INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **17** (2005) R1–R62 doi:10.1088/0953-8984/17/1/R01

TOPICAL REVIEW

On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion

B N J Persson[1,2], O Albohr[3], U Tartaglino[1,4,5], A I Volokitin[1] and E Tosatti[2,4,5]

[1] IFF, FZ-Jülich, 52425 Jülich, Germany
[2] International Centre for Theoretical Physics (ICTP), PO Box 586, I-34014 Trieste, Italy
[3] Pirelli Deutschland AG, 64733 Höchst/Odenwald, Postfach 1120, Germany
[4] International School for Advanced Studies (SISSA), Via Beirut 2, I-34014 Trieste, Italy
[5] INFM Democritos National Simulation Centre, Trieste, Italy

Received 6 August 2004, in final form 29 October 2004
Published 10 December 2004
Online at stacks.iop.org/JPhysCM/17/R1

Abstract

Surface roughness has a huge impact on many important phenomena. The most important property of rough surfaces is the surface roughness power spectrum $C(q)$. We present surface roughness power spectra of many surfaces of practical importance, obtained from the surface height profile measured using optical methods and the atomic force microscope. We show how the power spectrum determines the contact area between two solids. We also present applications to sealing, rubber friction and adhesion for rough surfaces, where the power spectrum enters as an important input.

(Some figures in this article are in colour only in the electronic version)

Contents

1. Introduction

Surface roughness has an enormous influence on many important physical phenomena such as contact mechanics, sealing, adhesion and friction. Thus, for example, experiments have shown that a substrate with a root mean square (rms) roughness of order $\sim$1 μm can already completely remove the adhesion between a rubber ball and a substrate, while nanoscale roughness will remove the adhesion between most hard solids, e.g., metals and minerals; in other words, roughness is the main reason that adhesion is usually not observed in most macroscopic phenomena. Similarly, rubber friction on most surfaces of practical interest, e.g., road surfaces, is mainly due to the roughness induced pulsating forces which act on the rubber surface as it slides over the substrate asperities.

Let us illustrate the importance of surface roughness with three modern applications. At present there is a strong effort to produce small mechanical devices, e.g., micromotors. The largest problem in the development of such devices is the adhesion and, during sliding, the friction and wear between the contacting surfaces [1]. As an example, in figure 1 we show the simplest possible microdevice, namely a micrometre cantilever beam. (Suspended micromachined structures such as plates and beams are commonly used in the manufacturing of pressure and accelerator sensors.) If the beam is sufficiently long or thin the free beam state in (a) will be unstable, and the bound state in (b) will correspond to the minimum free energy state [2]. Roughly speaking, the state (b) is stable if the energy of binding to the substrate is higher than the elastic energy stored in the bent beam. The energy of binding to the substrate can be strongly reduced by introducing (or increasing) the surface roughness on the substrate (see section 7.1). In addition, if the surfaces are covered by appropriate monolayer films, these surfaces can be made hydrophobic thus eliminating the possibility of the formation of (water) capillary bridges.

A second application is the formation of hydrophobic coatings on surfaces by creating the appropriate type of surface roughness[6]. This amounts to copying nature where many plant

[6] See http://www.botanik.uni-bonn.de/system/bionik_flash.html for information involving surface roughness in relation to hydrophobicity and surface self-cleaning in biological systems.

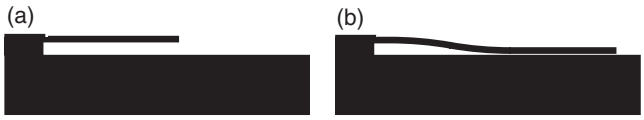

Figure 1. (a) A micrometre sized cantilever beam. (b) If the beam is too long or too thin the minimum free energy state corresponds to the beam partly bound to the substrate. Surface roughness lowers the binding energy (per unit area) and hence stabilizes the non-bonded state in (a).

Figure 2. A water droplet on a superhydrophobic surface: the droplet touches the leaf only at a few points and forms a ball. It completely rolls off at the slightest declination (see footnote 6). Reproduced with permission.

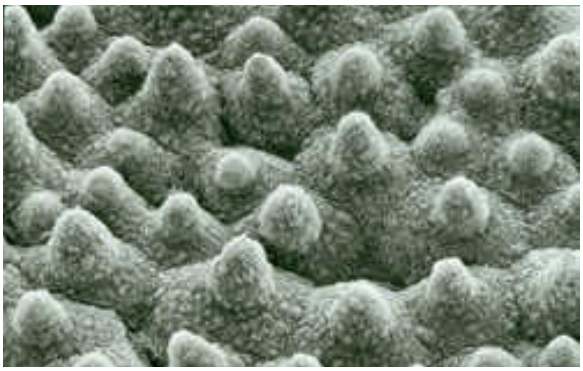

Figure 3. A leaf surface with roughness on several length scales optimized (via natural selection) for hydrophobicity and self-cleaning. Through the combination of microstructure (cells) and nanostructure (wax crystals) the water contact angle θ_0 is maximized (see footnote 6). Reproduced with permission.

surfaces are found to be highly hydrophobic (figure 2) as a result of the formation of special types of surface roughness (figure 3). The surface roughness allows air to be trapped between the liquid and the substrate, while the liquid is suspended on the tips of the asperities. Since the area of real liquid–substrate contact is highly reduced, the contact angle of the drop is determined almost solely by the surface tension of the liquid, leading to a very large contact angle. New commercial products based on this 'lotus effect' [3], such as self-cleaning paints and glass windows, have been produced.

Finally, we mention the effect of surface roughness on direct wafer bonding [4]. Wafer bonding at room temperature is due to relatively weak interatomic attraction forces, e.g.,

the van der Waals interaction or hydrogen bonding, giving (for perfectly flat surfaces) an interfacial binding energy of order 6 meV Å^{-2}. The wafer surface roughness is the most critical parameter determining the strength of the wafer bonding. In particular, when the surface roughness exceeds a critical value, the wafers will not bind at all, in agreement with the theory presented in section 7.1. Primary grade polished silicon wafer surfaces have rms roughness of order ∼0.1 nm when measured over a 10×10 μm^2 surface area, and such surfaces bind spontaneously. However, when the surface roughness amplitude is of order 1 nm the surfaces either bind (slowly) when squeezed together at high enough pressure, or they do not bind at all, depending on the detailed nature of the surface roughness power spectra.

Surfaces with 'ideal' roughness, e.g., prepared by fracture or by some growth process, have been studied intensively for many years [5–8]. However, much less information has been available for more common surfaces of engineering interest. In this article we discuss the nature of the power spectra of some surfaces of practical importance. As illustrations we discuss contact mechanics, sealing, rubber friction and adhesion.

This review is organized as follows. In section 2 we define the surface roughness power spectrum $C(q)$, and discuss some of its properties. In section 3 we present power spectra deduced from surface topography measurements for different surfaces of technological importance. Section 4 illustrates how the surface roughness power spectrum determines the contact area between elastic and viscoelastic solids. In section 5 we briefly consider the influence of surface roughness on sealing. In section 6 we discuss rubber friction on rough substrates, and show some results for how the friction depends on the surface roughness power spectrum. We also discuss the influence of tyre–road polishing and discuss rubber friction on wet road surfaces. In section 7 we discuss adhesion between rough surfaces and present two applications which illustrate the importance of surface roughness for adhesion in technology and biology. Section 8 contains our summary and an outlook.

2. The surface roughness power spectrum: definition and general properties

The influence of roughness on the adhesional and frictional properties introduced above is mainly determined by the surface roughness power spectrum $C(q)$ (or power spectral density) defined by [9]

$$C(q) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2x \, \langle h(\boldsymbol{x}) h(\boldsymbol{0}) \rangle \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}. \tag{1}$$

Here $\boldsymbol{x} = (x, y)$ and $z = h(\boldsymbol{x})$ is the substrate height measured from the average surface plane, defined so that $\langle h \rangle = 0$. The $\langle \cdots \rangle$ stands for ensemble averaging, i.e., averaging over a collection of different surfaces with identical statistical properties. We have assumed that the statistical properties of the substrate are translationally invariant, so that the correlation $\langle h(\boldsymbol{x}+\boldsymbol{x}_0)h(\boldsymbol{x}_0)\rangle$ does not depend on the choice of $\boldsymbol{x}_0$, but only on the in-plane distance vector $\boldsymbol{x}$. Alternatively the height correlation can be defined through a spatial average over $\boldsymbol{x}_0$:

$$\lim_{L\to\infty} \frac{1}{L^2} \int_{-L/2}^{+L/2} \mathrm{d}x_0 \int_{-L/2}^{+L/2} \mathrm{d}y_0 \, h(x_0 + x, y_0 + x) h(x_0, y_0).$$

The latter definition is quite popular in the engineering community [9], and it is closer to the experimental procedure for measuring the power spectrum. We notice, however, that the lateral size L of the surface area is always finite in any experimental measurement or numerical calculation. In some special cases, such as the ones discussed in section 3.1, the spatial average is not equivalent to the ensemble average for any practical choice of L. Finally we point out that the power spectrum can also be defined through the square modulus of the Fourier transform

of h; this is indeed the conventional approach adopted in the context of signal theory and electronics, and it explains the origin of the name 'power spectral density'. Nonetheless, such an approach is equivalent to equation (1) because of the Wiener–Khintchine theorem (see appendix C and particularly equation (C.6)).

