Wear, 26 (1973) 305–333

RANDOM PROCESS MODEL OF ROUGH SURFACES IN PLASTIC CONTACT

P. R. NAYAK*

Bolt Beranek and Newman Inc., Cambridge, Mass. 02138 (U.S.A.)

(Received February 23, 1973)

SUMMARY

The plastic contact of a rough surface and a hard, smooth flat is analyzed by modeling the rough surface as an isotropic, Gaussian, random process. The applicability of this model to the contact of two rough surfaces is discussed, and it is shown that the model is appropriate.

It is not necessary to analyze interactions of asperity pairs, with the attendant questions of their misalignment, the shape of their caps, etc. Instead, a model involving the interaction of the continuous surfaces is developed, which implicitly takes into account these geometrical factors, as well as allowing for the possibility of the coalescence of microcontacts as the normal pressure is increased.

Approximate relations between the density of finite contact patches, their mean area and mean circumference, and the normal pressure/hardness ratio are derived. These relations depend not only on the density and height distribution of maxima, but also on the shape of the Power Spectral Density of the surface. Many surfaces of interest are likely to give rise to multiply-connected contact patches at all except very high separations. The density of holes appearing within the contact patches as well as their area is estimated.

Results are derived for surfaces that may be partitioned into two components, one with a large r.m.s. value and a narrow roughness spectrum, and the other with a small r.m.s. value and an arbitrary spectrum. For these surfaces, the density of holes at small separations becomes equal to the density of finite contact patches; the area of the holes remains small, however. It is conjectured that for surfaces that may not be partitioned in this manner, conventional models of contact are inapplicable. Specifically, the contact patches are likely to be perforated by holes at all separations, the hole area being a significant fraction of the contact area. Unit events such as the contact or collision of asperities also appear to become meaningless

1. INTRODUCTION

In a recent paper[1] a two-dimensional random process model of a rough surface was developed, based on the work of Longuet-Higgins[2]. This model is used in an analysis of the static contact of plastic rough surfaces.

---

* Present address: Engineering Research Centre, TELCO, Poona, India.

A literature survey on the plastic contact of rough surfaces[3–5] shows that it has been considered necessary to use an asperity-based model of surface topography to obtain such statistics as the density of microcontacts and their mean area. An asperity-based model assumes the rough surface to consist of a mean plane with independent (*i.e.*, not touching) hills and valleys randomly distributed on it; it is usually assumed that the caps of the asperities are spherical.

This asperity model (with its attendant assumptions) is unnecessary, and is often invalid.

By assuming that the surface is a Gaussian random process, an assumption increasingly supported by experimental evidence, relationships are obtained between such statistics as the real area of contact, the density of contacts, their mean area, a shape factor for the contacts, and the normal pressure. For an isotropic surface, these relationships depend fairly strongly on the shape of the surface Power Spectral Density (PSD). In particular, a question of importance is whether or not the surface roughness can be partitioned into a dominant narrow-band roughness and a superposed, low-level, roughness. If this cannot be done, *i.e.*, if the surface PSD is truly broad-band with no dominant (narrow) band of wavelengths, it appears that conventional descriptions of contact in terms of convex, singly-connected contact patches are likely to be inapplicable.

An important part of the model of plastic contact proposed here is an observation of Pullen and Williamson[6] that when an incremental deformation of the asperity tips occurs, the incremental plastic volume reappears as a uniform rise in that part of the surface that is not in contact. However, certain other modes of plastic deformation could be assumed, including the popular (but wrong) one, that the plastically deformed volume simply vanishes; the quintessence of these models being that the non-contacting surface is not distorted.

Another important part of the model is another observation of Pullen and Williamson[6], that the apparent normal pressure is not proportional to the fraction (real area/apparent area). Through analysis and experiment, they demonstrate that at high apparent normal pressures the mean pressure on the microcontacts may be significantly greater than the hardness H without collapse of the asperities. A specific relation between the pressure and the real area is obtained in ref. 6 and is used here. Other empirically determined relations could be used in the model without destroying its validity.

What emerges from the analysis is a picture of plastic contact that explicitly exhibits the coalescence of many small contacts into fewer large ones as the normal pressure is increased. The predicted density of contacts is thus significantly lower, except at low pressures, than would be predicted by an asperity model of contact. It is shown that at small mean-plane separations, the microcontacts (or contact patches as they will be termed to make explicit the possibility that they are multiply connected) are perforated throughout by holes, the density of these holes being approximately equal to the density of contact patches. It is also shown that for most surfaces, the average contact patch is likely to be markedly noncircular in shape, at all except the lightest loads.

Though the analysis presented here and the attendant conclusions are strictly valid only for isotropic, Gaussian surfaces, an examination of the physical phenomena underlying the conclusions shows that they are likely to be qualitatively correct for non-Gaussian surfaces, too.

Section 2 briefly summarizes the observations and conclusions of Pullen and Williamson[6], the main goal being to demonstrate that the plastic contact of a rough surface and a hard flat at some separation is roughly equivalent (insofar as contact statistics are concerned) to the intersection of the rough surface and an imaginary plane at some other "equivalent" separation. It is further shown that the contact of two rough surfaces is equivalent to the contact of a "composite" rough surface and a hard flat, and therefore, to the intersection of that composite rough surface and an imaginary plane.

Section 3 examines in detail the intersection of an isotropic, Gaussian rough surface, and an imaginary plane, the main goals being to obtain the density of contact patches, to roughly assess their shape, and to tackle the question of whether or not the contact patches have holes in them.

Section 4 presents a brief example illustrating theory. Finally, an extensive discussion of the results of the model and of the assumptions it makes is presented in Section 5.

2. PLASTIC CONTACT OF ROUGH SURFACES

2.1. One rough surface and a hard, smooth flat

Some recent observations of Pullen and Williamson[6], discussed in some detail by Tallian[7] are of great use in the development of a random process model of rough surfaces in plastic contact.

Consider a rough surface contacting a hard, smooth flat, the separation of the flat and the mean-plane of the rough surface being y. If σ be the r.m.s. roughness of the surface, the dimensionless separation is $y^* = y/\sigma$. Pullen and Williamson found in a fascinating experimental study that the volume of metal plastically deformed during an incremental approach of the surfaces reappears as a uniform incremental rise in the untouched part of the surface. This mode of deformation results in the following conceptual model of contact. Assuming the lower mean plane P_1 in Fig. 1 to remain fixed (because of the redistribution of plastic volume), the upper plane P_2 is lowered until the dimensionless mean-plane separation is y^*. Simultaneously, the entire lower surface rises by an amount u, such that the volume of the rough surface per unit area eventually lying above the imaginary plane P_2 is equal to u. This conceptual model is, in turn equivalent to a model wherein the rough surface intersects the plane P_2, but at an effective dimensionless separation

$$z_0^* = y^* - u/\sigma ,$$

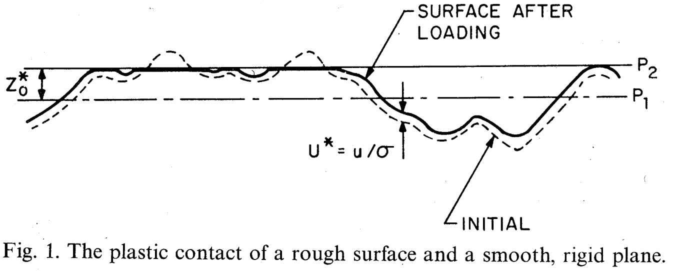


Fig. 1. The plastic contact of a rough surface and a smooth, rigid plane.

such that

$$y^* = z_0^* - \tfrac{1}{2} z_0^* \operatorname{erfc}(z_0^*/2^{\frac{1}{2}}) + (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2} z_0^{*2})$$

Figure 2 shows z_0^* as a function of y^*. Clearly, for a separation y^*, the real area of contact per unit mean-plane area is found from

$$A = \int_{z_0^*(y^*)}^{\infty} p(z^*) \mathrm{d}z^* ,$$

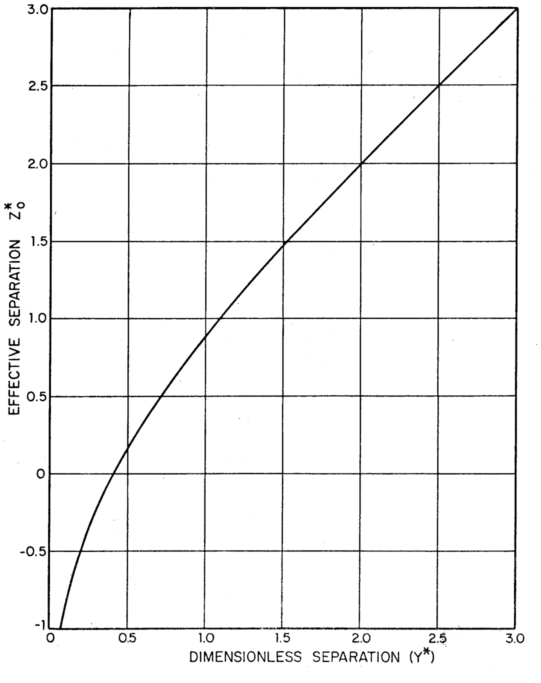
where $p(z^*)$ is the height-distribution of the surface. Taking this distribution to be Gaussian, one obtains

$$A = \tfrac{1}{2} \operatorname{erfc}[z_0^*(y^*)/2^{\frac{1}{2}}] .$$

Figure 3 shows A as a function of y^*.

The normal pressure is not proportional to A, due to the interaction between adjacent plastic microcontacts. On the basis of theory and experiment, it is shown by Pullen and Williamson that a good lower bound for the apparent normal pressure is obtained by setting

$$p/H \approx A/(1-A) .$$



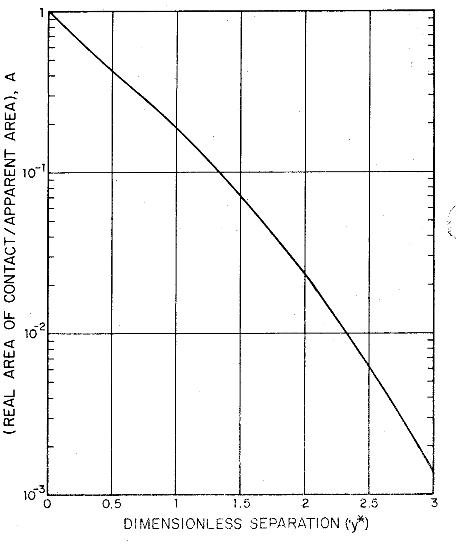


Fig. 2. The dimensionless effective separation z_0^* as a function of the apparent dimensionless separation y^*.

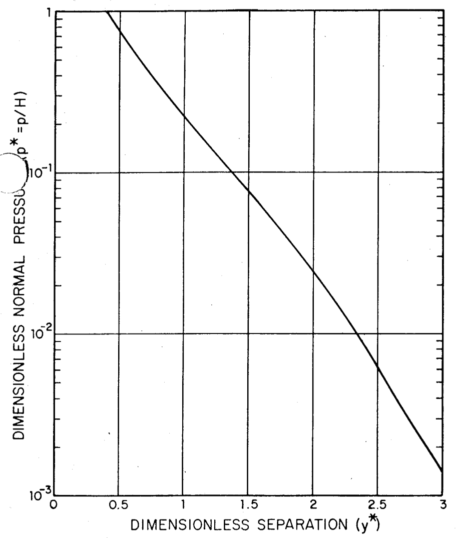
Fig. 3. The fractional real area of contact as a function of the dimensionless separation.

