathematical theory of plasticity*. Oxford: Oxford University Press.
- Hyun, S., Pei, L. & Molinari, J. F. 2004 Finite-element analysis of contact between elastic self-affine surfaces. *Phys. Rev.* **70**, 026117.
- Johnson, K. L. 1985 *Contact mechanics*. Cambridge: Cambridge University Press.
- Majumdar, A. & Bhushan, B. 1990 Role of fractal geometry in roughness characterization and contact mechanics of surfaces. *ASME J. Tribol.* **112**, 205–216.
- Majumdar, A. & Bhushan, B. 1991 Fractal model of elastic–plastic contact between rough surfaces. *ASME J. Tribol.* **113**, 1–11.

- Majumdar, A. & Tien, C. L. 1990 Fractal characterization and simulation of rough surfaces. *Wear* **136**, 313–327. (doi:10.1016/0043-1648(90)90154-3)
- Manners, W. & Gholami, B. 2005 Constriction resistance between materials with fractal patterns of contact. *Sci. Meas. Technol. Proc. IEEE* **152**, 161–168. (doi:10.1049/ip-smt:20045034)
- McCool, J. I. 1986 Comparison of models for the contact of rough surfaces. *Wear* **107**, 37–60. (doi:10.1016/0043-1648(86)90045-1)
- Mihailidis, A., Bakolas, V. & Drivakos, N. 2001 Subsurface stress field of a dry line contact. *Wear* **249**, 546–556. (doi:10.1016/S0043-1648(01)00542-7)
- Nayak, P. R. 1971 Random process model of rough surfaces. *ASME J. Lubr. Technol.* **93**, 398–407.
- Nix, W. D. & Gao, H. 1998 Indentation size effects in crystalline materials: a law for strain gradient plasticity. *J. Mech. Phys. Solids* **46**, 411–425. (doi:10.1016/S0022-5096(97)00086-0)
- Pei, L., Hyun, S., Molinari, J. F. & Robbins, M. O. In press. Finite element modeling of elasto-plastic contact between rough surfaces. *J. Mech. Phys. Solids*.
- Persson, B. N. J. 2001 Elastoplastic contact between randomly rough surfaces. *Phys. Rev. Lett.* **87**, 116101. (doi:10.1103/PhysRevLett.87.116101)
- Persson, B. N. J., Albohr, O., Tartaglino, U., Volokitin, A. I. & Tosatti, E. 2005 On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. *J. Phys: Condens. Matter* **17**, R1–R62. (doi:10.1088/0953-8984/17/1/R01)
- Polonsky, I. A. & Keer, L. M. 2000a A fast and accurate method for numerical analysis of elastic layered contacts. *ASME J. Tribol.* **122**, 30–35. (doi:10.1115/1.555323)
- Polonsky, I. A. & Keer, L. M. 2000b Fast methods for solving rough contact problems: a comparative study. *ASME J. Tribol.* **122**, 36–41. (doi:10.1115/1.555326)
- Sayles, R. S. 1996 Basic principles of rough surface contact analysis using numerical methods. *Tribol. Int.* **29**, 639–650. (doi:10.1016/0301-679X(96)00016-3)
- Sayles, R. S. & Thomas, T. R. 1978 Surface topography as a nonstationary random process. *Nature* **271**, 431–434. (doi:10.1038/271431a0)
- Sutcliffe, M. P. F. 1988 Surface asperity deformation in metal forming processes. *Int. J. Mech. Sci.* **30**, 847–868. (doi:10.1016/0020-7403(88)90010-0)
- Tabor, D. 1951 *Hardness of metals*. Oxford: Oxford University Press.
- Westergaard, H. M. 1939 Bearing pressures and cracks. *ASME J. Mech.* **6**, 49–53.
- Whitehouse, D. J. & Archard, J. F. 1970 The properties of random surface of significance in their contact. *Proc. R. Soc. A* **316**, 97–121.
- Willner, K. 2004 Elasto-plastic normal contact of three-dimensional fractal surfaces using halfspace theory. *ASME J. Tribol.* **126**, 28–33. (doi:10.1115/1.1631019)

As this paper exceeds the maximum length normally permitted,
the authors have agreed to contribute to production costs.