Together with translational invariance, in (1) we have also assumed that the statistical properties of the substrate are isotropic, so that $C(q)$ only depends on the magnitude $q = |\boldsymbol{q}|$ of the wavevector $\boldsymbol{q}$. Note that from (1) it follows that

$$\langle h(\boldsymbol{x})h(\boldsymbol{0})\rangle = \int \mathrm{d}^2q\, C(q)\mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}$$

so the root mean square roughness amplitude $\sigma = \langle h^2\rangle^{1/2}$ is determined by

$$\langle h^2\rangle = \int \mathrm{d}^2q\, C(q) = 2\pi \int_0^\infty \mathrm{d}q\, qC(q). \tag{2}$$

In reality, there will always be an upper limit and a lower limit to the q-integral in (2). Thus, the largest possible wavevector will be of order $2\pi/a$, where a is a short wavelength cut-off corresponding perhaps to some lattice constant, while the smallest possible wavevector is of order $2\pi/L$ where L is the linear size of the surface. In general, one may define a root mean square roughness amplitude which depends on the range of roughness (q_0, q_1) involved in the integral in (2):

$$\langle h^2\rangle(q_0, q_1) = 2\pi \int_{q_0}^{q_1} \mathrm{d}q\, qC(q). \tag{3}$$

For a randomly rough surface, when $h(\boldsymbol{x})$ is a Gaussian random variable, the statistical properties of the surface are completely defined by the power spectrum $C(q)$ (see appendix A). In this case the height probability distribution

$$P_\mathrm{h} = \langle\delta[h - h(\boldsymbol{x})]\rangle$$

will be a Gaussian:

$$P_\mathrm{h} = \frac{1}{(2\pi)^{1/2}\sigma}\mathrm{e}^{-h^2/2\sigma^2}.$$

The height distributions of many natural surfaces, e.g., surfaces prepared by fracture, or surfaces prepared by bombardment by small particles (e.g., sand blasting or ion sputtering), are usually nearly Gaussian. On the other hand, rough surfaces, e.g., prepared by fracture, and then (slightly) polished, display a non-symmetric height distribution (i.e., no symmetry as $h \to -h$) since the asperity tops have been polished more than the bottoms of the valleys. These surfaces, typically of considerable practical importance—see below—have non-Gaussian height distributions. For such surfaces it is interesting to study the *top*, C_T, and the *bottom*, C_B, power spectra defined by

$$C_\mathrm{T}(q) = \frac{1}{(2\pi)^2}\int \mathrm{d}^2x\, \langle h_\mathrm{T}(\boldsymbol{x})h_\mathrm{T}(\boldsymbol{0})\rangle\mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}} \tag{4a}$$

$$C_\mathrm{B}(q) = \frac{1}{(2\pi)^2}\int \mathrm{d}^2x\, \langle h_\mathrm{B}(\boldsymbol{x})h_\mathrm{B}(\boldsymbol{0})\rangle\mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}} \tag{4b}$$

where $h_\mathrm{T}(\boldsymbol{x}) = h(\boldsymbol{x})$ for $h > 0$ and zero otherwise, while $h_\mathrm{B}(\boldsymbol{x}) = h(\boldsymbol{x})$ for $h < 0$ and zero otherwise. These are 'rectified' profiles; see figure 4. It is easy to show that $C \approx C_\mathrm{T} + C_\mathrm{B}$. It is also clear by symmetry that for a surface prepared by fracture, $C_\mathrm{T}(q) = C_\mathrm{B}(q)$, since what is the top of one of the cracked block surfaces is the bottom of the other (opposite) crack surface, and vice versa; see figure 5. However, if the cracked surface is slightly polished then, because the contact pressure with the polishing object (e.g., sandpaper) is highest at the asperity tops,

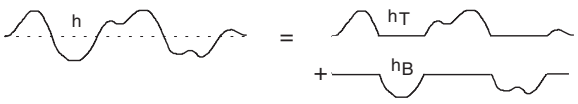

Figure 4. The surface profile $h(x)$ is decomposed into a top $h_{\mathrm{T}}(x)$ and a bottom $h_{\mathrm{B}}(x)$ profile.

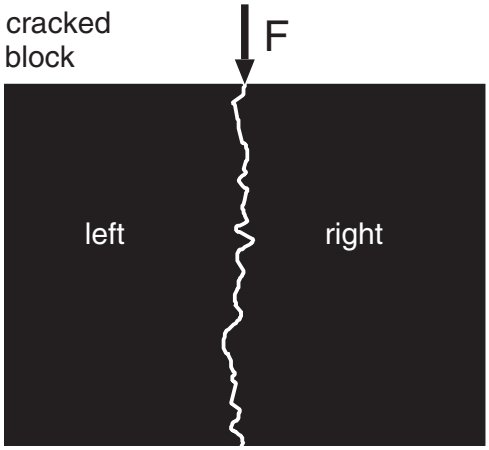

Figure 5. Rough surfaces prepared by crack propagation have surface roughness with statistical properties which must be invariant under the replacement of $h \to -h$. This follows from the fact that what is a valley on one of the crack surfaces (say the left) is an asperity with respect to the other crack surface (right). Thus the top and bottom power spectra must obey $C_{\mathrm{T}}(q) = C_{\mathrm{B}}(q)$.

the polishing at the asperity tops will be stronger than at the valley bottoms, and thus $C_{\mathrm{B}} > C_{\mathrm{T}}$. If n_{T} and n_{B} are the fractions of the nominal surface area (i.e., the surface areas projected on the xy-plane) where $h > 0$ and $h < 0$, respectively, with $n_{\mathrm{T}} + n_{\mathrm{B}} = 1$, then we may also define $C^*_{\mathrm{T}}(q) = C_{\mathrm{T}}/n_{\mathrm{T}}$ and $C^*_{\mathrm{B}} = C_{\mathrm{B}}/n_{\mathrm{B}}$. In general, $n_{\mathrm{T}} \approx n_{\mathrm{B}} \approx 0.5$ and for surfaces prepared by fracture $n_{\mathrm{T}} = n_{\mathrm{B}} = 0.5$. Roughly speaking, C^*_{T} would be the power spectrum resulting if the actual bottom profile (for $h < 0$) was replaced by a mirrored top profile (for $h > 0$). A similar statement holds for C^*_{B}.

Many surfaces tend to be nearly self-affine fractal. A self-affine fractal surface has the property that if part of the surface is magnified, with a magnification which in general is appropriately different in the direction perpendicular to the surface as compared to the lateral directions, then the surface 'looks the same', i.e., the statistical properties of the surface are invariant under the scale transformation (see appendix B). For a self-affine surface the power spectrum has the power-law behaviour

$$C(q) \sim q^{-2(H+1)},$$

where the Hurst exponent H is related to the fractal dimension D_{f} of the surface via [5] $H = 3 - D_{\mathrm{f}}$. Of course, for real surfaces this relation only holds in some finite wavevector region $q_0 < q < q_1$, and in a typical case $C(q)$ has the form shown in figure 6. Note that in many cases there is a roll-off wavelength q_0 below which $C(q)$ is approximately constant. We will discuss this point further, below.