By combining this equation with Fig. 3, a relation between $p^* \equiv p/H$ and the dimensionless separation y^* may be obtained, and is shown in Fig. 4.

Finally, the relation between the effective separation z_0^* and the dimensionless pressure p^* is shown to be

$$\operatorname{erfc}(z_0^*/2^{\frac{1}{2}}) = 2p^*/(1+p^*)\,.$$

This relation is shown in Fig. 5.



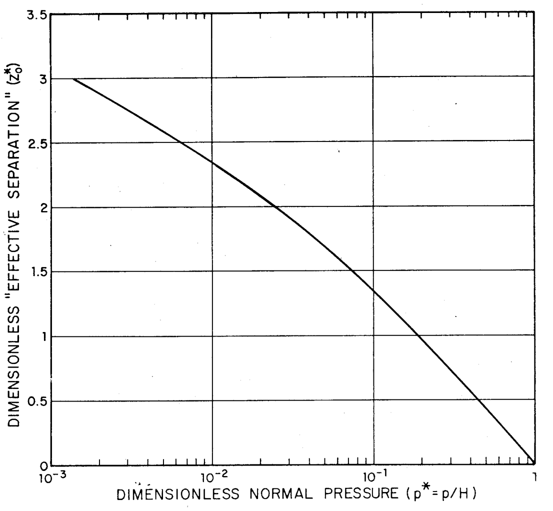


Fig. 4. Relation between the apparent dimensionless separation and the pressure to hardness ratio.

Fig. 5. Relation between the effective dimensionless separation and the pressure to hardness ratio.

Using Fig. 5, the effective separation may be determined once the normal pressure p and the metal hardness H are known.

The preceding model is central to the theory to be developed since it establishes that one may (at least roughly) consider the contact of a plastic rough surface and a hard flat to be statistically equivalent to the intersection of the rough surface and an imaginary plane. This implies that one need not consider the surface to be distorting during contact; distortion would be virtually impossible to take into account analytically. On the other hand, an analysis of the intersection of a rough surface and an imaginary plane is tractable under certain conditions, as will be shown in Section 3.

2.2. Two rough surfaces

Let z_1 and z_2 be the heights of the two surfaces. If y be the separation of their mean planes, the two surfaces, considered imaginary for the moment, intersect whenever $z_c \equiv (z_1+z_2) \geqslant y$. The quantity z_c defines a "composite" rough surface such that the contact of the two rough surfaces is equivalent to the contact of the composite surface and a hard flat at a separation y.

When z_1 and z_2 are Gaussian, z_c will also be Gaussian. Furthermore, when z_1 and z_2 are uncorrelated, the Power Spectral Density (PSD) of z_c will be the sum of the PSD's of z_1 and z_2. It is then a straightforward conclusion[1] that $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Thus, the dimensionless separation for the composite surface is $y^* = y/\sigma = y/(\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}$ and once this is obtained, the entire analysis of the preceding section is directly applicable to the present problem, and it follows that the contact of two rough surfaces is also equivalent to the intersection of a composite rough surface and an imaginary plane.

If z_1 and z_2 be isotropic, *i.e.*, if their PSD's have circular symmetry, then z_c will also be isotropic. Furthermore, designating the r.m.s. slope and r.m.s. second derivative of an arbitrary profile of an isotropic surface by σ' and σ'', it also follows[1] that

$$(\sigma_c')^2 = (\sigma_1')^2 + (\sigma_2')^2$$

and

$$(\sigma_c'')^2 = (\sigma_1'')^2 + (\sigma_2'')^2 .$$

With this groundwork, a detailed analysis of the intersection of an isotropic, Gaussian rough surface and an imaginary plane may be made.

3. INTERSECTION OF A ROUGH SURFACE AND A PLANE

3.1. Some known statistics of Gaussian rough surfaces

Let the rough surface height be $z(x_1, x_2)$, where (x_1, x_2) are Cartesian coordinates in the mean plane. We first define three parameters: σ, the r.m.s. height of a cross-sectional profile of the surface, σ', the r.m.s. slope of the profile, and σ'', its r.m.s. second derivative. (In the notation of ref. 1, we have $\sigma^2 = m_0$, $(\sigma')^2 = m_2$, $(\sigma'')^2 = m_4$.) Furthermore, define a dimensionless surface height by

$$z^* = z/\sigma , \tag{1}$$

and a parameter α by

$$\alpha = \frac{(\sigma\sigma'')^2}{(\sigma')^4} . \tag{2}$$

The parameter α is a measure of the breadth of the Power Spectral Density (PSD) of the surface roughness, *i.e.*, the range of wavelengths encountered in it. Large values of α indicate a broad PSD; small values, a narrow PSD. For an isotropic, Gaussian surface[2], $\alpha \geqslant 1.5$. Then, the following results hold for an isotropic, homogeneous, Gaussian surface.

(1) The probability density for the surface height is given by[1]

$$p(z^*) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}z^{*2}) . \tag{3}$$

(2) The probability density for heights of maxima* (*i.e.*, local maxima) is given by[1]

---

* Equation (4) corrects a typographical error appearing in the first term in the braces on the right-hand-side of eqn. (50) in ref. 1.

$$p_{\max}(z^*) = \frac{3^{\frac{1}{2}}}{2\pi}\left\{\left[\frac{3(2\alpha-3)}{\alpha^2}\right]^{\frac{1}{2}} z^* \exp(-C_1 z^*) + \frac{3(2\pi)^{\frac{1}{2}}}{2\alpha}(1+\operatorname{erf}\beta)(z^{*2}-1)\exp(-\tfrac{1}{2}z^{*2}) + (2\pi)^{\frac{1}{2}}\left[\frac{\alpha}{3(\alpha-1)}\right]^{\frac{1}{2}}(1+\operatorname{erf}\gamma)\exp[-\alpha z^{*2}/2(\alpha-1)]\right\}, \quad (4)$$

where

$$C_1 = \alpha/(2\alpha-3),$$

$$\beta = \left[\frac{3}{2(2\alpha-3)}\right]^{\frac{1}{2}} z^*$$

and

$$\gamma = \left[\frac{\alpha}{2(\alpha-1)(2\alpha-3)}\right]^{\frac{1}{2}} z^*. \quad (5)$$

(3) The density of all maxima (#/unit area) is given by[1,2]

$$D_{\max} = \frac{1}{6\pi\cdot 3^{\frac{1}{2}}}\left(\frac{\sigma''}{\sigma'}\right)^2. \quad (6)$$

(4) The fraction of mean-plane area corresponding to surface points above a height z_0^* is given by

$$A(z_0^*) = \int_{z_0^*}^{\infty} p(z^*)\mathrm{d}z^* = \tfrac{1}{2}\operatorname{erfc}(z_0^*/2^{\frac{1}{2}}). \quad (7)$$

Equation (3) has been used in deriving the final expression.

(5) The volume of the solid lying above a height z_0^* (per unit mean-plane area) is given by

$$V(z_0^*) = \sigma\int_{z_0^*}^{\infty} A(z^*)\mathrm{d}z^*. \quad (8)$$

3.2. Statistics of the intersection

We now proceed to derive various results regarding the density of finite contact patches that the surface has in a plane at height z_0^*, parallel to the mean plane. Each such patch is bounded by one closed contour if it is singly connected, and by many closed contours if it is multiply connected. To determine the density of finite contacts, we proceed as per Longuet-Higgins[2].

On a contour map of the surface, we may assign a direction ϕ to each point. The sign convention is that for any point, one stands on the contour through the point, facing uphill; then the local ϕ is the angle between the tangent vector pointing to the right, and any other fixed line, say the x_1-axis. Then[2] for any closed curve in the mean plane, the total change in ϕ in one complete circuit is given by

$$\Delta\phi = 2\pi\times\{\text{No. of maxima} + \text{No. of minima} - \text{No. of saddle points enclosed by the closed curve}\}. \quad (9)$$

By letting the curve grow in size indefinitely, Longuet-Higgins shows that for a homogeneous surface, the following result must hold:

$$D_{max} + D_{min} - D_{sp} = 0\,, \tag{10}$$

where D_{max}, D_{min} and D_{sp} are the total densities of summits, minima, and saddle points, respectively.

Now if the surface is Gaussian with zero mean, it is statistically symmetrical about the mean plane. Thus

$$D_{max} = D_{min}\,. \tag{11}$$

Combining eqns. (10) and (11), we have

$$\tfrac{1}{2}D_{sp} = D_{max} = D_{min}\,. \tag{12}$$

Instead of considering arbitrary curves, we may consider closed contours bounding finite contact patches. If the contact patches are singly connected, as in Fig. 6(a), then clearly, for the contour bounding the patch, $\Delta\phi = 2\pi$. Since all maxima, minima and saddle points within such a patch also lie above z_0^*, eqn. (9) may be reduced to

$$\{\text{No. of maxima} + \text{No. of minima} - \text{No. of saddle points within one finite singly-connected path}\} = 1\,. \tag{13}$$

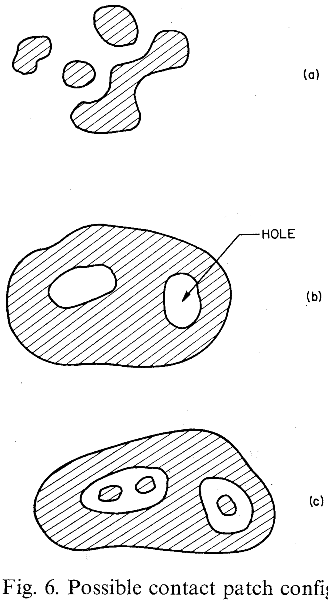


Fig. 6. Possible contact patch configurations.

Next, we consider a contact patch of the type shown in Fig. 6(b). It is possible to show that for such a patch,

$$\{\text{No. of (maxima} + \text{minima} - \text{saddle points) above } z_0^*\} + \{\text{No. of holes within the patch}\} = 1\,. \tag{14}$$

Suppose all the contact patches were of the type shown in Fig. 6(b),

(Fig. 6(a) being a particular case). Then, simply by counting patches and writing eqn. (14) for each patch, one arrives at the following result.

$$d_c(z_0^*) = d_{max}(z_0^*) + d_{min}(z_0^*) - d_{sp}(z_0^*) + d_H(z_0^*)\,, \tag{15}$$

where $d_c(z_0^*)$ is the density of finite contact patches at z_0^*, $d_H(z_0^*)$ is the density of holes at z_0^*, and d_{max}, d_{min} and d_{sp} are the densities of maxima, minima and saddle-points above z_0^*, respectively.

Finally, consider a contact patch such as that shown in Fig. 6(c). Equation (15) applies even in the presence of such contacts, if each of the little islands within the holes is counted as a separate contact patch. There is reasonable doubt that these islands can usefully be interpreted as separate contacts, but the following analysis leading to the density of contact patches is sufficiently approximate so that the preceding interpretation makes little quantitative difference. This is particularly true in view of the radically different qualitative view of the process of contact that emerges from the analysis, compared to existing models.

