Tribol Lett (2013) 49:539–543
DOI 10.1007/s11249-012-0094-6

ORIGINAL PAPER

Multiscale Surfaces and Amontons' Law of Friction

J. R. Barber

Received: 4 December 2012 / Accepted: 24 December 2012 / Published online: 26 January 2013
© Springer Science+Business Media New York 2013

Abstract The theories of Archard, Greenwood, and more recently Persson all predict that the area of actual contact between two rough surfaces will be approximately proportional to the applied normal force, though these authors make different assumptions and approach the problem from different points of view. Here we discuss the nature of these approximations and show that the common conclusion follows from the multiscale nature of the surface profile. In particular, it is shown that whatever assumption is made about the nature of friction on the microscale, the macroscale frictional behaviour will always approximate Amontons' law if the surface has sufficient multiscale content.

Keywords Contact · Friction · Multiscale · Surface roughness

1 Introduction

Over a wide range of conditions, the frictional force F opposing sliding of two solid bodies is observed to be approximately proportional to the normal contact force N. This empirical approximation has come to be known as Amontons' law, or Coulomb's law of friction. Of course numerous situations can be found where it is not satisfied, even in an approximate sense, particularly in cases where the contacting bodies are small, but it is sufficiently close to the truth in macroscopic geometries to find an extensive use in engineering practice.

J. R. Barber (✉)
Department of Mechanical Engineering, University of Michigan, 2350 Hayward Street, Ann Arbor, MI 48109-2125, USA
e-mail: jbarber@umich.edu

Many authors have sought to explain the physical basis of Amontons' law, one of the more convincing early explanations being that due to Bowden and Tabor [1], who argued that due to inevitable surface roughness, contact will be restricted to a few areas of "actual contact." They likened the resulting microcontacts to hardness indentations and hence argued that the total area of actual contact $\mathcal{A}$ would be given by $\mathcal{A} = N/H$, where H is the hardness of the softer material. If the local shear stress needed to shear the resulting junctions between the bodies is τ_S, the frictional force is then obtained as $F = \tau_S N/H$, implying a coefficient of friction $\mu = \tau_S / H$—i.e., as the ratio of two material properties.

An early criticism of Bowden and Tabor's theory was that the material near areas of actual contact would deform plastically on first loading, but that work-hardening and residual stress would reduce the likelihood of plastic deformation on subsequent loading, such as during sliding. This argument was supported by measurements of wear between sliding surfaces which suggest that the detachment of a wear particle is an extremely rare consequence of asperity interaction. Two authors in particular (Archard [2] and Greenwood [3]) attempted to derive an explanation of Amontons' law that was not dependent on the assumption of plastic deformation, both appealing to the characteristic shape of the contacting rough surfaces, but approaching the problem from different points of view.

1.1 The Greenwood and Williamson Theory

Greenwood and Williamson's theory [3] is the better known (it was cited 142 times in 2011 alone) and is based on the representation of the rough surface as a set of idealized asperities with an appropriate height distribution. In particular, these authors showed that for the special case of an exponential height distribution, the effect of increasing the normal force N is to increase the number of asperity contacts, but that the distribution with regard to size remains unchanged. As a consequence, all quantities that can be written as the sum of contributions from the separate asperities must be proportional to the number of contacts and hence to the normal force. In particular, the friction force and the electrical contact conductance are both so proportional. A more realistic Gaussian height distribution leads to a relatively weak dependence of friction coefficient on N.

The principal difficulty with Greenwood's theory is that it assumes that the asperity contacts are sufficiently far apart to act independently, and it requires that we decide what constitutes an asperity—notably, at what length scale the asperities are to be defined. What appears to be a single asperity at one scale may resolve into a cluster of smaller asperities at a finer scale. These issues were further discussed by Greenwood in [4, 5].

1.2 The Archard Theory

Archard [2] took as his starting point the multiscale nature of rough surfaces. In particular, he compared the elastic contact problem for a single Hertzian spherical contact, with that for a sphere carrying a collection of smaller spherical asperities, and a three-scale surface, as shown schematically in Fig. 1. He assumed that the normal force carried by each small asperity would be equal to that proportion of the total force as would be carried by the surrounding "catchment area" at the next coarser scale, and thereby showed that the relation between $\mathcal{A}$ and N approaches asymptotically to linearity as the number of scales is increased. It is notable that Archard's model is identifiable as a fractal description of a rough surface, though the fractal approach to rough surface contact was not to be formally enunciated for another 30 years. As with Greenwood's theory, the Archard model predicts linearity between macroscopic quantities of interest, but the constant of proportionality is unfortunately dependent on the scale at which the model is terminated. For example, Archard's prediction of $\mathcal{A}$ for a given normal force N decreases as more scales are added. Nowadays, we would explain this by stating that the elastic actual contact area is itself a fractal [6, 7].