Finally, note that while the root mean square roughness is usually dominated by the longest wavelength surface roughness components, higher order moments of the power spectra such as the average slope or the average surface curvature are dominated by the shorter wavelength components. For example, assuming a self-affine fractal surface, equation (3) gives

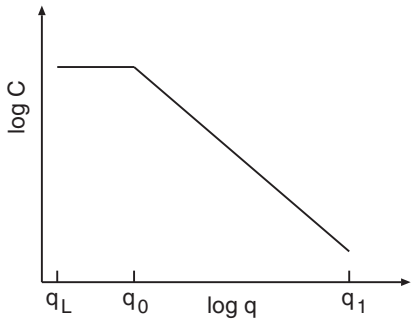

Figure 6. The surface roughness power spectrum of a surface which is a self-affine fractal for $q_0 < q < q_1$. The long distance roll-off wavevector q_0 and the short distance cut-off wavevector q_1 depend on the system under consideration. The slope of the $\log C$–$\log q$ plot for $q > q_0$ determines the fractal exponent of the surface (see the text). The lateral size L of the available surface region determines the smallest possible wavevector $q_L = 2\pi/L$.

$$\langle h^2 \rangle (q_0, q_1) \sim \int_{q_0}^{q_1} \mathrm{d}q\, q^{-2H-1} \sim q_0^{-2H} - q_1^{-2H} \approx q_0^{-2H}$$

if $q_1/q_0 \gg 1$. However, the average slope and the average curvature have additional factors of q^2 and q^4, respectively, in the integrand of the q-integral, and these quantities are therefore dominated by the large q (i.e., short wavelength) surface roughness components (see appendix C).

3. The surface roughness power spectrum: experimental results

In this section we present power spectra for different surfaces of practical importance. The power spectra have been calculated using equations (1), (4*a*) and (4*b*) (see appendix D), where the height profile $h(\boldsymbol{x})$ has been measured using either optical methods[7] or by atomic force microscopy [8].

3.1. Surfaces produced by crack propagation

Figure 7 shows the power spectra $C(q)$ for three freshly cleaved stone surfaces, namely a granite and two basalt stone surfaces. Here, and in what follows, we show the power spectra on a log–log scale (where log stands for the base 10 logarithm). Note that the granite and basalt surfaces, in spite of the rather different mineral microstructures (see below), give identical power spectra within the accuracy of the measurement. It has been stated (see, e.g., [10]) that surfaces produced by crack propagation have self-affine fractal structure with the universal fractal dimension $D_\mathrm{f} \approx 2.2$. However, our measured $\log C$–$\log q$ relations are not perfectly straight lines, i.e., the surfaces in the length scale range studied cannot be accurately described as self-affine fractals. Moreover, the average slope of the curves in figure 7 corresponds to the fractal dimension $D_\mathrm{f} \approx 2$ rather than 2.2.

The similarity of the power spectra for the basalt and granite surfaces in figure 7 is striking. Granite and basalt both result from magma and have similar compositions, consisting mainly of minerals from the silicate group. However, granite originates from magma which was trapped deep in the crust, taking a very long time to cool down and crystallize into solid rock. As a result granite is a coarse-textured rock where individual mineral grains are easily seen. Basalt, on the

[7] Many methods for measurement of surface topography have been developed; see, e.g., http://www.michmet.com, http://www.solarius-inc.com or http://www.schmitt-ind.com for experimental equipment and some illustrative results.

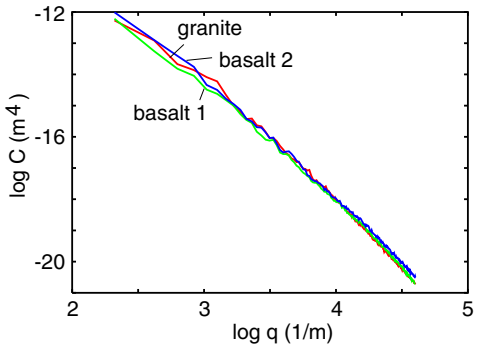

Figure 7. The surface roughness power spectra for two freshly cleaved basalt surfaces and a fresh granite surface.

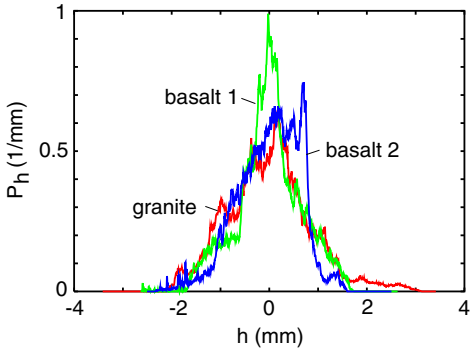

Figure 8. The height distribution P_h for two freshly cleaved (cobblestone) basalt surfaces and a fresh granite surface. Note the random non-Gaussian nature of the height profiles.

other hand, results from fast cooling of magma from, e.g., volcanic eruptions, is fine grained, and nearly impossible to resolve into the individual mineral grains without magnification. In spite of these differences, the surface roughness power spectra of freshly cleaved surfaces are nearly identical. This may indicate some kind of universal power spectrum for surfaces resulting from cleaving of mineral stones of different types, a point that could merit further investigation.

Note that there is no roll-off region for these fracture-produced surfaces, whose behaviour appears fractal-like up to the longest length scale studied, determined by the lateral size L of the surfaces (or of the regions experimentally studied, of the order of 1 cm in our case), i.e., with reference to figure 6, $q_0 = q_L \equiv 2\pi/L$. One consequence of this is that the rms roughness amplitude is determined mainly by the $\lambda \sim L$ wavelength fluctuations of the surface height, and will therefore depend on the size L of the surface. Furthermore, the height distribution P_h obtained for any given realization of the rough surface will not be Gaussian, but will exhibit random fluctuations as compared to other realizations (see figure 8, which illustrates this point for the three stone surfaces discussed above). However, the ensemble averaged height distribution (not shown) should be Gaussian or nearly Gaussian. Thus, when there is no roll-off region in the measured power spectra, averaging over the surface area is not identical to ensemble averaging. However, when there is a roll-off wavevector, $q_0 = 2\pi/\lambda_0$, and if the surface is studied over a region with the lateral size $L \gg \lambda_0$, ensemble averaging and averaging over the surface area $L \times L$ will give identical results for P_h, and the rms roughness amplitude will be independent of L for $L \gg \lambda_0$.

3.2. Polished crack surfaces

In the past, cobblestones, made of granite or basalt, were frequently used for road surfaces. However, these surfaces do not exhibit good frictional properties against rubber. In particular, with increasing time of use, the cobblestone surfaces become polished by slipping tyres (see section 6.2), and that polishing results in a reduced rubber–road friction, even for dry driving conditions[8]. Figure 9 illustrate this polishing effect. It shows the power spectrum of a heavily used (basalt) cobblestone, and of two freshly cleaved surfaces (from figure 7), from the same

[8] Cobblestones made from porphyry (a volcanic rock), thanks to the difference in hardness of the minerals that it contains (see section 6.2), have better polishing properties than cobblestones made of granite and basalt; see, e.g., http://www.bourgetbros.com.

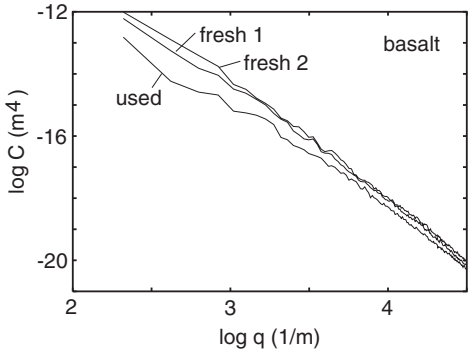

Figure 9. The surface roughness power spectra $C(q)$ for two freshly cleaved cobblestone (basalt) surfaces, and for a wear-polished (used) surface.

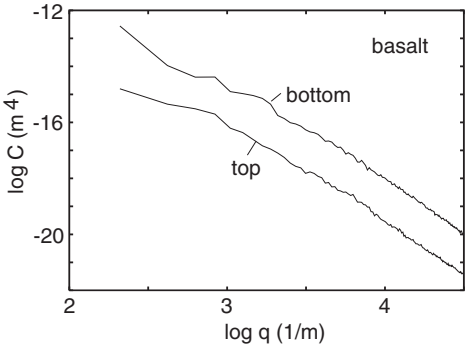

Figure 10. The top C_{T}^{*} and the bottom C_{B}^{*} surface roughness power spectra $C(q)$ for a used cobblestone (basalt) surface.

cobblestone. At long wavelength the power spectrum of the heavily used surface is nearly one decade smaller than that of the freshly prepared surfaces. The effect of polishing is further emphasized by calculating the top and bottom power spectra, C_{T}^{*} and C_{B}^{*}, as shown in figure 10. The top power spectrum of the worn surface is a factor $\sim$30 times smaller than the bottom spectrum for *all wavevectors*. As anticipated, the asymmetry arises from the higher polishing of asperities relative to valleys. It is of course crucial to take this polishing effect into consideration when designing road surfaces, and we will discuss this point further in section 6.2.