In sum, eqn. (15) will be used to estimate the density of finite contact patches at any separation z_0^*. To be specific, a finite contact patch is defined as a finite area of intersection including all points such that it is possible to travel from any point within it to any other point within it without crossing a contour line at z_0^*. These contact patches may be multiply connected, and they very likely will not be convex.

It is known[1] that surfaces with a broad-band PSD will have minima above the mean plane, just as they will have maxima below the mean plane. The presence of these minima ensures the occurrence of holes. At any height z_0^*, these holes will occur due to minima lying below z_0^* which are of sufficient depth so that the corresponding valleys intersect the z_0^* plane. If one knew something about the depths of these valleys, the following approximate analysis would be unnecessary; such information, however, would be contained only in higher-order autocorrelation functions or power spectra, and would not be easy to come by, both from the point of view of obtaining the empirical data, and of interpreting them. We therefore proceed with our approximate analysis.

For a surface that is symmetric about the mean plane, we know that the density of minima above z_0^* is equal to the density of maxima below $(-z_0^*)$. Thus

$$d_{max}(z_0^*) - d_{min}(z_0^*) = D_{max}\left\{\int_{-\infty}^{-z_0^*} p_{max}(z^*)\mathrm{d}z^* + \int_{z_0^*}^{\infty} p_{max}(z^*)\mathrm{d}z^*\right\}, \tag{16}$$

where $p_{max}(z^*)$ is as given in eqn. (4). Since the integral of p_{max} over $(-\infty, \infty)$ is unity, eqns. (12), (15) and (16) may be combined to obtain

$$d_c(z_0^*) = D_{max}\left\{1 - \int_{-z_0^*}^{z_0^*} p_{max}(z^*)\mathrm{d}z^* - 2\int_{z_0^*}^{\infty} p_{sp}(z^*)\mathrm{d}z^*\right\} + d_H(z_0^*)\,, \tag{17}$$

where p_{sp} is the probability density for heights of saddle points. This equation is still very general, the only assumptions being those of symmetry about the mean plane, and of the contact patches being finite. Thus, the preceding equation would not apply to either asymmetric surfaces or to two-dimensional surfaces, where the contacts are all of infinite length. It would apply quite generally to all other surfaces. We now proceed to study isotropic, Gaussian surfaces, with the anticipation that what is learned about them will apply qualitatively to other surfaces.

By using the techniques of random process theory[1], it is shown in Appendix I that for a Gaussian, isotropic, random surface,

$$p_{sp}(z^*) = \frac{1}{(2\pi)^{\frac{1}{2}}}\left(\frac{\alpha}{\alpha-1}\right)^{\frac{1}{2}} \exp[-\alpha z^{*2}/2(\alpha-1)]\,. \tag{18}$$

Introducing eqns. (4) and (18) into eqn. (17), we obtain, for isotropic, Gaussian surfaces,

$$d_c(z_0^*) = (2\pi)^{-\frac{3}{2}}(\sigma'/\sigma)^2\, z_0^* \exp(-\tfrac{1}{2}z_0^{*2}) + d_H(z_0^*)\,. \tag{19}$$

Since this result is central to the model being developed, discussion of it is in order. First, if it is known that holes do not occur, then the density of finite contact patches would be given by the first term on the right of eqn. (19). One would then conclude that the density depends only on σ' (r.m.s. profile slope) and σ (r.m.s. profile height) and that the density of finite contacts would be zero at $z_0^*=0$; contact would occur in one or more unbounded, singly connected patches. Intuitively one knows this to be unlikely for most surfaces, and it is concluded that the density of holes is not negligible. It is shown in the following analysis that the density of holes is negligible only when the surface has a narrow PSD, *i.e.*, when only a narrow range of wavelengths is present in the roughness. Such surfaces are not often found in practice.

If one assumes that $d_H(z_0^*)=0$, then eqn. (19) yields a lower bound for the density of contacts; not a particularly useful bound in the vicinity of $z_0^*=0$, but meant to illustrate the procedure to be used to estimate the density of holes. An upper bound for the density is obtained by assuming that each maximum above z_0^* gives rise to a separate contact:

$$d_c(z_0^*) \leqslant D_{max} \int_{z_0^*}^{\infty} p_{max}(z^*)\,dz^*\,. \tag{20}$$

Thus $d_c(z_0^*)$ is bounded as follows

$$\frac{1}{(2\pi)^{\frac{3}{2}}}\left(\frac{\sigma'}{\sigma}\right)^2 z_0^* \exp(-\tfrac{1}{2}z_0^{*2}) \leqslant d_c(z_0^*) \leqslant D_{max} \int_{z_0^*}^{\infty} p_{max}(z^*)\,dz^*\,. \tag{21}$$

These bounds are not particularly good and would be improved if one were to obtain bounds on the density of holes. In general, if one were to obtain lower and upper bounds $d_H^{LB}(z_0^*)$ and $d_H^{UB}(z_0^*)$, then $d_c(z_0^*)$ would be bounded as follows:

$$d_H^{LB}(z_0^*) \leqslant d_c(z_0^*) - (2\pi)^{-\frac{3}{2}}(\sigma'/\sigma)^2\, z_0^* \exp(-\tfrac{1}{2}z_0^{*2}) \leqslant d_H^{UB}(z_0^*)\,. \tag{22}$$

To obtain the bounds on d_H, it is necessary to assume that the surface roughness $z(x_1, x_2)$ can be written in the form

$$z = z_1 + z_2\,, \tag{23}$$

where z_1 has a PSD such that $\alpha_1 \approx 1.5$ (*i.e.*, the PSD is approximately a ring, with only a narrow range of wavelengths) and such that

$$\sigma_1^2 \gg \sigma_2^2\,, \tag{24}$$

where σ_1 and σ_2 are the r.m.s. values of z_1 and z_2. The partitioning indicated in eqn. (23) may be done in general when the surface PSD has a dominant annular ridge at some wavelength. This feature will often be present on surfaces resulting from some feature of the machining process. It is particularly important that the partitioning be done on the basis of the surface PSD and not the profile PSD, since it has been shown[8] that the two PSD's give rise to radically different ideas of the spectral content of the roughness.

If the partitioning additionally be done in such a way that z_1 and z_2 are uncorrelated[9], then the following results obtain:

$$\left.\begin{aligned} \sigma^2 &= \sigma_1^2+\sigma_2^2 \approx \sigma_1^2 \\ \sigma'^2 &= \sigma_1'^2+\sigma_2'^2 \\ \sigma''^2 &= \sigma_1''^2+\sigma_2''^2 \end{aligned}\right\}, \tag{25}$$

where the subscripts 1 and 2 refer to the z_1 and z_2 components. Since $\sigma_1 \approx \sigma$, we also have $z_0^* \equiv z_0/\sigma \approx z_0/\sigma_1 \equiv z_1^*$.

Now consider the intersection of the dominant z_1 component and a plane at z_0^*. Since $\alpha_1 \approx 1.5$, there are relatively few minima above $z_0^*=0$ [as can be seen from eqn. (4), and from the fact that $p_{\min}(z^*)=p_{\max}(-z_0^*)$]. Thus, one would expect relatively few holes to appear in the contact patches. The holes are caused by the superposed z_2 roughness and its valleys and the density of holes will depend mainly on the ratio σ_2/σ_1. If σ_2 is very small compared to σ_1, the perturbation caused by z_2 will be negligible. Some holes will nevertheless appear around the contours of the z_1-contact patches. To put this argument on a more quantitative level, consider the interference between the z_0^* plane and the z_1 roughness, given by

$$e = z_1(x_1, x_2) - z_0 \,. \tag{26}$$

When the z_2 roughness is superposed, some of its valleys around the point (x_1, x_2) will be deeper than e, and will cause holes to appear in the contact patch in which the point (x_1, x_2) lies. We now obtain bounds on the density of these holes.

3.2.1. Upper bound on the density of holes

When the z_2 roughness is superimposed on the z_1 roughness, local minima appear. If the gradient of z_1 is ∇z_1, these minima are caused by those points of z_2 where $\nabla z_2 = -\nabla z_1$. These minima will cause holes if $z_1 \geqslant z_0$ and if $z_2 \leqslant z_0 - z_1$, although it is not clear how many holes will be caused, since each hole may contain more than one local minimum. However, an upper bound on the density of holes is obtained by assuming that each local minimum caused by z_2 gives rise to a hole if it is deep enough.

Using the techniques of ref. 1, it may be shown that the density of points on z_2 satisfying the requirements for a minimum except that

$$\frac{\partial z_2}{\partial x_1} = -\frac{\partial z_1}{\partial x_1}, \quad \frac{\partial z_2}{\partial x_2} = -\frac{\partial z_1}{\partial x_2}$$

is given by

$$D_\zeta^2 = D_{\min}^2 \exp\left[-\frac{1}{2\sigma_2'^2}(\xi_2^2+\xi_3^2)\right] \tag{27}$$

where D^2_{min} is the density of minima of z_2, given by

$$D^2_{min} = \frac{1}{6\pi \cdot 3^{\frac{1}{2}}} \left(\frac{\sigma''_2}{\sigma'_2}\right)^2 , \tag{28}$$

and ξ_2, ξ_3 are defined by

$$\xi_2 = \frac{\partial z_1}{\partial x_1}, \quad \xi_3 = \frac{\partial z_1}{\partial x_2}. \tag{29}$$

Furthermore, the probability density for the heights of these points is simply

$$p^2_\zeta(z_2) = p^2_{min}(z_2) , \tag{30}$$

where $p^2_{min}(z_2)$ is the probability density for the heights of true minima of z, given by

$$p^2_{min}(z_2) = p^2_{max}(-z_2) , \tag{31}$$

p^2_{max} being given by eqn. (4).

In eqns. (27) and (30), the subscript ζ signifies points with a given gradient $\zeta = (\xi_2^2 + \xi_3^2)^{\frac{1}{2}}$.

The fraction of the mean plane area where z_1, ξ_2 and ξ_3 lie in the ranges $(z_1, z_1 + dz_1)$, $(\xi_2, \xi_2 + d\xi_2)$, $(\xi_3, \xi_3 + d\xi_3)$ respectively is given by[1]

$$dA = p(z_1)p(\xi_2)p(\xi_3)dz_1 d\xi_2 d\xi_3 . \tag{32}$$

An upper bound on the number of holes in dA is thus given by

$$d^{UB}_H = D^2_\zeta dA = D^2_s p(z_1)p(\xi_2)p(\xi_3)dz_1 d\xi_2 d\xi_3 \int_{z_2=z_0-z_1}^{-\infty} p^2_\zeta(z_2)dz_2 , \tag{33}$$

and an upper bound on the total number of holes is obtained by integrating this expression, letting z_1 vary over (z_0, ∞), and ξ_2, ξ_3 over $(-\infty, +\infty)$:

$$d^{UB}_H(z_0) = \int_{z_1=z_0}^{\infty} \iint_{\xi_2,\xi_3=-\infty}^{\infty} D^2_\zeta p(z_1)p(\xi_2)p(\xi_3)dz_1 d\xi_2 d\xi_3 \int_{z=z_0-z_1}^{-\infty} p^2_\zeta(z_2)dz_2 \tag{34}$$

Using the following expressions[1] for $p(\xi_2)p(\xi_3)$

$$p(\xi_2)p(\xi_3) = \frac{1}{2\pi\sigma'_1} \exp\left[-\frac{1}{2\sigma_1'^2}(\xi_2^2 + \xi_3^2)\right] ,$$

and introducing eqns. (27), (30) and (31) into Eqn. (33), one obtains

$$d^{UB}_H(z_0^*) = D^2 \min(\sigma'_2/\sigma')^2 \int_{z_0^*}^{\infty} p(z_1^*)dz_1^* \int_{\sigma_1(z_1^*-z_0^*)/\sigma_2}^{\infty} p^2_{max}(z_2^*)dz_2^* , \tag{35}$$

where

$$z_1^* = z_1/\sigma_1, \quad z_0^* = z_0/\sigma_1, \quad z_2^* = z_2/\sigma_2 . \tag{36}$$

Note that $p(z_1^*)$ is Gaussian with zero mean and unit standard deviation, whereas D^2_{min} and p^2_{max} will depend on σ_2, σ'_2 and σ''_2, in the manner indicated in eqns. (4) and (6). Thus, eqn. (28) gives an upper bound on the density of holes that may be completely determined from the surface PSD.