1.3 A Comparison

Both theories depend on the fact that rough surfaces have multiscale properties, since the height distribution in Greenwood's theory results from the asperities "riding" on a longer wavelength waviness. However, the two theories differ in important respects, notably the way in which the fine scale asperities are assumed to be loaded. Think of Archard's two-scale surface as comprising a set of non-linear springs (the fine scale asperities) riding on a coarse scale asperity. The normal displacement of each asperity peak will actually depend both on the force on that asperity and on the underlying coarse-scale deformation, influenced by the forces on the other asperities. Greenwood's theory neglects the second of these two effects, which amounts to assuming that the coarse scale asperity is rigid. By contrast, Archard assumes that the force carried by each fine scale asperity is determined by the local contact pressure at the next coarser scale, which is equivalent to the opposite assumption—i.e., that the fine scale asperities are rigid and only the coarse scale deforms elastically. This is arguably the most significant distinction between the two theories. The truth of course lies somewhere between these extremes, but can only be investigated using numerical methods.

2 Fractal Theories

Although Archard's model has fractal geometry, it is a very specific kind of fractal and probably is not a good description of an actual rough surface. However, the same strategy can be applied to more general surfaces. Ciavarella et al. [7] developed a model that is a direct outgrowth of Archard's, but based on the two-dimensional Weierstrass series representing a fractal surface profile of specified fractal dimension. They followed Archard in determining the force applied to each scale of the model based on the mean contact pressure at the previous scale. Their results exhibited the same features as in Archard's model, including a tendency to linearity between normal force and actual contact area, but with a constant that decreases as fine scales are added.

Persson's theory [8] can also be seen as an application of Archard's protocol for load distribution, but applied to an

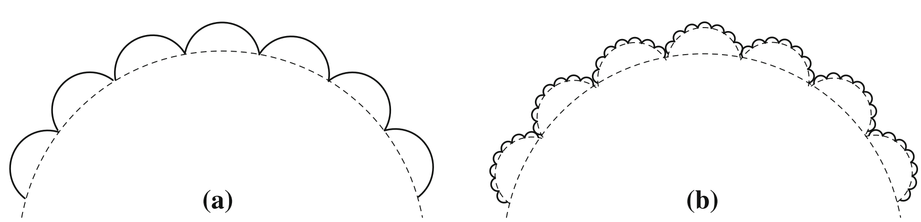

Fig. 1 Archard's two scale (**a**) and three-scale (**b**) surfaces

infinitesimal additional segment $\delta\omega$ of the power spectral density (PSD) of the surface profile $\Phi^P(\omega)$. This theory necessitates a further approximation—that the effect of this infinitesimal segment can be assessed using properties of the "full contact" solution, but it leads to a very elegant relation between the probability $\mathcal{P}(p)$ that a given point experience a contact pressure p and the "truncated" PSD—i.e., a spectral description of the profile with a sharp fine scale cutoff. Under this assumption, $\mathcal{P}(p)$ must satisfy the diffusion equation

$$\frac{\partial^2 \mathcal{P}}{\partial p^2} = 2\frac{\partial \mathcal{P}}{\partial V}, \tag{1}$$

where

$$V = \left(\frac{E^*}{2}\right)^2 \int_0^{\omega_1} \omega^2 \Phi^P(\omega) d\omega \tag{2}$$

is the variance of the pressure distribution needed to maintain full contact with the PSD truncated at the wavenumber ω_1. In this equation, E^* is the composite contact modulus defined by

$$\frac{1}{E^*} = \frac{(1-\nu_1^2)}{E_1} + \frac{(1-\nu_2^2)}{E_2}, \tag{3}$$

where E_1, E_2, ν_1, ν_2 are Young's modulus and Poisson's ratio respectively for the materials of the two contacting bodies. Equations (1, 2) are here presented in the modified form developed by Greenwood and Manners [9].