3.3. Surfaces with long distance roll-off

As pointed out above, surfaces prepared by fracture have no natural long distance cut-off and the rms roughness amplitude increases continuously and without limit as the probed surface area increases. This is similar to Brownian motion where the mean square displacement increases without limit (as $\sim t^{1/2}$) as the time t increases. However, most surfaces of engineering interest have a long distance cut-off or roll-off wavelength λ_0 corresponding to a wavevector $q_0 = 2\pi/\lambda_0$, as shown in figure 6. For example, if a flat surface is sand blasted for some time the resulting rough surface will have a long distance roll-off length, which increases with the time of sand blasting. Similarly, if atoms or particles are deposited on an initially flat surface the resulting rough surface will have a roll-off wavelength which increases with the deposition time, as has been studied in detail in recent growth models. Another way to produce a surface with a long distance roll-off wavelength is to prepare the solid as a conglomerate of small particles. A nominally flat surface of such a solid has still roughness on length scales shorter than the diameter of the particles, which therefore may act as a long distance roll-off wavelength. We illustrate this here with a solid produced by sintering together corundum particles at high temperature and pressure (figure 11), and for a sandpaper surface (figure 13). For both surfaces the height distribution P_{h} is smooth (see figures 12 and 14), since averaging over a surface area with lateral size $L \gg \lambda_0$ is equivalent to ensemble averaging.

The sandpaper surface in figure 13 was studied using the AFM at two different resolutions over square areas of 20×20 and 100×100 $\mu\mathrm{m}^2$ as indicated by the two different curves in figure 13. The height distribution P_{h} (and hence also the rms roughness amplitude) calculated from these two different measurements over different surface areas—see figure 14—are nearly identical, as expected when L is larger than the roll-off length λ_0.

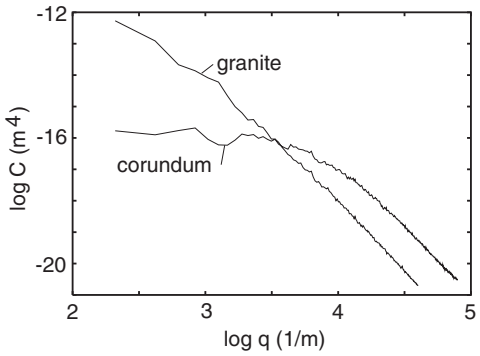

Figure 11. The surface roughness power spectra for a fresh granite surface and a fresh sintered corundum surface.

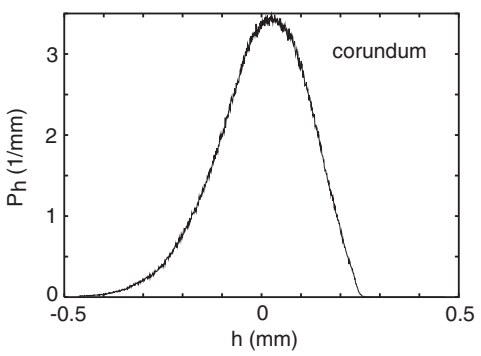

Figure 12. The height distribution P_{h} as a function of the height h for a sintered corundum surface.

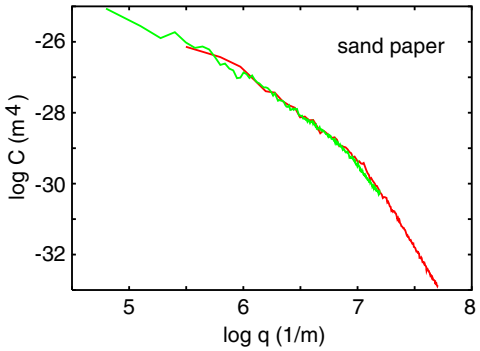

Figure 13. The surface roughness power spectra $C(q)$ for a sandpaper surface. The two curves are based on the height profiles measured with an AFM at two different spatial resolutions over 20×20 μm^2 (black (red)) and 100×100 μm^2 (grey (green)) square areas.

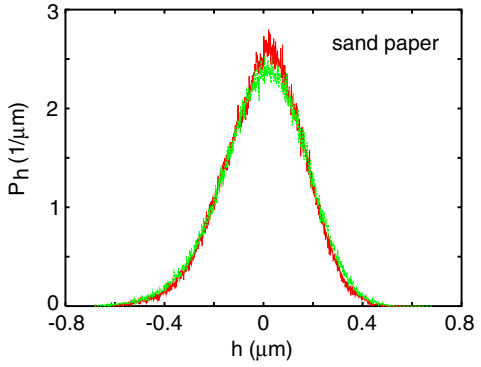

Figure 14. The surface roughness height probability distribution P_{h} for a sandpaper surface. The two curves are based on the height profiles measured with an AFM at two different spatial resolutions over 20×20 μm^2 (black (red)) and 100×100 μm^2 (grey (green)) square areas.

3.4. Road surfaces

Asphalt and concrete road surfaces have nearly perfect self-affine fractal power spectra, with a very well defined roll-off wavevector $q_0 = 2\pi/\lambda_0$ of order 1000 m^{-1}, corresponding to $\lambda_0 \approx 1$ cm, which reflects the largest stone particles used in the asphalt. This is illustrated in figure 15 for two different asphalt surfaces. From the slope of the curves for $q > q_0$ one can deduce the fractal dimension $D_{\mathrm{f}} \approx 2.2$, which is typical for asphalt and concrete road surfaces. The height distributions of the two asphalt surfaces are shown in figure 16. Note that the rms roughness amplitude of surface 2 is nearly twice as high as for surface 1. Nevertheless, the tyre–rubber friction is slightly higher on the road surface 1 because it has slightly larger power spectra for most values of q in figure 15. Thus there is in general *no* direct correlation between the rms roughness amplitude and the rubber friction on road surfaces, as will be further discussed in section 6.

Many attempts have been made to relate rubber friction on road surfaces to the so-called 'sand filling number'. The sand filling number is the amount of very fine-grained sand needed to fill out all the road surface cavities in a given surface area. However, no correlation between

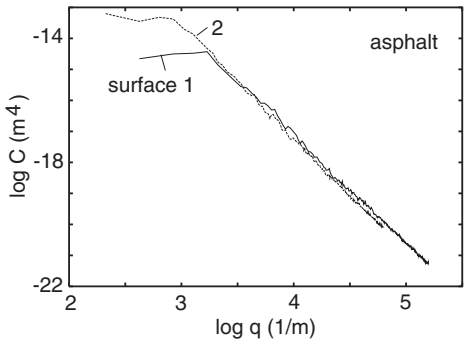

Figure 15. The surface roughness power spectra $C(q)$ for two asphalt road surfaces.

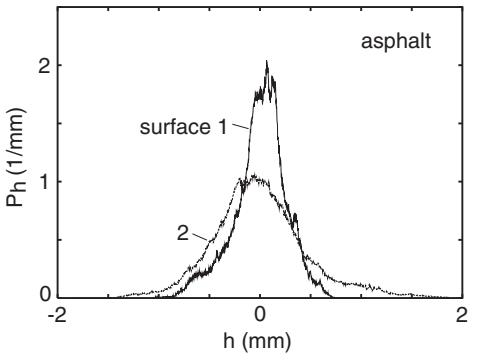

Figure 16. The height distribution P_h for two different asphalt road surfaces.

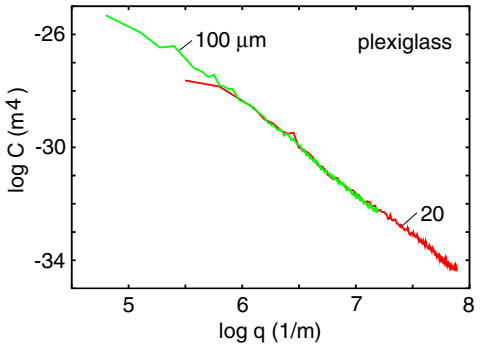

Figure 17. The surface roughness power spectra $C(q)$ for a plexiglass surface. The two curves are based on the height profiles measured with an AFM at two different spatial resolutions over 20 × 20 and 100 × 100 μm^2 square areas.

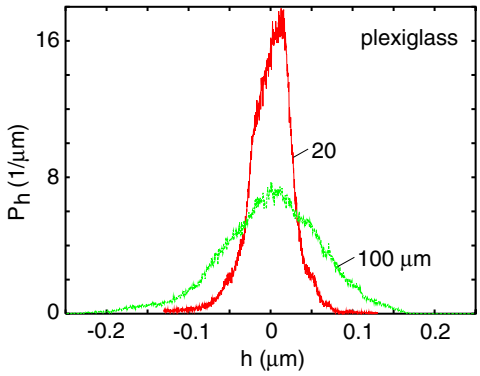

Figure 18. The surface roughness height probability distribution P_h for a plexiglass surface. The two curves are based on the height profiles measured with an AFM at two different spatial resolutions over 20 × 20 and 100 × 100 μm^2 square areas.

the sand filling number and rubber friction on dry road surfaces has been found [11]. In the light of modern rubber friction theories, this result is not unexpected since the rubber friction depends on the power spectrum for *all wavevectors*, while only the long wavelength components contribute appreciably to the sand filling number. Thus, for example, for the two asphalt surfaces in figure 15, the sand filling factor of road 2 is nearly twice as large as that for road 1, but the rubber friction is slightly higher on road surface 1.