3.2.2. Lower bound on the density of holes

A lower bound on the number of holes Δd_H within an area dA is obtained by assuming all the holes to be singly connected. This argument is similar to the one used in obtaining the lower bound in eqn. (21). However, as in eqn. (27), the densities of all stationary points of z_2 are multiplied by the factor $\exp[-(\xi_2^2+\xi_3^2)/2\sigma_2'^2]$ due to the local gradient caused by the z_1-roughness. The resulting expression is

$$\Delta d_H^{LB} = \frac{1}{(2\pi)^{\frac{3}{2}}} (\sigma_2'/\sigma_2)^2 p(z_1) p(\xi_2) p(\xi_3) \left(\frac{z_1 - z_0}{\sigma_2}\right)$$
$$\times \exp\left\{ - \frac{1}{2\sigma_2'^2} (\xi_2^2 + \xi_3^2) - \tfrac{1}{2}\left(\frac{z_1 - z_0}{\sigma_2}\right)^2 \right\} dz_1\, d\xi_2\, d\xi_3 \,.$$

Integrating over z_1, ξ_2 and ξ_3, one obtains

$$d_H^{LB}(z_0^*) = \frac{1}{(2\pi)^2} \left(\frac{\sigma_2'}{\sigma_2}\right)^2 \left(\frac{\sigma_2'}{\sigma'}\right)^2 \left(\frac{\sigma_1}{\sigma_2}\right) \int_{z_0^*}^{\infty} (z_1^* - z_0^*)$$
$$\times \exp\left\{ -\tfrac{1}{2} \left[z_1^{*2} + \left(\frac{\sigma_1}{\sigma_2}\right)^2 (z_1^{*2} - z_0^{*2}) \right] \right\} dz_1^* \,.$$

If, as has been assumed, $\sigma_1 \gg \sigma_2$, and if in addition, $\sigma_2 z_0^* \ll \sigma_1$, this reduces to

$$d_H^{LB}(z_0^*) \approx \frac{1}{(2\pi)^2} \left(\frac{\sigma_2'}{\sigma_2}\right)^2 \left(\frac{\sigma_2'}{\sigma'}\right)^2 \left(\frac{\sigma_2}{\sigma_1}\right) \exp\left(-\tfrac{1}{2} z_0^{*2}\right). \qquad (37)$$

This completes the determination of the bounds on $d_H(z_0^*)$. In summary, the density of finite contact patches is bounded as in eqn. (22), d_H^{UB} being given by eqn. (35) and d_H^{LB} by eqn. (37).

Before proceeding to examine the value of the bounds just obtained, we consider the implications of the model of the process of contact developed so far.

3.3. Physical interpretation of the surface contact model

When the surface may be partitioned, we know that the density of finite contact patches is given by eqn. (19); intuitively we also know that most of the holes are caused by the z_2-roughness. An alternative approach to the one used is to argue that singly-connected patches are produced by the z_1-roughness. The z_2-roughness causes perforations to appear in these patches, but this by itself does not cause the number of contact patches to change. In addition, however, the z_2-roughness causes numerous clusters of microcontacts to appear outside the large contact patches of the z_1-roughness. The density of these microcontacts added to the density of large contact patches must equal the density of finite contacts, given by eqn. (19). We now demonstrate by a heuristic argument that this is indeed approximately true for the model.

Assuming that the z_1-roughness causes few holes, the density of large contact patches (d_{LCP}) is found from eqn. (19) by substituting σ_1 for σ, σ_1' for σ', and by setting $d_H = 0$:

$$d_{LCP}(z_0^*) \approx (2\pi)^{-\frac{3}{2}} (\sigma_1'/\sigma)^2 z_0^* \exp\left(-\tfrac{1}{2} z_0^{*2}\right). \qquad (38)$$

Here, it has been assumed (as before) that $\sigma_1 \approx \sigma$.

Denote by d_M the density of small microcontacts caused by the z_2-roughness. Then in an area dA where the separation of the hard flat and the z_1-roughness is $(z_0 - z_1)$, eqn. 19 and the argument of Section 3.2.2. indicate that the following relation must hold:

$$\Delta(d_M - d_H) = \frac{1}{(2\pi)^{\frac{3}{2}}} (\sigma_2'/\sigma_2)^2 p(z_1) p(\xi_2) p(\xi_3) \left(\frac{z_0 - z_1}{\sigma_2}\right) \times \exp\left\{-\frac{1}{2\sigma_2'^2}(\xi_2^2 + \xi_3^2) - \frac{1}{2}\left(\frac{z_0 - z_1}{\sigma_2}\right)^2\right\} dz_1 d\xi_2 d\xi_3 .$$

Integrating over z_1^*, ξ_2 and ξ_3, it may be shown that when $\sigma_2 \ll \sigma_1$, the resulting expression is

$$d_M(z_0^*) - d_H(z_0^*) \approx (2\pi)^{-\frac{3}{2}} (\sigma_2'/\sigma)^2 (\sigma_2'/\sigma')^2 z_0^* \exp(-\tfrac{1}{2} z_0^{*2}) . \tag{39}$$

Finally, the density of finite contacts is

$$d_c(z_0^*) = d_{LCP}(z_0^*) + d_M(z_0^*) = (2\pi)^{-\frac{3}{2}} (\sigma_1'/\sigma)^2 z_0^* \exp(-\tfrac{1}{2} z_0^{*2}) + d_M(z_0^*) \tag{40}$$

which, upon using eqns. (39) and (25), reduces to eqn. (19) except for the term: $-(2\pi)^{-\frac{3}{2}} (\sigma_1'/\sigma)^2 (\sigma_2'/\sigma')^2 z_0^* \exp(-z_0^{*2}/2)$; this term, however, is never more than one-fourth of the first term in eqn. (19), and the error in the heuristic approach is small. (The error arises because of the assumption that eqn. (15) applies locally to the z_2-roughness superimposed on the z_1 roughness. Equation (15) does not hold exactly because the local roughness is inhomogeneous.) This constitutes a verification of the claim that large contact patches are produced by the z_1-roughness, small microcontacts and holes being produced by the z_2-roughness.

The following approximate picture of the process of contact results. Large contact patches, whose density is given by eqn. (38), appear. They are perforated by holes and surrounded by clusters of small microcontacts, the densities of these satisfying eqn. (39). In particular, the density of holes is bounded by d_H^{UB} and d_H^{LB}, given by eqns. (35) and (37). These holes and microcontacts appear in the main near the peripheries of the large contact patches. However, holes may also appear near the middle of a patch, if the patch appears on a z_1 peak of large radius, and the interference there is small.

If the area of the microcontact clusters resulting from the z_2 roughness is small compared to the area of the main contact patches, one would conclude that the density of finite contacts given by eqn. (19) would be misleading to a certain extent. Now the overall real area of contact A is given by eqn. (7). On the other hand, the area of the holes caused by the z_2-roughness (or the microcontacts) is approximately given by

$$A_2 = \int_{z_0^*}^{\infty} p(z_1^*) dz_1^* \int_{-\infty}^{-(\sigma_1/\sigma_2)(z_1^* - z_0^*)} p(z_2^*) dz_2^* . \tag{41}$$

Since both $p(z_1^*)$ and $p(z_2^*)$ are Gaussian,

$$A_2 \approx \frac{(\sigma_2/\sigma_1)}{4\pi} \exp(-\tfrac{1}{2} z_0^{*2}) . \tag{42}$$

Finally, the area of the large contact patches (excluding the holes) is approximately

$$A_{\mathrm{LCP}} = A - A_2 = A\,(1 - A_2/A)\,. \tag{43}$$

The ratio of A_2 to A is

$$\frac{A_2}{A} = \frac{2\sigma_2/\sigma_1}{\pi}\,\frac{\exp\left(-\frac{1}{2}z_0^{*2}\right)}{\mathrm{erfc}\,(z_0^*/2^{\frac{1}{2}})}\,. \tag{44}$$

When $z_0^*=0$, $A_2/A=(\sigma_2/\pi\sigma_1)\ll 1$. When $z_0^*\to+\infty$, $A_2/A\to(2/\pi)^{\frac{1}{2}}\times(\sigma_2/\sigma_1 z_0^*)\ll 1$. When $z_0^*\to-\infty$, $A_2/A\to(\sigma_2/2\pi\sigma_1)\exp[-(\frac{1}{2})z_0^{*2}]\ll 1$. Thus, for a surface that may be partitioned, the hole area is always small compared to the area of the main contact patches.

It is also instructive to compare the mean areas of the large contact patches and of the z_2-microcontacts. Designating the mean area by $\bar{a}$, we have, for the large patches,

$$\bar{a}_1 = \frac{A_{\mathrm{LCP}}}{\text{density}} \approx \frac{(2\pi)^{\frac{3}{2}}}{2z_0^*}\left(\frac{\sigma}{\sigma_1'}\right)^2 \mathrm{erfc}\,(z_0^*/2^{\frac{1}{2}})\exp\left(\tfrac{1}{2}z_0^{*2}\right). \tag{45}$$

On the other hand, an upper bound on $\bar{a}_2$ is obtained by assuming the density of z_2-microcontacts (or holes) to be equal to $d_{\mathrm{H}}^{\mathrm{LB}}$. Thus, from eqns. (37) and (42),

$$\bar{a}_2 < \pi(\sigma_2/\sigma_2')^2(\sigma'/\sigma_2')^2\,. \tag{46}$$

Comparing $\bar{a}_1$ and $\bar{a}_2$, we find

$$\frac{\bar{a}_1}{\bar{a}_2} > \frac{(2\pi)^{\frac{1}{2}}}{z_0^*}\left(\frac{\sigma}{\sigma_1'}\right)^2\left(\frac{\sigma_2'}{\sigma_2}\right)^2\left(\frac{\sigma_2'}{\sigma'}\right)^2 \mathrm{erfc}\,(z_0^*/2^{\frac{1}{2}})\exp\left(\tfrac{1}{2}z_0^{*2}\right). \tag{47}$$

As $z_0^*\to 0$, it may be seen that $\bar{a}_1 \gg \bar{a}_2$. This is in part due to the fact that the density of large patches$\to 0$, and their mean area$\to\infty$. As $z_0^*\to\infty$, eqn. (47) results in

$$\frac{\bar{a}_1}{\bar{a}_2} > \frac{2}{z_0^{*2}}\left(\frac{\sigma}{\sigma_2}\right)^2\left(\frac{\sigma_2'}{\sigma_1'}\right)^2\left(\frac{\sigma_2'}{\sigma'}\right)^2. \tag{48}$$

It appears, therefore, that at large separations, it is not possible to conclude in general that the holes or microcontacts due to the z_2-roughness are of small size compared to the z_1-patches. However, when the surface roughness is such that $\sigma_2'=\mathcal{O}(\sigma_1')$, then $\bar{a}_1 \gg \bar{a}_2$.