Persson [8] argues that the square-root asymptotic behaviour of the contact pressure at the edge of an area of actual contact demands that $\mathcal{P}(p,V) \to 0$ as $p \to 0$. If the nominal contact pressure p_0 is uniform, we also have the initial condition $\mathcal{P}(p,0) = \delta(p - p_0)$ which is equivalent to the release of an "instantaneous point source" at the point (p_0,0) in the semi-infinite domain $p > 0$, initially at zero. These conditions and Eq. (1) define a classical diffusion problem whose solution is

$$\mathcal{P}(p,V) = \frac{1}{\sqrt{2\pi V}}\left[\exp\left(-\frac{(p-p_0)^2}{2V}\right) - \exp\left(-\frac{(p+p_0)^2}{2V}\right)\right], \tag{4}$$

where the boundary condition $\mathcal{P}(0,V) = 0$ is satisfied by the location of a negative image source at the point ($-p_0$, 0) in the infinite domain. The proportion of the nominal contact area in actual contact can then be obtained as

$$\mathcal{A}(V) = \int_0^\infty \mathcal{P}(p,V)dp = \operatorname{erf}\left(\frac{p_0}{\sqrt{2V}}\right). \tag{5}$$

If the truncation limit ω_1 is extended, introducing more of the PSD, V increases, and for sufficiently large V we can approximate the error function by the first term in its power series expansion, obtaining

$$\mathcal{A}(V) \to \frac{2p_0}{\sqrt{2\pi V}}, \tag{6}$$

showing that, as in Archard's model, (i) the total area of actual contact approaches linearity with the applied normal force (here p_0 per unit nominal area), and (ii) the constant of proportionality decreases without limit as further fine scale detail is added to the description of the surface ($V \to \infty$).

Notice also that in the fine-scale limit ($\sqrt{V} \gg p_0$), we can approximate the probability distribution (4) as

$$\mathcal{P}(p,V) \to \frac{p_0 p}{V}\sqrt{\frac{2}{\pi V}}\exp\left(-\frac{p^2}{2V}\right), \tag{7}$$

which is virtually indistinguishable from the exact expression for $p_0/\sqrt{2V} < 0.05$.

3 Amontons' Friction Law

Both Archard and Greenwood focussed their attention on the relation between the normal contact force N and the area of actual contact $\mathcal{A}$, implicitly retaining Bowden and Tabor's assumption that the friction force during sliding would be proportional to $\mathcal{A}$. However, for Greenwood's exponential distribution it is not necessary to make this argument. Each asperity is in a state that depends only on the individual asperity compression and since the population distribution of such states remains constant as N is increased (only the total number increases), it follows immediately that the friction force will be linearly proportional to N. We could assume absolutely any nonlinear relation between normal compression and friction force at the fine asperity scale and we would still recover linearity at the macroscale.

A similar result can be deduced for Archard's sphere-on-sphere model. The friction force at each asperity can only depend on the normal force at that asperity, but again we can allow this relation to be arbitrarily nonlinear. Using Archard's protocol to determine the forces on the various fine scale asperities in the multiscale surface leads to progressively more linear relations as more scales are added.

We do not give these derivations here, but instead go straight to the more general Persson theory, in which the equivalent of a nonlinear local frictional behaviour would be for the frictional traction (force per unit area q) on the microscale to be an arbitrary (and hence generally nonlinear) function $q(p)$ of the local normal pressure p. It follows immediately that the total frictional force per unit nominal contact area is

$$q_0 = \int_0^{\infty} \mathcal{P}(p, V) q(p) dp. \quad (8)$$

Now as long as $p_0/\sqrt{2V} < 0.05$, we can replace $\mathcal{P}(p, V)$ by the approximate expression (7), obtaining

$$q_0 = \frac{p_0}{V}\sqrt{\frac{2}{\pi V}} \int_0^{\infty} p \exp\left(-\frac{p^2}{2V}\right) q(p) dp, \quad (9)$$

from which we deduce that q_0 varies linearly with p_0 for all functions $q(p)$. In other words, Amontons' law of friction follows inevitably from the multiscale character of the rough surfaces, provided only that in regions of actual contact there exists some (but arbitrary) relation between the local normal pressure and the shear traction during sliding.

3.1 Fractal Surfaces

The results of the previous section are expressed in terms of the variance of the contact pressure and are not necessarily restricted to fractal surfaces, provided the surface has sufficient multiscale features to ensure that $\sqrt{V}/p_0$ is sufficiently large. However, to assess the accuracy of the approximation (9), it is convenient to consider a quasi fractal surface with a low frequency cutoff at ω_0 and a profile PSD of the form

$$\Phi^P(\omega) = C\left(\frac{\omega}{\omega_0}\right)^{-\beta}; \quad \omega > \omega_0, \quad (10)$$

where the exponent β is related to the fractal dimension D of the profile by $\beta = 5 - 2D$ [10]. Notice that $1 < D < 2$ and hence $1 < \beta < 3$. The cutoff frequency ω_0 might here represent a longest wavelength in the surface roughness, or reflect the finite dimensions of the contacting bodies.