3.5. Other surfaces of practical interest

Finally, let us consider two other surfaces of practical importance. Figure 17 shows the power spectrum of a plexiglass surface measured (using an AFM) at two different resolutions over two different surface areas 20 × 20 and 100 × 100 μm^2 wide. This surface does not exhibit a roll-off wavevector in the wavevector range studied and the height distributions deduced from the two different surface areas differ strongly—see figure 18—with the rms roughness amplitude being more than twice as large for the measurement over the larger surface area. The log C–log q relation in figure 17 is a nearly perfect straight line, and the slope corresponds to the fractal dimension $D_f \approx 2.4$.

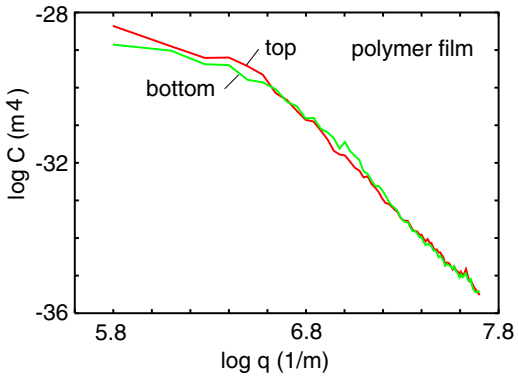

Figure 19. The top C_T^* and the bottom C_B^* surface roughness power spectra for a polymer film spin coated on a very smooth substrate and dried. From [12].

Figure 19 shows the top and bottom power spectra C_T^* and C_B^* for a polymer film which was spin coated and dried on a flat silicon surface. In this case there is a roll-off wavelength $\lambda_0 \approx 1\ \mu$m which is probably related to the average polymer film thickness. Note that within the accuracy of the experiment, $C_T^* = C_B^*$, proving qualitatively that the short wavelength roughness in the large valleys is very similar to the short wavelength roughness on the large asperities. The $\log C$–$\log q$ relation in figure 19 for $\log q > 6.6$ is a nearly perfect straight line, and the slope corresponds to a roughness exponent $H = 1.6$. This is larger than unity, which implies that the surface cannot be described as fractal at any scale.

4. Contact mechanics

Practically all macroscopic bodies have surfaces with roughness on many different length scales. When two bodies with nominally flat surfaces are brought into contact, real (atomic) contact will only occur in small randomly distributed areas, and the area of real contact is usually an extremely small fraction of the nominal contact area. We can visualize the contact regions as small areas where asperities from one solid are squeezed against asperities of the other solid; depending on the conditions the asperities may deform elastically or plastically.

How large is the area of *real* contact between a solid block and a substrate? This fundamental question has extremely important practical implications. For example, it determines the contact resistivity and the heat transfer between the solids. It is also of direct importance for wear and sliding friction [13], e.g., the rubber friction between a tyre and a road surface, and has a major influence on the adhesive force between two solid blocks in direct contact.

Contact mechanics has a long history. The first study was presented by Hertz [14]. He gave the solution for the frictionless normal contact of two elastic bodies of quadratic profile. He found that the area of real contact ΔA varies non-linearly with the load or squeezing force: $\Delta A \propto F_N^{2/3}$. In 1957 Archard [15] applied the Hertz solution to the contact between rough surfaces and showed that for a simple fractal-like model, where small spherical bumps (or asperities) were distributed on top of larger spherical bumps and so on, the area of real contact varies *nearly linearly* with F_N. A similar conclusion was reached by Greenwood and Williamson [16–18] who again assumed asperities with spherical summits (of identical radius) with a Gaussian distribution of heights, as sketched in figure 20(b). A more general contact mechanics theory has been developed by Bush *et al* [19]. They approximated the summits as paraboloids and applied the classical Hertzian solution for the deformation. The height

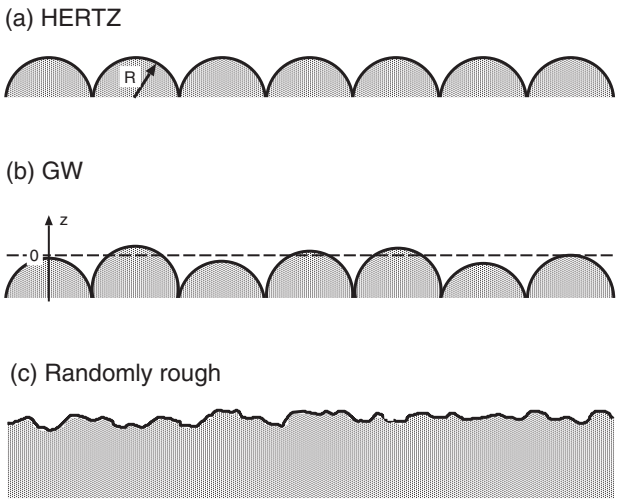

Figure 20. Three models of 'rough' surfaces. In case (a) all the 'asperities' are equally high and have identical radius of curvature. In this case, according to the Hertz contact theory, the area of real contact ΔA between a solid with a flat surface and the surface depends non-linearly on the squeezing force (or load) F_N according to $\Delta A \sim F_N^{2/3}$. If the asperities have instead random height distribution as in (b) then, for small F_N, ΔA is *nearly* proportional to the squeezing force. If the surface roughness is random with 'asperities' of different heights and curvature radii as in (c), the area of real contact for small F_N is *exactly* proportional to the squeezing force.

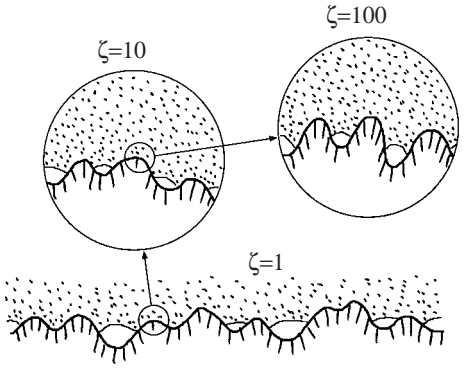

Figure 21. A rubber block (dotted area) in adhesive contact with a hard rough substrate (dashed area). The substrate has roughness on many different length scales and the rubber makes partial contact with the substrate on all length scales. When a contact area is studied at low magnification ($\zeta = 1$) it appears as if complete contact occurs in the macroasperity contact regions, but when the magnification is increased it is observed that in reality only partial contact occurs.

distribution was described by a random process, and they found that at low squeezing force F_N the area of real contact increases linearly with F_N.

Figure 21 shows the contact between two solids at increasing magnification ζ. At low magnification ($\zeta = 1$) it looks as if complete contact occurs between the solids at many *macroasperity* contact regions. When the magnification is increased and smaller length scale roughness is detected, it can be observed that only partial contact occurs at the asperities. In fact, if there were no short distance cut-off the true contact area would eventually vanish. In

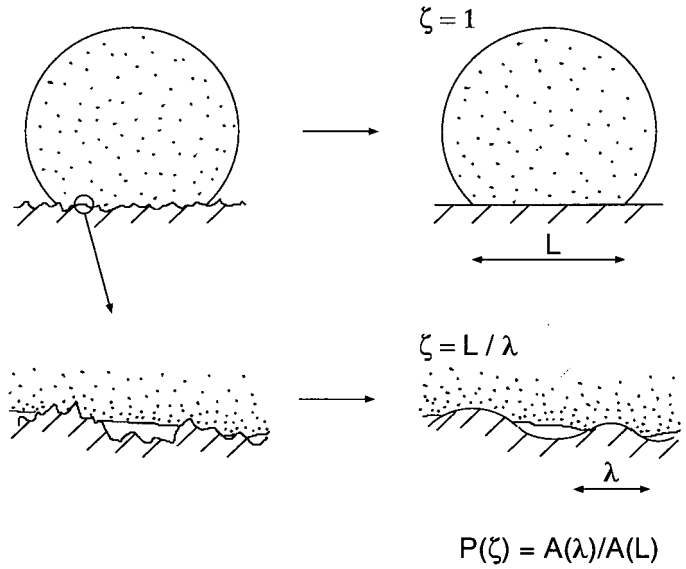

Figure 22. An elastic ball squeezed against a hard, rough, substrate. Left: the system at two different magnifications. Right: the area of contact $A(\lambda)$ on the length scale λ is defined as the area of real contact when the surface roughness on length scales shorter than λ has been removed.

reality, a short distance cut-off always exists, e.g., the interatomic distance. In many cases the local pressure in the contact regions at the asperities may become so high that the material yields plastically before reaching the atomic dimension. In these cases the size of the real contact area will be determined mainly by the yield stress of the solid.