One might conclude on the basis of the preceding analysis that for surfaces that may be partitioned, the holes or microcontacts resulting from the z_2-roughness are unimportant from the viewpoint of either load-bearing or electrical/thermal contact conductance problems. However, they may be significant in problems involving the frequency of asperity collisions (wear, fatigue) or lubrication, the holes providing traps for the lubricant. This latter possibility was communicated by Dr. Tibor Tallian of SKF Industries as a possible explanation for the static load-bearing capacity of lubricant films. Further insight into the lubrication problem at small separations (high pressures) is gained by noting that even though the area of the holes is small, the contact patches tend to be very elongated and of complex shape, making it difficult for the lubricant trapped in the open spaces to flow out.

Discussion of partitionable surfaces yields some insight into the contact of surfaces that may not be so partitioned. We may still write $z=z_1+z_2$ with $\alpha_1 \approx 1.5$, but it may no longer be assumed that $\sigma_2 \ll \sigma_1$ (or that $\sigma_1 \approx \sigma$). One may to a certain extent consider surfaces that may be partitioned, and then let the ratio (σ_2/σ_1) increase. Clearly, the area of the holes and of the z_2 microcontacts, as well as their density, will increase. Moreover, the holes will begin to have islands of contact within them, and so on. Eventually, as one approaches a truly broadband surface, it will no longer be permissible to think in terms of isolated, singly-connected microcontacts, even at large separations. The real area of contact will look like pieces of Swiss cheese. One may put these ideas in terms of level-crossings of a profile of the surface. For large z_0^*, the length of the dwells below z_0^* has a large mean value[2]. It is usually thought that these dwells correspond to areas of no contact on the surface between contact patches. This interpretation is correct for narrow-band or partitionable surfaces. However, for surfaces with a flat PSD, these dwells may equally well correspond to spaces of no contact within contact patches. It is then no longer permissible to think in terms of unit events such as asperity contacts and collisions, and a new model of contact needs to be developed. We do not, in this paper, attempt to develop such a model. Instead, we confine attention in the remainder of this paper to narrow-band and partitionable surfaces.

First, an analysis of the bounds on the density of holes obtained in Sections 2.2.1. and 2.2.2.

3.4. Analysis of the bounds

We now exclusively consider surfaces that may be partitioned, narrow-band surfaces obviously falling into this category. In order to make the following discussion of the bounds on the density of microcontacts explicit, we consider two limiting cases: (a) when z_2 is broad-band (*i.e.*, $\alpha_2 \gg 1.5$), and (b) when z_2 is also narrow-band (*i.e.*, $\alpha_2 \approx 1.5$).

(a) z_2 is broad-band, i.e., $\alpha_2 \gg 1.5$

In this case, $p^2_{\max}(z_2^*)$ in eqn. (35) is found from eqn. (4) to be approximately Gaussian:

$$p^2_{\max}(z_2^*) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}z_2^{*2}). \tag{49}$$

Introducing this into eqn. (35), and again requiring $\sigma_2 z_0^* \ll \sigma_1$, the following bound is obtained:

$$d_{\mathrm{H}}(z_0^*) \leqslant d_{\mathrm{H}}^{\mathrm{UB}}(z_0^*) \approx \frac{D^2_{\max}}{2\pi}\left(\frac{\sigma_2}{\sigma_1}\right)\left(\frac{\sigma_2'}{\sigma'}\right)^2 \exp\left(-\tfrac{1}{2}z_0^{*2}\right). \tag{50}$$

Now $D^2_{\max}$ is found from Eqn. (6) to be

$$D^2_{\max} = \frac{1}{6\pi\cdot 3^{\frac{1}{2}}}\left(\frac{\sigma_2''}{\sigma_2'}\right)^2. \tag{51}$$

Combining eqns. (22), (49) and (50), we obtain the following upper bound on the density of finite contacts:

$$d_{\mathrm{c}}(z_0^*) \leqslant d_{\mathrm{c}}^{\mathrm{UB}}(z_0^*) = (2\pi)^{-\frac{3}{2}}(\sigma'/\sigma)^2 z_0^* \exp(-\tfrac{1}{2}z_0^{*2}) + (12\times 3^{\frac{1}{2}}\pi^2)^{-1}(\sigma_2''/\sigma')^2(\sigma_2/\sigma_1)\exp(-\tfrac{1}{2}z_0^{*2}). \tag{52}$$

The lower bound is obtained by combining eqns. (22) and (37):

$$d_c(z_0^*) \geqslant d_c^{LB}(z_0^*) = (2\pi)^{-\frac{3}{2}}(\sigma'/\sigma)^2 z_0^* \exp(-\tfrac{1}{2}z_0^{*2}) + (2\pi)^{-2}(\sigma_2'/\sigma_2)^2(\sigma_2/\sigma_1)(\sigma_2'/\sigma')^2 \exp(-\tfrac{1}{2}z_0^{*2}). \qquad (53)$$

To evaluate these bounds, we first compare the two terms within each bound, and then compare the two bounds. The ratio of the two terms in eqn. (52) is

$$R \equiv \frac{\text{second term}}{\text{first term}} = \frac{\alpha}{3(6\pi)^{\frac{1}{2}} z_0^*} \left(\frac{\sigma_2''}{\sigma''}\right)^2 \left(\frac{\sigma_2}{\sigma_1}\right). \qquad (54)$$

For small values of z_0^*, R is very large. This indicates the likelihood that the density of holes (second term) approximately equals the density of finite contact patches. Moreover, the analysis of Section 3.3, indicates that for small z_0^* the finite contact patches are almost all due to the z_2-roughness, the large contact patches due to the z_1-roughness no longer being bounded in size. For $z_0^* \to \infty$, $R \to 0$. To determine when the density of holes is negligible, we require $R \leqslant 0.1$, and obtain

$$z_0^* \geqslant 0.77\alpha(\sigma_2/\sigma_1)(\sigma_2''/\sigma'')^2. \qquad (55)$$

Above this value of z_0^*, the effect of the z_2-roughness is negligible; below it, the presence of the holes and z_2-microcontacts begins to be important.

The ratio of the two terms in eqn. (53) is

$$Q = \frac{\text{second term}}{\text{first term}} = \frac{1}{z_0^*(2\pi)^{\frac{1}{2}}} \left(\frac{\alpha}{\alpha_2}\right) \left(\frac{\sigma_2''}{\sigma''}\right)^2 \left(\frac{\sigma_2}{\sigma_1}\right). \qquad (56)$$

Again, according to the lower bound, the density of holes equals the density of finite contacts as $z_0^* \to 0$. As $z_0^* \to \infty$, the density of holes again becomes negligible. $Q \geqslant 0.1$ when

$$z_0^* \leqslant 4 \left(\frac{\sigma_2}{\sigma}\right) \left(\frac{\alpha}{\alpha_2}\right) \left(\frac{\sigma_2''}{\sigma''}\right)^2. \qquad (57)$$

Comparing eqns. (55) and (57), we may say with certainty for $z_0^* > 0.77\alpha (\sigma_2/\sigma)(\sigma_2''/\sigma'')^2$, the density of holes is less than 10% of the density of finite contacts; for $z_0^* < 4(\sigma_2/\sigma)(\alpha/\alpha_2)(\sigma_2''/\sigma'')^2$, the density of holes is more than 10%. There is a shadow zone between these two values of z_0^* where little is known. When $\alpha_2 \approx 5$, the two bounds coincide. The entire analysis is therefore not applicable for $\alpha_2 < 5$. (It was originally assumed that α_2 was sufficiently large as to give a Gaussian p_{max}^2 with zero mean, say $\alpha_2 > 20$.)

Finally, we compare the two bounds themselves. From eqns. (52) and (53),

$$S \equiv \frac{d_c^{UB}(z_0^*)}{d_c^{LB}(z_0^*)} = \frac{z_0^* + 0.077\alpha(\sigma_2''/\sigma'')^2(\sigma_2/\sigma_1)}{z_0^* + 0.4(\alpha/\alpha_2)(\sigma_2''/\sigma'')^2(\sigma_2/\sigma_1)}. \qquad (58)$$

First, we must have $S \geqslant 1$. This leads to $\alpha_2 \gtrsim 5$, as was previously noted. To determine when the bounds are close, we require $S < 1.1$. This condition certainly prevails if

$$z_0^* > 0.77\alpha(\sigma_2/\sigma_1)(1 - 5.7/\alpha_2)(\sigma_2''/\sigma'')^2. \qquad (59)$$

When α_2 is very large, this reduces to eqn. (55), which is the constraint for the density of holes to be certainly negligible, in which case the closeness of the bounds is not surprising. For a certain range of values of α_2, however, eqn. (59) indicates that the bounds will be close even though the density of holes is not negligible.

When $z_0^*=0$, eqn. (58) becomes

$$S = 0.192\alpha_2, \quad z_0^* = 0 .$$

If $\alpha_2 \gg 1.5$, the bounds at $z_0^*=0$ give little information, since they differ by a large factor. Some information may nevertheless be obtained by comparing the upper bound with the density of contact patches that would be predicted by an independent-asperity model. According to the latter, the density of contacts, denoted by d_c^{ASP}, would be

$$d_c^{ASP}(z_0^*=0) = D_{max} \int_0^{\infty} p_{max}(z^*)dz^* \geq \tfrac{1}{2} D_{max} , \tag{60}$$

where D_{max} is the density of maxima, given by eqn. (6). Combining eqns. (52) and (60), we obtain

$$\frac{d_c^{UB}(0)}{d_c^{ASP}(0)} \leq \frac{3\times3^{\frac{1}{2}}}{\pi\alpha_2} \left(\frac{\sigma_2}{\sigma}\right) \left(\frac{\sigma_2''}{\sigma''}\right)^2 \ll 1 . \tag{61}$$

Thus, the density of finite contacts is definitely smaller than the density predicted by the asperity model, and by quite a large factor.

(b) z_2 is narrow-band, i.e., $\alpha_2 \approx 1.5$

In this case, $p_{max}^2(z_2^*)$ may be shown from eqn. (4) to be

$$p_{max}^2(z_2^*) = \begin{cases} (6/\pi)^{\frac{1}{2}} \exp(-\frac{1}{2}z_2^{*2})[z_2^{*2}-1+\exp(-z_2^{*2})] , & z_2^* \geq 0 \\ 0 , \quad z_2^*<0 . \end{cases} \tag{62}$$

Proceeding as in the preceding case, the following approximate upper bound results:

$$d_c(z_0^*) \leq d_c^{UB}(z_0^*) = (2\pi)^{-\frac{3}{2}}(\sigma'/\sigma)^2 z_0^* \exp(-\tfrac{1}{2}z_0^{*2}) + (5/36\pi^2)(\sigma_2/\sigma)(\sigma_2''/\sigma')^2 \exp(-\tfrac{1}{2}z_0^{*2}) . \tag{63}$$

The lower bound remains the same as in eqn. (53).