The RMS roughness amplitude σ is defined by

$$\sigma^2 = \int_0^{\infty} \Phi^P(\omega) d\omega = \frac{C\omega_0}{(\beta - 1)}, \quad (11)$$

from Manners and Greenwood [9] Eq. (2), and hence the constant C in (10) can be written

$$C = \frac{(\beta - 1)\sigma^2}{\omega_0}. \quad (12)$$

Substituting (10, 12) into (2), we then obtain

$$V = C\left(\frac{E^*}{2}\right)^2 \int_{\omega_0}^{\omega_1} \omega^2 \left(\frac{\omega}{\omega_0}\right)^{-\beta} d\omega = \left(\frac{\beta - 1}{3 - \beta}\right)\left(\frac{E^*\sigma\omega_0}{2}\right)^2 \left[\left(\frac{\omega_1}{\omega_0}\right)^{3-\beta} - 1\right] \quad (13)$$

and it follows that the criterion $p_0/\sqrt{2V} < 0.05$ is equivalent to

$$\frac{p_0}{E^*} < 0.05\sigma\omega_0 \sqrt{\frac{\beta - 1}{2(3 - \beta)}\left[\left(\frac{\omega_1}{\omega_0}\right)^{3-\beta} - 1\right]} \quad (14)$$

To fix ideas, consider a profile with $D = 1.5$, $\beta = 2$ with an RMS roughness $\sigma = 1$ μm and a cutoff wavenumber $\omega_0 = 1\ \mathrm{mm}^{-1}$. The criterion (14) is then satisfied for nominal pressures $p_0 < 10^{-3}E^*$ (which is a reasonably wide range of pressures if we are not to have bulk yielding of the material) as long as the short wavelength cutoff $\omega_1 > 800\omega_0$.

3.2 Plastic deformation

Equation (1) is based on the assumption that the contact remains elastic and this is clearly unrealistic at extremely high contact pressures p. Persson [11] presented a modified version of his theory in which he argued that the probability distribution $\mathcal{P}(p, V)$ should be truncated at some value S by imposing a second boundary condition $\mathcal{P}(S, V) = 0$. Persson refers to S as being the yield stress S_Y of the material, though in the spirit of Bowden and Tabor [1], it might be more appropriate to use the hardness which is approximately equal to $3S_Y$. We also note that Bower [12], using the methodology of Ciavarella et al. [7] to investigate the elastic-plastic contact of a surface defined by the Weierstrass profile, showed that the constraint of adjacent highly clustered contacts caused the effective mean pressure in the contact zones to increase to a limit of approximately $5.8S_Y$ under predominantly plastic conditions.

The boundary conditions $\mathcal{P}(0, V) = 0, \mathcal{P}(S, V) = 0$ can be satisfied by superposing an appropriate sequence of positive and negative instantaneous sources in the infinite domain, leading to the solution

$$\mathcal{P}(p, V) = \frac{1}{\sqrt{2\pi V}} \sum_{n=-\infty}^{\infty} \left[\exp\left(-\frac{(p - p_0 - 2nS)^2}{2V}\right) - \exp\left(-\frac{(p + p_0 - 2nS)^2}{2V}\right)\right], \quad (15)$$

Also, the proportion of the nominal contact area that is in a plastic state is equal to the integral of the flux across the boundary $p = S$ and this is obtained by analogy with equation (5) as

$$\mathcal{A}_Y = \sum_{n=0}^{\infty} \operatorname{erfc}\left(\frac{(2n + 1)S - p_0}{\sqrt{2V}}\right) - \sum_{n=0}^{\infty} \operatorname{erfc}\left(\frac{(2n + 1)S + p_0}{\sqrt{2V}}\right). \quad (16)$$

These relations take simpler forms in the limit where $p_0/\sqrt{2V} \ll 1$, namely

$$\mathcal{P}(p, V) \rightarrow \frac{2p_0}{V\sqrt{2\pi V}} \sum_{n=-\infty}^{\infty} (p + 2nS) \exp\left(-\frac{(p + 2nS)^2}{2V}\right), \tag{17}$$

and

$$\mathcal{A}_Y \rightarrow \frac{4p_0}{\sqrt{2\pi V}} \sum_{n=0}^{\infty} \exp\left(-\frac{((2n+1)S)^2}{2V}\right), \tag{18}$$

both of which are linearly proportional to the nominal pressure p_0. Thus, if we generalize the relation $q = q(p)$ to include a different value of shear traction q_Y in the plastically deformed regions $\mathcal{A}_Y$, we would obtain

$$q_0 = \int_0^S \mathcal{P}(p, V) q(p) dp + \mathcal{A}_Y q_Y \tag{19}$$

and in the limit (17, 18), this would lead to Amontons' law $q_0 = \mu\, p_0$ as before.