4.1. Elastic contact mechanics

From contact mechanics (see, e.g., [18]) it is known that in the frictionless contact of elastic solids with rough surfaces the contact stresses depend only upon the shape of the gap between them before loading. Thus, without loss of generality, the actual system may then be replaced by a flat elastic surface (elastic modulus E and Poisson ratio ν, related to the original quantities via $(1-\nu^2)/E = (1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2$) in contact with a rigid body having a surface roughness profile which results in the same undeformed gap between the surfaces.

One of us (Persson) has recently developed a theory of contact mechanics [20], valid for randomly rough (e.g., self-affine fractal) surfaces. In the context of rubber friction, which motivated this theory, mainly elastic deformation occurs. However, the theory can also be applied when both elastic and plastic deformations occur in the contact areas. This case is of course relevant to almost all materials other than rubber.

The basic idea behind the new contact theory is that it is very important not to exclude *a priori* any roughness length scale from the analysis. Thus, if $A(\lambda)$ is the (apparent) area of contact on the length scale[9] λ (see figure 22), then we study the function $P(\zeta) = A(\lambda)/A(L)$ which is the relative fraction of the surface area where contact occurs on the length scale

[9] We define the apparent contact area $A(\lambda)$ on the length scale λ to be the area of real contact if the surface was smooth on all length scales shorter than λ. That is, considering the Fourier transform of the surface profile, all the components whose wavevector is larger that $2\pi/\lambda$ have to be set to 0, and the contact area with this new surface is by definition $A(\lambda)$.

$\lambda = L/\zeta$ (where $\zeta \geqslant 1$), with $P(1) = 1$. Here $A(L) = A_0$ denotes the macroscopic contact area (L is the diameter of the macroscopic contact area, so $A_0 \approx L^2$).

Consider the system at the length scale $\lambda = L/\zeta$, where L is the diameter of the nominal contact area. We define $q_L = 2\pi/L$ and write $q = q_L\zeta$. Let $P(\sigma, \zeta)$ denote the stress distribution in the contact areas under the magnification ζ. The function $P(\sigma, \zeta)$ satisfies the differential equation (see [20])

$$\frac{\partial P}{\partial \zeta} = f(\zeta)\frac{\partial^2 P}{\partial \sigma^2} \tag{5}$$

where $f(\zeta) = G'(\zeta)\sigma_0^2$, σ_0 being the average pressure in the nominal contact area, and

$$G(\zeta) = \frac{\pi}{4}\left(\frac{E^*}{\sigma_0}\right)^2 \int_{q_L}^{\zeta q_L} \mathrm{d}q\, q^3 C(q), \tag{6}$$

with $E^* = E/(1-\nu^2)$.

Equation (5) is a diffusion-type equation, where time is replaced by magnification ζ, and the spatial coordinate with the stress σ (and where the 'diffusion constant' depends on ζ). Hence, when we study $P(\sigma, \zeta)$ on shorter and shorter length scales (corresponding to increasing ζ), the function $P(\sigma, \zeta)$ will become broader and broader in σ-space. We can take into account that detachment will occur when the local stress reaches $\sigma = 0$ (we assume no adhesion) via the boundary condition [21]

$$P(0, \zeta) = 0. \tag{7}$$

In order to solve equation (5) we also need an 'initial' condition. This is determined by the pressure distribution at the lowest magnification $\zeta = 1$. If we assume a constant pressure σ_0 in the nominal contact area, then $P(\sigma, 1) = \delta(\sigma - \sigma_0)$.

We assume that only elastic deformation occurs (i.e., the yield stress $\sigma_Y \to \infty$). In this case

$$P(\zeta) = \int_0^\infty \mathrm{d}\sigma\, P(\sigma, \zeta).$$

When adhesion is taken into account, tensile stresses can occur at the interface between the two solids, and the boundary condition (7) is no longer valid [22]; see section 7.1. It is straightforward to solve (5) with the boundary conditions $P(0, \zeta) = 0$ and $P(\infty, \zeta) = 0$ to get

$$P(\zeta) = \frac{2}{\pi}\int_0^\infty \mathrm{d}x\, \frac{\sin x}{x} \mathrm{e}^{-x^2 G(\zeta)} = \mathrm{erf}\left(\frac{1}{2\sqrt{G}}\right). \tag{8}$$

Note that for small load σ_0, $G \gg 1$ and in this case (8) reduces to $P(\zeta) \approx P_1(\zeta)$ where

$$P_1(\zeta) = [\pi G(\zeta)]^{-1/2}. \tag{9}$$

Since $G \sim 1/\sigma_0^2$ it follows that the area of real contact is *proportional* to the load for small load. Using (8) and (9) we can write in a general case

$$P(\zeta) = \mathrm{erf}\left(\frac{\sqrt{\pi}}{2} P_1(\zeta)\right). \tag{10}$$

The physical meaning of the diffusion-like equation (5) is as follows. When the system is studied at the lowest magnification $\zeta = 1$ no surface roughness can be observed and the block makes (apparent) contact with the substrate everywhere in the nominal contact area. In this case, if we neglect friction at the interface, the stress at the interface will everywhere equal the applied stress σ_0—see figure 23(a)—so the distribution will initially be delta function

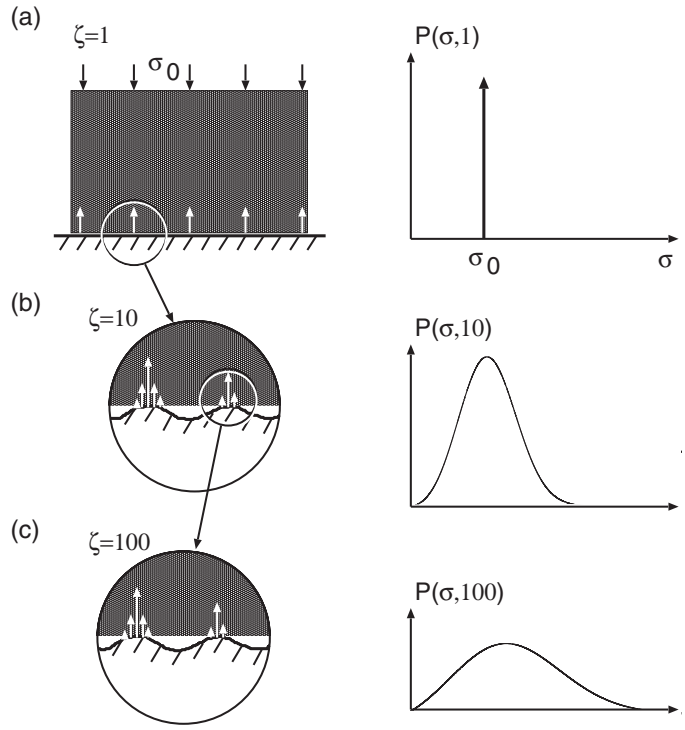

Figure 23. The stress distribution $P(\sigma, \zeta)$ in the contact region between a (rigid) block and an elastic substrate at increasing magnification ζ. At the lowest (engineering) magnification $\zeta = 1$ the substrate surface looks smooth and the block makes (apparent) contact with the substrate in the whole nominal contact area. As the magnification increases, we observe that the area of (apparent) contact decreases, while the stress distribution becomes broader and broader.

like, $P(\sigma, 1) = \delta(\sigma - \sigma_0)$. Increasing the magnification, we include surface roughness with wavelength down to $\lambda = L/\zeta$, and here one may observe some non-contact regions as shown in figure 23(b). Since the stress must go continuously to zero at the edges of the boundary between the contact and non-contact regions, it follows that the stress distribution $P(\sigma, \zeta)$ will have a tail extending the whole way down to zero stress as indicated in figure 23(b) (right). There will also be a tail toward larger stresses $\sigma > \sigma_0$ because the average stress must be equal to σ_0. This distribution broadens as in a diffusion problem. With increasing magnification, the stress distribution will broaden further and without limit as indicated in figure 23(c) (right).