The ratio of the terms in eqn. (63) is

$$R = \frac{\text{second term}}{\text{first term}} = \frac{5}{18z_0^*}\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \alpha \left(\frac{\sigma_2}{\sigma}\right) \left(\frac{\sigma_2''}{\sigma''}\right)^2 . \tag{64}$$

To determine when the density of holes is less than one-tenth of the density of large contact patches, we require $R<0.1$, and find that this occurs if

$$z_0^* \geq 2.77\alpha(\sigma_2/\sigma)(\sigma_2''/\sigma'')^2 . \tag{65}$$

From the lower bound analysis in case (a), we observe that the density of holes is greater than one-tenth of the density of large contact patches when z_0^* satisfies eqn. (57). The constraint on z_0^* becomes

$$z_0^* \leqslant 4(\alpha/\alpha_2)(\sigma_2/\sigma)(\sigma_2''/\sigma'')^2 . \tag{66}$$

The region of uncertainty between the two values of z_0^* given by eqns. (65) and (66) is now quite small, since $\alpha_2 \approx 1.5$.

Finally, the ratio of the upper and lower bounds in this case is

$$S = \frac{d_c^{UB}(z_0^*)}{d_c^{LB}(z_0^*)} = \frac{z_0^* + 0.277\alpha(\sigma_2/\sigma)(\sigma_2''/\sigma'')^2}{z_0^* + 0.4(\alpha/\alpha_2)(\sigma_2/\sigma)(\sigma_2''/\sigma'')^2} . \tag{67}$$

This ratio is always greater than unity, and the analysis is therefore applicable in all cases. The bounds are within 10% of each other when $S < 1.1$, leading to

$$z_0^* \geqslant 2.77\alpha(\sigma_2/\sigma)(\sigma_2''/\sigma'')^2(1 - 1.59/\alpha_2) . \tag{68}$$

The ratio has its maximum value at $z_0^* = 0$, where $S = 0.69\alpha_2$. Thus the bounds are fairly close for all z_0^*. For $\alpha_2 = 1.5$, the bounds are certainly within 10% of each other for $z_0^* \geqslant 0$.

(c) The general case

An examination of eqns. (55) and (65) indicates that the density of holes is likely to be less than 10% of the density of large contact patches (caused by the z_1-roughness) when

$$z_0^* > m\alpha(\sigma_2/\sigma)(\sigma_2''/\sigma'')^2$$

where m varies from 0.77 when z_2 is broad-band to 2.77 when z_2 is narrow-band. The density of holes is certainly greater than 10% of the density of large contact patches when

$$z_0^* \leqslant 4(\alpha/\alpha_2)(\sigma_2/\sigma)(\sigma_2''/\sigma'')^2 .$$

Finally, the two bounds will be within 10% of each other if

$$z_0^* \geqslant p\alpha(\sigma_2/\sigma_1)(\sigma_2''/\sigma'')^2(1 - q/\alpha_2) ,$$

where p and q vary from $p = 0.77$, $q = 5.7$ when $\alpha_2 \gg 1$ to $p = 2.79$, $q = 1.59$ when $\alpha_2 \approx 1.5$.

3.5. The shape of the large contact patches

Some idea of the shape of the large contact patches may be gained by the use of a dimensionless shape factor, which allows a comparison of the contact patches with ellipses (or circles). The reason for choosing ellipses for comparison is that the contact patch on an asperity can be assumed to be roughly elliptic for small values of interference. The purpose of the following analysis is to evaluate this assumption for large values of interference.

The shape factor for a homogeneous array of closed curves is defined here by

$$F = 4\pi A d / l^2 , \tag{69}$$

where A is the area enclosed (per unit mean plane area), d the density of closed curves (#/unit area) and l the length of the curves per unit area.

To determine the value of the shape factor for an array of ellipses in a reasonably simple manner, it is necessary to assume that they all have the same eccentricity e. Let a be the semimajor axis with probability density $p(a)$, mean m_a, and variance about the mean σ_a^2. Then it may be shown that

$$F_E(e) = \frac{\pi^2}{4} \frac{(1-e^2)^{\frac{1}{2}}}{[K(e)]^2} \left\{ \left(\frac{\sigma_a}{m_a} \right)^2 + 1 \right\}, \tag{70}$$

where $K(e)$ is the complete elliptic integral of the first kind.

A lower bound on F_E is obtained by setting $\sigma_a = 0$, *i.e.*, by taking all the ellipses to be of the same size:

$$F_E(e) \geqslant F_E^{LB}(e) = \frac{\pi^2}{4} \frac{(1-e^2)^{\frac{1}{2}}}{[K(e)]^2}. \tag{71}$$

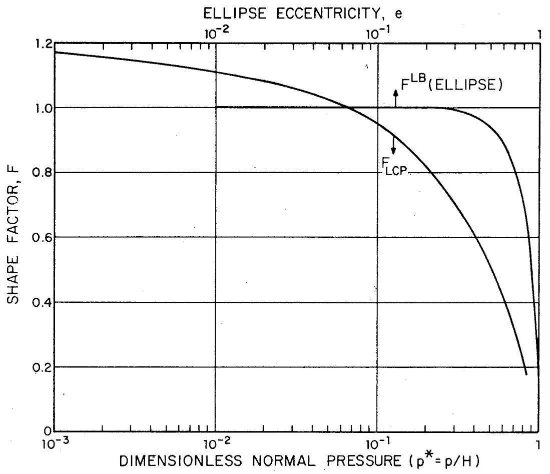


Fig. 7. Shape factor for an array of similar ellipses; upper bound on the shape factor of the large contact patches.

This lower bound is shown in Fig. 7. For $e = 0$, *i.e.*, for an array of identical circles, F_E^{LB} $(e = 0) = 1$. One may now evaluate the shape factor for the true contact patches and compare it with F_E^{LB}. If $F_{LCP} < 1$, it is immediately clear that not all the contact patches are circular. For a given value of eccentricity, $e = e_0$, if $F_{LCP} < F_E^{LB}(e_0)$, it is similarly clear that at least some of the large contact patches must be more elongated than ellipses with eccentricity e_0. Unfortunately, the shape factor does not contain too much information beyond this. It does not indicate, for example, the possible variations in the eccentricity e. Furthermore, it must be remembered that the preceding comparison is artificial to the extent that it assumes the contact patches to be roughly elliptic. They are likely to be quite irregular and non-convex, particularly at small separations, as shown by the gap maps of Williamson[10].

To determine F_{LCP}, one needs to know l_{LCP}, the length of the outer boundaries of the large contact patches. This may be approximately determined as follows. Corrsin[11] and Longuet-Higgins[2] have shown that the total length of the contours at z_0^* per unit mean-plane area for an isotropic Gaussian surface is

$$l = \tfrac{1}{2}(\sigma'/\sigma)\exp(-\tfrac{1}{2}z_0^{*2}). \tag{72}$$

If one assumes that the boundaries of the small microcontacts and holes do not intersect the boundaries of the contact patches caused by the z_1-roughness, then eqn. (72) may be applied to z_1 to obtain

$$l_{\mathrm{LCP}} \approx \tfrac{1}{2}(\sigma_1'/\sigma_1)\exp(-\tfrac{1}{2}z_0^{*2}). \tag{73}$$

Now the analysis of Section 3.3. indicates that the area of the large contact patches is

$$A_{\mathrm{LCP}} \approx \tfrac{1}{2}\operatorname{erfc}(z_0^*/2^{\frac{1}{2}}). \tag{74}$$

Combining eqns. (38), (69), (73) and (74), one obtains

$$F_{\mathrm{LCP}} = (8/\pi)^{\frac{1}{2}} z_0^* \operatorname{erfc}(z_0^*/2^{\frac{1}{2}})\exp(\tfrac{1}{2}z_0^{*2}). \tag{75}$$

The small microcontacts and holes and the large contact patches will intersect to a degree. This has two effects: first, some holes and microcontacts will vanish, implying that our estimates are too high; and second, the large contact patches will have sinuous boundaries, with a total length somewhat larger than that given by eqn. (14). Thus the shape factor in eqn. (75) is an upper bound on the true shape factor which is shown in Fig. 7. It may also be interpreted as the shape factor for smoothed large contact patches.

One can qualitatively assess the sinuousness of the large contact patches as follows. The total length of all contour lines is given by eqn. (72). The length of the boundaries of the small microcontacts and holes is $l_2 = l - l_{\mathrm{LCP}}$. According to the preceding argument, the expression in eqn. (73) is a lower bound on l_{LCP}. Thus, since $\sigma_1 \approx \sigma$,

$$l_2 \leqslant \tfrac{1}{2}[(\sigma' - \sigma_1')/\sigma]\exp(-\tfrac{1}{2}z_0^{*2}). \tag{76}$$

An alternative procedure for calculating l_2 is as follows: assume that l_1 and l_2 do not intersect. Upon applying eqn. (72) to the z_2-roughness over a small area dA where the separation is $(z_0 - z_1)$ and then integrating over z_1—a procedure similar to that used in Section 3.2.1. and 3.2.2.—we find

$$l_2 = \tfrac{1}{2}(\sigma_2'/\sigma)\exp(-\tfrac{1}{2}z_0^{*2}). \tag{77}$$

There is obviously a discrepancy between eqns. (77) and (76). Since

$$\frac{\sigma_2'}{\sigma' - \sigma_1'} = \frac{\sigma' + \sigma_1'}{\sigma_2'} > 1,$$

l_2 in eqn. (77) is always larger than the upper bound in eqn. (76). The greater the discrepancy, the more sinuous do we expect the boundaries of the large contact patches to be. The discrepancy is large when $\sigma_2' \ll \sigma_1'$, and small when $\sigma_2' \gg \sigma_1'$. The reason for this is that when $\sigma_2' \ll \sigma_1'$, holes (and small microcontacts) will tend to form only around the boundaries of the large contact patches, with a greater chance of intersection. When $\sigma_2' \gg \sigma_1'$, on the other hand, holes can form well within the large contact patches, with a decreased likelihood of intersection.

4. EXAMPLE

To illuminate the preceding somewhat abstract discussion, we now consider a simple example of an isotropic, Gaussian surface. This example does not necessarily have a counterpart in nature, and is meant merely as an illustration.