One might take issue with Persson's boundary condition $\mathcal{P}(S, V) = 0$, since in contrast to the condition at $p = 0$, there is no asymptotic argument to justify a square-root bounded approach to the limiting yield condition and indeed Bower's numerical elasto-plastic solution for the Weierstrass profile suggests rather that the limiting state is approached asymptotically rather than suddenly. However, it seems likely that any modification in the evolution of $\mathcal{P}$ near the plastic limit will also result in a linear relation between q_0 and p_0 when $p_0/\sqrt{2V} \ll 1$.

3.3 Size effects

At very small length scales, material properties such as yield strength and hardness tend to increase due to the reduced probability of finding suitably oriented dislocations or slip planes. Such effects can be incorporated at least in a heuristic sense in the present argument by making S (and possibly q_Y) in equations (15–19) a function of ω_1 and hence of V. Similarly, noting that the analyses both of Ciavarella et al. [7] and Gao and Bower [12] predict bifurcation of the contact area down to the wavelength identified with ω_1, we could describe a dependence of the frictional traction $q(p)$ on the size of typical asperity contacts by making q a function of both p and V. It is clear from equations (9, 17–19) that the linear relation between q_0 and p_0 would be retained under these generalizations.

4 Conclusions

We have shown that the Archard and Greenwood and Williamson theories can be considered as opposite extremes in regard to their treatment of interaction between adjacent asperities, but despite this, both theories predict Amonton's law of friction if the surfaces have sufficient levels of multiscale roughness, regardless of the friction mechanism assumed at the finest scale. The same result follows from both Persson's theory and Ciavarella's solution for the Weierstrass profile, though Persson's argument is based on results derived from a "full contact" analysis, whereas Ciavarella's (which requires unrealistic scale separation) shows that the problem becomes dominated by partial contact states at the fine scale. A similar conclusion follows from the extension of these two theories into the elastic-plastic régime. We conclude that Amonton's law is a direct consequence of the multiscale nature of most rough surfaces, and hence that it is not surprising that significant deviations from the law are observed in nanoscale and AFM experiments.

References

1. Bowden, F.P., Tabor, D.: The Friction and Lubrication of Solids. Clarendon Press, Oxford (1950)
2. Archard, J.F.: Elastic deformation and the laws of friction. Proc. R. Soc. Lond. **A243**, 190–205 (1957)
3. Greenwood, J.A., Williamson, J.B.P.: The contact of nominally flat surfaces. Proc. R. Soc. Lond. **A295**, 300–319 (1966)
4. Greenwood, J.A.: A unified theory of surface roughness. Proc. R. Soc. Lond. **A393**, 133–157 (1984)
5. Greenwood, J.A., Wu, J.J.: Surface roughness and contact: an apology. Meccanica **36**, 617–630 (2001)
6. Borri-Brunetto, M., Carpinteri, A., Chiaia, B.: Lacunarity of the contact domain between elastic bodies with rough boundaries. In: Frantziskonis G. (ed.) Probamat 97, Probabilities and Materials. Kluwer, Dordrecht (1997)
7. Ciavarella, M., Demelio, G., Barber, J.R., Jang, Y-H.: Linear elastic contact of the Weierstrass profile. Proc. R. Soc. Lond. **A456**, 387–405 (2000)
8. Persson, B.N.J.: Theory of rubber friction and contact mechanics. J. Chem. Phys. **115**, 3840–3861 (2001)
9. Manners, W., Greenwood, J.A.: Some observations on Persson's diffusion theory of elastic contact. Wear **261**, 600–610 (2006)
10. Majumdar, A., Bhushan, B.: Characterization and modeling of surface roughness and contact mechanics. In: Handbook of Micro/Nano Tribology, pp.109–165. CRC Press, New York (1995)
11. Persson, B.N.J.: Elastoplastic contact between randomly rough surfaces. Phys. Rev. Lett. **87**, Art. No. 116101 (2001)
12. Gao, Y.F., Bower, A.F.: Elastic-plastic contact of a rough surface with Weierstrass profile. Proc. R. Soc. Lond. **A462**, 319–348 (2006)