The theory presented above predicts that the area of contact increases linearly with the load for small load. In the standard theory of Greenwood and Williamson [17] this result holds only approximately and a comparison of the prediction of their theory with the present theory is therefore difficult. Bush *et al* [19] have developed a more general and accurate contact theory. They assumed that the rough surface consists of a mean plane with hills and valleys randomly distributed on it. The summits of these hills are approximated by paraboloids, whose distributions of heights and principal curvatures are obtained from the random process theory. This is to be compared with the GW assumption that the caps of the asperities are spherical, each having the same mean radius of curvature. As a result of the more random nature of the surface, Bush *et al* found that at small load the area of contact depends linearly on the load

according to

$$A = \kappa \frac{F_N}{E^*} \left(\int d^2q \, q^2 C(q) \right)^{-1/2} \tag{11}$$

where F_N is the normal load, $E^* = E/(1-\nu^2)$ and $\kappa = (2\pi)^{1/2}$. This result is very similar to the prediction of the present theory where, for small load, from (6) and (9), A is again given by equation (11) but now with $\kappa = (8/\pi)^{1/2}$. Thus our contact area is a factor of $2/\pi$ smaller than the one predicted by the theory of Bush *et al*. Both the theory of Greenwood and Williamson and that of Bush *et al* assume that the asperity contact regions are independent. However, as discussed in [21], for real surfaces (which always have surface roughness on many different length scales) this will never be the case even at a very low nominal contact pressure. As we argued [21], this may be the origin of the difference of $2/\pi$ between our theory (which assumes roughness on many different length scales) and the result of Bush *et al*.

The predictions of the contact theories of Bush *et al* [19] and Persson [20] were compared to numerical calculations (see [21, 23]). Borri-Brunetto *et al* [24] studied the contact between self-affine fractal surfaces using an essentially exact numerical method. They found that the contact area is proportional to the squeezing force for small squeezing forces. Furthermore, it was found that the slope $\alpha(\zeta)$ of the line $A = \alpha(\zeta)F$ decreased with increasing magnification ζ. This is also predicted by the analytical theory (equation (11)). In fact, good agreement was found between theory and computer simulations for the change in the slope with magnification and its dependence on the fractal dimension D_f.

Hyun *et al* performed a finite-element analysis of contact between elastic self-affine surfaces. The simulations were done for a rough elastic surface contacting a perfectly rigid flat surface. The elastic solid was discretized into blocks and the surface nodes form a square grid. The contact algorithm identified all nodes on the top surface that attempt to penetrate the flat bottom surface. The total contact area A was obtained by multiplying the number of penetrating nodes by the area of each square associated with each node. As long as the squeezing force was so small that the contact area remained below 10% of the nominal contact area, $A/A_0 < 0.1$, the area of real contact was found to be proportional to the squeezing force in accordance with equation (11). Figure 24 shows the Hyun *et al* results for the factor κ in (11) as a function of Hurst's exponent H for self-affine fractal surfaces. The two horizontal lines are predictions of the theories of Bush *et al* (solid line) and Persson (dashed line). The agreement with the analytical predictions is quite good considering the ambiguities in the discretization of the surface. The algorithm only considers nodal heights and assumes that contact of a node implies contact over the entire corresponding square. This procedure would be accurate if the spacing between nodes were much smaller than the typical size of asperity contacts. However, the majority of the contact area consists of clusters containing only one or a few nodes. Since the number of large clusters grows as $H \to 1$, this may explain why the numerical results approach Persson's prediction in this limit.

Hyun *et al* also studied the distribution of connected contact regions and the contact morphology. In addition, the interfacial stress distribution was considered, and it was found that the stress distribution remained non-zero as the stress $\sigma \to 0$. This violates the boundary condition (7) that $P(\sigma, \zeta) = 0$ for $\sigma = 0$. However, it has been shown analytically [21] that for 'smooth' surface roughness this latter condition must be satisfied, and we believe that the violation of this boundary condition in the numerical simulations may reflect the way the solid was discretized and the way the contact area was defined in the Hyun *et al* numerical procedure.

Elastic contact theory and numerical simulations show that in the region where the contact area is proportional to the squeezing force, the stress distribution at the interface is independent

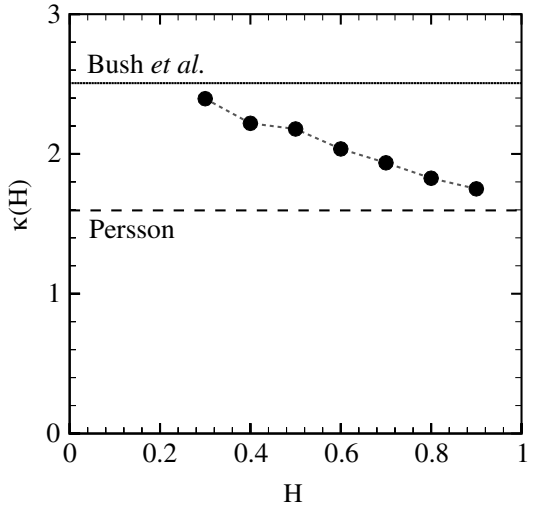

Figure 24. Dots: the factor κ as a function of Hurst's exponent H for self-affine fractal surfaces. The two horizontal lines are the predictions of the theories of Bush *et al* (solid line) and Persson (dashed line). From [23].

of the squeezing force. In addition, for an infinite system the distribution of sizes of the contact regions does not depend on the squeezing force (for small squeezing forces). Thus, when the squeezing force increases, new contact regions are formed in such a way that the distribution of contact regions and the pressure distribution remain unchanged. This is the physical origin of Coulomb's friction law which states that the friction force is proportional to the normal (or squeezing) force [13], and which usually holds accurately as long as the block–substrate adhesional interaction can be neglected [25].

4.2. Surface stiffness of fractal surfaces

The contact behaviour of realistic surfaces with random multiscale roughness remains largely unknown. Recently experimental results of the surface stiffness for self-affine fractal surfaces were presented [26]. A hard micrometric flat probe surface was squeezed against different substrates, and the vertical displacement u was measured as a function of the squeezing force F. It was found that the surface stiffness $S = \mathrm{d}F/\mathrm{d}u$ depends remarkably on the fractal dimension of the substrate surface, *decreasing by more than an order of magnitude as the substrate fractal dimension is increased by only* 10%. As expected, the surface stiffness was nearly proportional to the inverse of the rms roughness amplitude σ. Figure 25 shows the surface stiffness as a function of the squeezing force F for fractal substrate surfaces produced from two different materials (+ and $\times$ symbols, respectively). Note the strong decrease in the surface stiffness S with increasing fractal dimension D_f of the substrate.

4.3. Viscoelastic contact mechanics

The contact between a viscoelastic solid and hard, randomly rough, substrates is a topic of great practical importance, e.g., for pressure sensitive adhesives, rubber friction and rubber seals. When a viscoelastic solid is squeezed with a constant force against a rough substrate, the area of real contact will increase monotonically with the contact time; see figure 26. Since

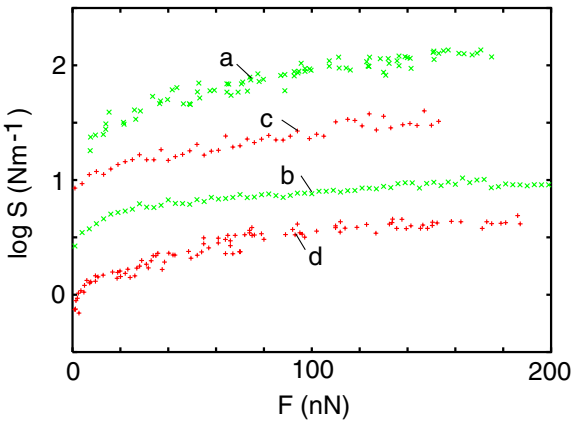

Figure 25. The surface stiffness $S = \mathrm{d}F/\mathrm{d}u$ as a function of the squeezing force F for a flat hard micrometre sized surface in contact with self-affine fractal substrates made from two different materials (+ and × symbols, respectively). The fractal dimension D_f and the rms roughness amplitude σ of the surfaces are: (a) $D_\mathrm{f} = 2.10$, $\sigma = 20$ nm; (b) 2.30, 80 nm; (c) 2.15, 36 nm and (d) 2.26, 160 nm. From [26].