Let the profile spectrum on some arbitrary scale of height and wavenumber k be given by

$$\Phi_{zz}(k) = \tfrac{1}{2}|k| \exp(-\tfrac{1}{2}k^2) . \tag{78}$$

For this spectrum,

$$\sigma^2 = 1 , \quad \sigma'^2 = 1.25 , \quad \sigma''^2 = 3.75 , \quad \alpha = 2.4 .$$

One could presumably assume the density of holes to be negligible, since α is near 1.5. Instead, consider components z_1 and z_2 with spectra

$$\Phi_{z_1z_1}(k) = \begin{cases} \Phi_{zz}(k) , & |k| \leqslant k_0 \\ 0 , & |k| > k_0 \end{cases} \tag{79}$$

and

$$\Phi_{z_2z_2}(k) = \Phi_{zz}(k) - \Phi_{z_1z_1}(k) . \tag{80}$$

Straightforward computations show that

$$\begin{aligned} \sigma_1^2 &= 1 - E , \\ \sigma_1'^2 &= 1.25\,(1\text{-ERFC}) - k_0 E , \\ \sigma_1''^2 &= 3.75\,(1\text{-ERFC}) - k_0 E(3 + k_0^2) , \end{aligned} \tag{81}$$

where

$$E \equiv \exp(-\tfrac{1}{2}k_0^2) , \quad \text{ERFC} = \text{erfc}(k_0/2^{\frac{1}{2}}) . \tag{82}$$

Then

$$\begin{aligned} \sigma_2^2 &= E , \\ \sigma_2'^2 &= 1.25\text{ERFC} + k_0 E , \\ \sigma_2''^2 &= 3.75\text{ERFC} + k_0 E(3 + k_0^2) . \end{aligned} \tag{83}$$

We now require $\sigma_1 \approx 10\sigma_2$ in order to ensure a good partition. This leads to $k_0 \approx 3.3$; with this value, eqns. (84) and (86) become

$$\sigma_1^2 = 0.99 , \quad \sigma_1'^2 = 1.22 , \quad \sigma_1''^2 = 3.47 , \quad (\alpha_1 = 2.3)$$

and

$$\sigma_2^2 = 0.01 , \quad \sigma_2'^2 = 0.03 , \quad \sigma_2''^2 = 0.28 , \quad (\alpha_2 = 3.1) .$$

Although one could obtain the necessary bounds by using the correct expression for $p_{\max}^2$ with $\alpha_2 = 3.1$, we assume that z_2 is approximately narrow-band. The following results may then be obtained from the analysis of Section 3.

(1) The density of holes is less than 10% of the density of large contact patches if $z_0^* \geqslant 0.05$.

(2) The density of holes is more than 10% of the density of large contact patches if $z_0^* \leqslant 0.023$.

(3) If $z_0^* \geqslant 0.025$, the density of finite contacts is within 10% of the following value:

$$d_c(z_0^*) = 0.08 z_0^* \exp(-\tfrac{1}{2} z_0^{*2}) .$$

(4) The shape factor for the (smoothed) large contact patches is

$$F_{LCP} = 1.63 z_0^* \operatorname{erfc}(z_0^*/2^{\frac{1}{2}}) \exp(-\tfrac{1}{2} z_0^{*2}) .$$

(5) The mean area of the large contact patches is

$$\bar{a}_1 = 6.45 \operatorname{erfc}(z_0^*/2^{\frac{1}{2}}) \exp(-\tfrac{1}{2} z_0^{*2}/2^{\frac{1}{2}}) \leqslant 6.45 .$$

(6) The mean area of the small microcontacts (or holes) is bounded as follows:

$$16.3 \leqslant \bar{a}_2 \leqslant 41.6 .$$

This last result is surprising, since it indicates that though the holes and microcontacts are few in number, they are on the average quite large. This result is somewhat analogous to that for a sinusoidal surface; if the surface is stretched in the mean plane without a change in the vertical scale, the density of contacts decreases to zero while the average size of the contacts increases indefinitely. Thus, the present analysis may indicate that $\bar{a}_2 \to \infty$ as $z_2 \to 0$: this is because the limit of a perfectly smooth surface may be approached in many ways. If $\sigma_2'' \to 0$ faster than σ_2', the average contact size at a given separation will increase indefinitely.

In this example, the density of holes is probably underestimated somewhat at low values of z_0^*; this is because z_1 is not truly narrow-band. For $\alpha_1 = 2.3$, eqn. (4) indicates[1] that there will be a significant number of minima of z_1 up to $z_0^* \approx 0.5$. Thus the assumption that no holes are caused by z_1 is likely to lead to error for $z_0^* < 0.5$. If the partitioning be done in such a way that α_1 is closer to 1.5, (σ_2/σ_1) becomes larger. In this case, an examination of the integrals appearing in eqn. (35) and in the expression leading to eqn. (37) indicate that our expressions for both the upper and lower bounds (Section 3.4) will be too high.

Thus, if we let $k_0 = 2.4$ in eqn. (79), we find $\sigma_1^2 = 0.962$, $\sigma_1'^2 = 1.14$, $\sigma_1''^2 = 2.89$, $\alpha_1 \approx 2$; $\sigma_2^2 = 0.038$, $\sigma_2'^2 = 0.11$, $\sigma_2''^2 = 0.86$, $\alpha_2 \approx 2.7$, and $\sigma_1/\sigma_2 = 5$. In this case, the density of holes is predicted to be less than 10% of the density of large contact patches for $z_0^* \geqslant 0.03$, using the approximate expression for the upper bound in eqn. (63). The error in the predicted density of holes due to z_2 will tend to counteract the assumption that no holes are produced by z_1. The closeness of the constraint on z_0^* for the number of holes to be negligible obtained with two different partitions indicates that for the spectrum of eqn. (80), one may for usual loads neglect the occurrence of holes.

5. DISCUSSION

It has been shown that a fairly rigorous and detailed analysis of plastic contact may be made for surfaces satisfying the following requirements:

(1) The surface is Gaussian. This implies that the joint probability density of the height, the two first derivatives and the three second derivatives is Gaussian; it also implies that the surface is symmetric about its mean plane.

(2) The surface is isotropic.

(3) The surface PSD is such that the roughness may be partitioned into two components z_1 and z_2 such that z_1 is narrow-band and $\sigma_1^2 \gg \sigma_2^2$.

For such surfaces, useful results are derived in Section 3 for bounds on the densities, mean areas and shape factors of large finite contacts, small microcontacts and holes (craters) at a given separation or equivalently, at a given pressure/hardness ratio.

What of surfaces that do not satisfy the preceding requirements? If the surface is neither isotropic nor Gaussian it is possible, in principle, to make an analysis similar to the present one, if the height distributions and densities of stationary points are somehow obtained—implying a detailed experimental study of each surface. With such a study, however, it would be feasible to generate contour maps of the surface and analyze the process of contact numerically: the point of an analysis such as that presented in this paper being that it requires empirical inputs that are relatively easy to obtain.

If the surface is Gaussian but anisotropic, the present analysis can be duplicated. The statistics of stationary points can be obtained from the surface PSD, or alternatively, from a few profile PSD's. This has not been attempted as yet, but is certainly practical.

One may qualitatively assess the effects of anisotropy by comparing the statistics of the intersection of an isotropic (two-dimensional) Gaussian surface with an imaginary plane with those for a one-dimensional Gaussian surface. For the latter, the contact patches are all of infinite length. Their density is given by[2]

$$d_c'(z_0^*) = 2\pi^{-1}(\sigma'/\sigma)\exp(-\tfrac{1}{2}z_0^{*2}).$$

Denoting the density of finite contacts for the (partitionable) isotropic surface by $d_c^2(z_0^*)$, we have, for large z_0^*,

$$d_c^2(z_0^*) \approx (2\pi)^{-\frac{3}{2}}(\sigma'/\sigma)^2\, z_0^*\exp(-\tfrac{1}{2}z_0^{*2}).$$

Suppose we compare two surfaces with equal σ and σ'. Then

$$\frac{(d_c')^2}{d_c^2} = \frac{(2\pi)^{\frac{1}{2}}}{z_0^{*2}}\exp(-\tfrac{1}{2}z_0^{*2}).$$

Evidently, the one-dimensional surface has far fewer contact patches than the isotropic surface. Thus, one could argue that the effects of increasing anisotropy are (a) to elongate the contact patches, and (b) to cause contact patches to coalesce and decrease in number.

It is pointless to guess at the effects of the departure from Gaussian-ness: these could be extremely varied; at the very least, contact statistics may be expected to depend on parameters other than σ, σ' and σ''. Instead, we briefly consider the third requirement above, that of partitionability.

It might appear at first sight that this is a stringent requirement, since it requires the roughness to have a dominant narrow-band component. However, narrowness of the surface spectrum does not imply a profile spectrum that looks narrow band. For example, if the surface spectrum is a ring delta-function:

$$\Phi_s(k_1,k_2) = (2\pi)^{-1}\,\delta(k-k_0)\,, \qquad k \equiv (k_1^2+k_2^2)^{\frac{1}{2}}\,,$$

the corresponding profile spectrum is

$$\Phi_p(k_1) = \begin{cases} \pi^{-1}(k_0^2 - k_1^2)^{-\frac{1}{2}}, & |k_1| \leqslant k_0 \\ 0, & |k_1| > k_0 . \end{cases}$$

This is a spectum that has an (integrable) singularity at $k_1 = k_0$; more important, however, is the fact that the spectrum is fairly flat at low values of k_1. For example, $\Phi(k_0/2) \approx 1.16\,\Phi(0)$, $\Phi(3k_0/4) \approx 1.5\,\Phi(0)$. Thus a spectrum that is fairly flat at low wavenumbers (long wavelengths), has a peak at some wavenumber, and then decays rapidly may be considered to be narrow-band. An example of this kind of spectrum is

$$\Phi_p(k_1) = \tfrac{1}{2}|k| \exp(-k^2/2)$$

discussed in Section 4. For this spectrum, $\alpha = 2.4$. A maximum occurs at $k = 1$, where $\Phi_p = 0.303$. The half-power bandwidth is $\Delta k \approx 1.6$, the half-power points being $k \approx 0.3$, 1.9.

Alternatively, a band-limited spectrum of the form

$$\Phi_p(k_1) = \begin{cases} 1, & |k| \leqslant k_0 \\ 0, & |k| > k_0 \end{cases}$$

has $\alpha = 1.8$. This again represents an approximately narrow-band surface.

Finally, consider a profile spectrum of the form

$$\Phi_p(k_1) = (1 + k_1^{2n})^{-1}, \qquad n\text{: integer} .$$

From the definitions of σ^2, σ'^2, σ''^2, it is possible to show that

$$\sigma^2 = \frac{(\pi/n)}{\sin(\pi/2n)}, \quad \sigma'^2 = \frac{(\pi/n)}{\sin(3\pi/2n)}, \quad \sigma''^2 = \frac{(\pi/n)}{\sin(5\pi/2n)} .$$

These hold when $n > 0$, $n > 1$ and $n > 2$ respectively. We assume $n \geqslant 3$.

From the definition of α, we find

$$\alpha = \frac{1 - \cos(3\pi/n)}{\cos(2\pi/n) - \cos(3\pi/n)} .$$

The following values of α are obtained:

n	3	4	5	6
α	4.00	2.41	2.12	2

Furthermore, as $n \to \infty$, $\alpha \to 1.8$, decreasing monotonically from $n = 3$. A profile with this spectrum represents an approximately narrow-band surface, if $n \geqslant 4$.

Thus, if a profile spectrum can be approximated in a band that contributes significantly to the mean-square level by one of the above spectra (or by a variety of others), it may be partitioned.

On the other hand, if the surface spectrum is clearly not partitionable in the necessary manner, the argument of Section 3.3 indicates that conventional descriptions of contact—involving encounters between clearly defined asperities—may not be applicable. In the face of a considerable lack of experimental data on what such surfaces look like, it appears fruitless to speculate on what new descriptions and concepts might emerge.

In conclusion comments on some other assumptions underlying the theory and on comparisons with experimental data are presented.

A major assumption in the theory is that the entire contacting area is plastically deformed and moreover, that this plastic deformation occurs in the manner postulated by Pullen and Williamson[6] and summarized in Section 2. The important ingredient of the Pullen/Williamson model is that the statistics of contact are the same as those for the intersection of the undeformed rough surface and an imaginary plane. Of secondary importance is their observation that the pressure/hardness ratio p^* is related to the fractional area in contact A by $p^* = A/(1-A)$. Other relations could be used, since they are merely used to calculate the pressure at a given separation, and do not enter into the analysis of contact statistics.

It is possible to develop a heuristic constraint ensuring that most of the surface does deform plastically. It has been shown in Section 3 that for partitionable surfaces most of the area in contact is due to the z_1-roughness and it is possible to demonstrate that at separations $z_0^* \geqslant 1$, the asperities are more-or-less independent (*i.e.*, nonintersecting). Suppose one assumes that this is so; then the density of large contact patches must equal the density of summits lying above z_0^*:

$$d_{\mathrm{LCP}}(z_0^*) = \int_{z_0^*}^{\infty} p'_{\max}(z_0^*)\,\mathrm{d}z^* .$$

Upon introducing eqn. (64) into this equation, we obtain

$$d_{\mathrm{LCP}}(z_0^*) \approx (2\pi)^{-\frac{3}{2}}(\sigma_1'/\sigma)\, z_0^* \exp(-\tfrac{1}{2}z_0^{*2}) \times \{1+(\pi/6)^{\frac{1}{2}} \exp(\tfrac{1}{2}z_0^{*2})\, \mathrm{erfc}\,(z_0^* \tfrac{3^{\frac{1}{2}}}{2}\} ,$$

a result which differs from eqn. (38) by less than about 10% if $z_0^* \geqslant 1$, thereby proving the hypothesis of independent asperities.

Now consider an asperity with a summit height z^*. It will deform plastically if[4]

$$z^* - z_0^* \geqslant (R/\sigma)(H/E') , \tag{84}$$

where R is the inverse of the mean curvature at the summit, H is the hardness of the asperity, and E' is defined by

$$\frac{1}{E'} = \frac{1-\nu^2}{E_1} + \frac{1-\nu_2^2}{E_2} ,$$

E_1, E_2 and ν_1, ν_2 being the Young's moduli and Poisson's ratios for the two surfaces.

It is further possible to show (by an examination of the joint-probability density of summit height and mean curvature[1]) that when $\alpha_1 = 1.5$, R is given by

$$R = (\tfrac{3}{2})^{\frac{1}{2}}(\sigma_2'' z^*) \tag{85}$$

for all asperities with height z^*; the distribution of curvature for summits of a given height being extremely narrow.

Upon combining eqns. (84) and (85) [and noting that $\sigma\sigma_1'' \approx \frac{3}{2}^{\frac{1}{2}}(\sigma_1')^2$], we find that for a given separation z_0^*, those asperities will deform plastically whose height z^* satisfies the following inequality:

$$z^* \geqslant \tfrac{1}{2}[z_0^* + \{z_0^{*2} + 4(H/E'\sigma')^2\}^{\frac{1}{2}}] . \tag{86}$$

It follows that almost all asperities in contact will deform plastically if $2(H/E'\sigma_1') \ll z_0^*$. If we require that

$$E'\sigma_1'/H \geqslant 5 , \tag{87}$$

then inequality (86) becomes

$$z^* \geqslant z_0^*(1 + 0.04/z_0^{*2}) \approx z_0^* \text{ if } z_0^* \geqslant 1 .$$

Thus if inequality (87) is satisfied, and if $z_0^* \geqslant 1$, then the assumption that most of the surface deforms plastically is justified. The constraint on σ_1', namely

$$\sigma_1' \geqslant 5(H/E') ,$$

does not involve the mean summit radius, as does Greenwood and Williamson's plasticity index[3,10], although both involve the ratio (H/E'). The criterion developed is reducible to that proposed by Halliday[12], but the criterion involves the r.m.s. profile slope of the z_1-roughness, and not of the entire z-roughness.

Comparisons of the theory with experimental data appear difficult since there appear to be no studies reporting both densities of contacts and profile spectra. Qualitatively, however, some similarities and some differences are found in a comparison with the data of Uppal, Probert and Thomas[13]. They report, for example, that except at the highest loads, most contacts are noncircular; they demonstrate the coalescence of contacts, as well as the presence of holes in the large contact patches. Further, they find that the density of contacts has a maximum at some separation. This is predicted by theory; for the results of Section 3 indicate the density of contacts probably satisfies a relation of the form

$$d_c(z_0^*) \propto \exp(-\tfrac{1}{2}z_0^*)(z_0^* + C) ,$$

where C is a constant that is usually small compared to one. It is then easy to demonstrate that $d_c(z_0^*)$ has a maximum at

$$z_0^* = \tfrac{1}{2}[(4 + c^2)^{\frac{1}{2}} - c] \leqslant 1 .$$

Here, a significant difference between theory and experiment is found, for the data of Uppal *et al.* suggest that the maximum occurs at a value of $z_0^* > 1$. No reasons for this discrepancy are known at present.

APPENDIX I

THE PROBABILITY DENSITY FOR HEIGHTS OF SADDLE POINTS

Let $z(x_1, x_2)$ be the height of the surface. Define

$$z^* = z/\sigma ,$$

$$[t_1, t_2, t_3] = \frac{3^{\frac{1}{2}}}{\sigma''}\left[\tfrac{1}{2}\left(\frac{\partial^2 z}{\partial x_1^2}+\frac{\partial^2 z}{\partial y_1^2}\right), \frac{\partial^2 z}{\partial x_1 \partial x_2}, \tfrac{1}{2}\left(\frac{\partial^2 z}{\partial x_1^2}-\frac{\partial^2 z}{\partial x_2^2}\right)\right]. \tag{A.1}$$

It is shown in ref. 1 that at stationary points, where $\partial z/\partial x_1 = \partial z/\partial x_2 = 0$, the joint-probability function for z^*, t_1, t_2 and t_3 is given by

$$P_{\text{sta}}(z^*, t_1, t_2, t_3) = \frac{C_1^{\frac{1}{2}}(\sigma''/\sigma')^2}{3(2\pi)^3}\exp(-C_1\xi^{*2})|t_1^2-t_2^2-t_3^2| \times \exp[-\tfrac{1}{2}(C_1 t_1^2+t_2^2+t_3^2+C_2 t_1 \xi^*)], \tag{A.2}$$

where

$$C_1 = \alpha/(2\alpha-3), \quad C_2 = C_1(12/\alpha)^{\frac{1}{2}}. \tag{A.3}$$

Saddle-points are defined by

$$\infty \leqslant t_1 \leqslant \infty, \quad t_2^2+t_3^2 \geqslant t_1^2. \tag{A.4}$$

Thus,

$$P_{\text{sp}}(z^*, t_1) = \frac{C_1^{\frac{1}{2}}}{3(2\pi)^3}\left(\frac{\sigma''}{\sigma'}\right)^2 \exp[-C_1\xi^{*2}-\tfrac{1}{2}(C_1 t_1^2+C_2 t_1\xi^*)] \times I, \tag{A.5}$$

where

$$I = \iint_{t_2^2+t_3^2 \geqslant t_1^2} |t_1^2-t_2^2-t_3^2| \exp.[-\tfrac{1}{2}(t_1^2+t_2^2)]\,\mathrm{d}t_2\,\mathrm{d}t_3.$$

The integral I is found to be equal to $4\pi\exp(-t_1^2/2)$, and by integrating eqn. (A.5) over t_1, we obtain [after using eqn. (A.3)]

$$P_{\text{sp}}(z^*) = \frac{2}{3\times 3^{\frac{1}{2}}}\left(\frac{\alpha}{\alpha-1}\right)^{\frac{1}{2}}\frac{(\sigma''/\sigma')^2}{(2\pi)^{\frac{3}{2}}}\exp\left[-\frac{\alpha\xi^{*2}}{2(\alpha-1)}\right]. \tag{A.6}$$

The density of saddle-points is obtained by integrating eqn. (A.6):

$$D_{\text{sp}} = \int_{-\infty}^{\infty} P_{\text{sp}}(z^*)\,\mathrm{d}z^* = \frac{2}{6\pi\cdot 3^{\frac{1}{2}}}(\sigma''/\sigma')^2. \tag{A.7}$$

Comparing eqn. (A.7) with eqn. (6), we find that eqn. (12) is confirmed:

$$D_{\text{sp}} = 2D_{\text{sum}}.$$

Finally, the probability density for heights of saddle-points is found by dividing $P_{\text{sp}}(z^*)$ by D_{sp}:

$$p_{\text{sp}}(z^*) = (2\pi)^{-1}\left(\frac{\alpha}{\alpha-1}\right)^{\frac{1}{2}}\exp\left[-\frac{\alpha\xi^{*2}}{2(\alpha-1)}\right], \tag{A.8}$$

which is eqn. (21).

REFERENCES

1 P. R. Nayak, Random process model of rough surfaces, *J. Lubric. Technol.*, 93, (3) (1971) 398–407.
2 M. S. Longuet-Higgins, The statistical analysis of a random moving surface, *Phil. Trans. Roy. Soc.*, A249 (1957) 321–387.

3 J. A. Greenwood and J. H. Tripp, The contact of two nominally flat rough surfaces, *Proc. Inst. Mech. Engrs.*, 185, 48/71 (1970–71) 625–633.
4 P. K. Gupta and N. H. Cook, Statistical analysis of mechanical interaction of rough surfaces, *J. Lubric. Technol.*, 94 (1) (1972) 19–26.
5 T. Hisakado, On the mechanism of contact of solid surfaces, *Bull. Jap. Soc. Mech. Engrs.*, 12 (54) (1969) 1537–1544.
6 J. Pullen and J. B. P. Williamson, On the plastic contact of rough surfaces, *Proc. Roy. Soc. (London)*, A327 (1569) (1972) 159–173.
7 T. E. Tallian, The theory of partial elastohydrodynamic contacts, *Wear*, 21 (1) (1972) 49–102.
8 P. R. Nayak, Some aspects of surface roughness measurement, *Wear*, 26 (1973) 165.
9 S. H. Crandall, Distribution of maxima in the response of an oscillator to random excitation, *J. Acoust. Soc. Amer.*, 47 (1970) 1338–1342.
10 J. B. P. Williamson and R. T. Hunt, The microtopography of solid surfaces, *Research Rept. No. 59, Burndy Corp. Research Div.*, Norwalk, Conn., (1968).
11 S. Corrsin, A measure of the area of a homogeneous random surface in space, *Quart. Appl. Math.*, 12 (1954) 404–408.
12 J. S. Halliday, Surface examination by reflection electron microscopy, *Proc. Inst. Mech. Engrs.*, 169 (1955).
13 A. H. Uppal, S. D. Probert and T. R. Thomas, The real area of contact between a rough and a flat surface, *Wear*, 22 (1972) 163–183.