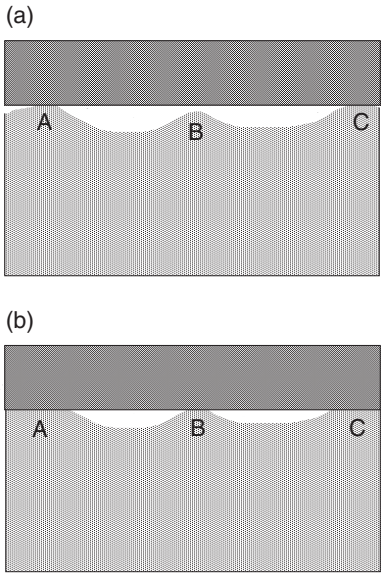

Figure 26. With increasing time, the contact area between the viscoelastic substrate and a flat hard solid surface increases through growth of existing asperity contact areas (A and C), as well as by the formation of new asperity contact areas (B).

rubber-like materials have a wide distribution of relaxation times, the area of real contact will usually increase over a very long time period (which, e.g., could be a year or more). Since the pull-off force depends on the area of real contact, contact theories for viscoelastic solids are important for estimating how the pull-off force (or tack) depends on the applied squeezing pressure and the squeezing time; see section 4.4.

Rubber-like materials have elastic modulus $E(\omega)$ which depends strongly on frequency ω and temperature T. Thus, at very low frequencies or high temperatures they behave as very soft *rubbery* materials, with elastic moduli typically in the range 0.01–1 MPa. At high frequencies or low temperatures they behave instead as hard glassy materials with elastic moduli of order 1 GPa or more. Thus, as a function of increasing frequency (or decreasing temperature) the

elastic modulus may increase by a factor of 1000 or more. The transition from the rubbery region to the glassy region is very wide, usually extending over more than three frequency decades. In a contact experiment, the inverse of the contact time is a characteristic frequency; thus for long contact times rubber behaves as a *soft* solid and the contact area is *large*, while for short contact times it is relatively *hard* and the contact area is *small*.

When a viscoelastic solid is squeezed with a constant force against a hard rough substrate, the contact area will increase continuously with time [27, 28]. The relative contact area $P(\zeta, t) = A(\zeta, t)/A_0$ is approximately given by equation (10):

$$P(\zeta, t) = \mathrm{erf}\left(\frac{\sqrt{\pi}}{2} P_1(\zeta, t)\right) \tag{12}$$

where [27]

$$P_1(\zeta, t) = \frac{2}{\pi Q(\zeta)} \int_{-\infty}^{\infty} \mathrm{d}\omega \, \frac{\tilde{\sigma}_0(\omega)}{E^*(\omega)} \mathrm{e}^{-\mathrm{i}\omega t}, \tag{13}$$

denotes the relative contact area to linear order in σ_0. Here $\tilde{\sigma}_0(\omega)$ is the Fourier transform of the squeezing pressure $\sigma_0(t)$ and

$$Q(\zeta) = \left(\int_0^{q_L \zeta} \mathrm{d}q \, q^3 C(q)\right)^{1/2}.$$

Now, let us assume that $\sigma_0(t) = \sigma_0$ for $0 < t < t_1$ and zero otherwise. In this case

$$\mathrm{i}\tilde{\sigma}_0(\omega) = \frac{\sigma_0/2\pi}{0^+ - \mathrm{i}\omega} \left(1 - \mathrm{e}^{\mathrm{i}\omega t_1}\right). \tag{14}$$

Substituting this into (13) gives

$$\begin{aligned} P_1(\zeta, t) &= \frac{2\sigma_0}{\pi Q(\zeta)} \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \, \frac{1 - \mathrm{e}^{\mathrm{i}\omega t_1}}{-\mathrm{i}\omega} \frac{\mathrm{e}^{-\mathrm{i}\omega t}}{E^*(\omega)} \\ &= \frac{2\sigma_0}{\pi^2 Q(\zeta)} \mathrm{Re} \int_0^{\infty} \mathrm{d}\omega \, \frac{1 - \mathrm{e}^{\mathrm{i}\omega t_1}}{-\mathrm{i}\omega} \frac{\mathrm{e}^{-\mathrm{i}\omega t}}{E^*(\omega)}. \end{aligned} \tag{15}$$

This result does not, of course, depend on t_1 as long as $t < t_1$ (causality). Thus we are free to choose for t_1 any value larger than the time t under consideration. The equations presented above are only valid as long as the contact area increases with time, which is the case, e.g., if a constant squeezing force is applied at time $t = 0$. As an illustration, in figure 27 we show the relative contact area calculated from equation (15) using the measured viscoelastic modulus for two types of rubber. The rubber block is squeezed with the nominal pressure $\sigma_0 = 0.1$ MPa against a 'steel' surface. We show results for two typical rubbers used in the construction of tyres, namely a tread rubber compound and a rubber compound from the tyre rim region involved in the tyre air sealing (see section 5). The substrate was a self-affine fractal with the fractal dimension $D_\mathrm{f} = 2.2$ and with the rms roughness amplitude 1 μm. The tread compound exhibits the largest contact area because it is elastically softer than the tyre rim compound. Note also that the change in the contact area is only a factor ∼2–4 when the contact time changes by ∼9 decades.

4.4. Tack

Pressure sensitive adhesives (PSA) are used in many important applications, e.g., for Scotch tapes, Post-it pads and self-adhesive labels and envelopes [29, 30]. The adhesive consists of a very thin layer (usually of order ∼20–100 μm) of a very soft, weakly cross-linked rubber compound. The low frequency elastic modulus is typically only 0.01 MPa which is ∼100 times

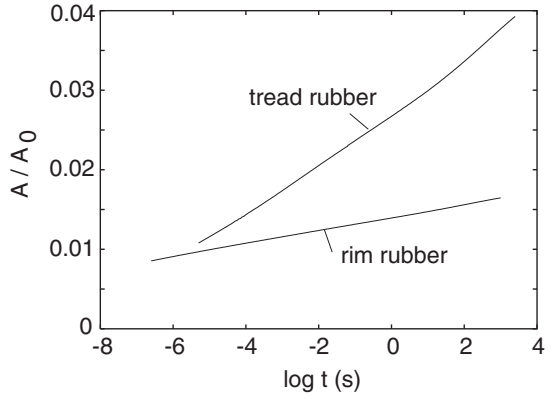

Figure 27. The relative contact area (obtained from equations (12) and (15)) at the magnification $\zeta = 10^5$, as a function of the logarithm of the contact time t, for the nominal pressures $\sigma_0 = 0.1$ MPa, for a tread tyre rubber compound and for a compound used in the tyre–rim area, at the temperature $T = 60\,^{\circ}$C. The substrate is self-affine fractal with the fractal dimension $D_f = 2.2$, with the rms roughness 1 μm and with the roll-off wavevector (see figure 6) $q_0 = 6 \times 10^4$ m^{-1} (this is a typical q_0 for a polished steel surface). The magnification $\zeta = 1$ refers to the length scale $\lambda_0 = 2\pi/q_0 \approx 0.1$ mm, so the magnification $\zeta = 10^5$ corresponds to the length scale (resolution) $\lambda \approx 1$ nm.

lower than that of the rubber used for tyres. As a result of the low elastic modulus, nearly complete contact will occur in the apparent contact area even for relatively low squeezing pressures and large surface roughness. This is in contrast to the case for tyre rubber which under similar condition would give a contact area of only a few per cent of the nominal contact area, as discussed in section 6.1.

We carried out calculations (using equation (15)) of the time-dependent contact area for a standard rubber tack compound (acrylic PSA, i.e. poly(2-ethylhexyl acrylate), with 2% acrylic acid (PEHA-AA)).

Figure 28 shows the results at $T = 20\,^{\circ}$C when the PSA is squeezed in contact with 'steel' surfaces with rms roughness amplitudes $h_0 = 0.25$, 1 and 4 μm. Figure 28(a) shows the logarithm of the relative contact area, as a function of the logarithm of the contact time, for a contact pressure $\sigma_0 = 0.01$ MPa. Figure 28(b) shows the logarithm of the relative contact area after 1 s of contact, as a function of the logarithm of the contact pressure.

It is interesting to compare the results in figure 28 with the experimental data reported in [31] and shown in figure 29. The data show the dependence of the maximum pull-off force F on the contact time and contact pressure for smooth and rough PSA on the same smooth steel surface. The tack film is a standard polymer compound similar to PEHA. In [31] no numerical values of the rms roughness were presented for any of the surfaces studied. If one assumes, as a first approximation, that the pull-off force is proportional to the area of real contact, and moreover that the smooth and rough PSA films in the measurements correspond, say, to the surfaces in figure 28 with the rms roughnesses 0.25, 1 and 4 μm, then the agreement between theory and the experiment is remarkably good.

5. Seals

Surface roughness is an important factor which influences the rate of leakage through seals. The exact mechanism of roughness induced leakage is not well understood [32]. In this section we present a new way of looking at this problem [27].

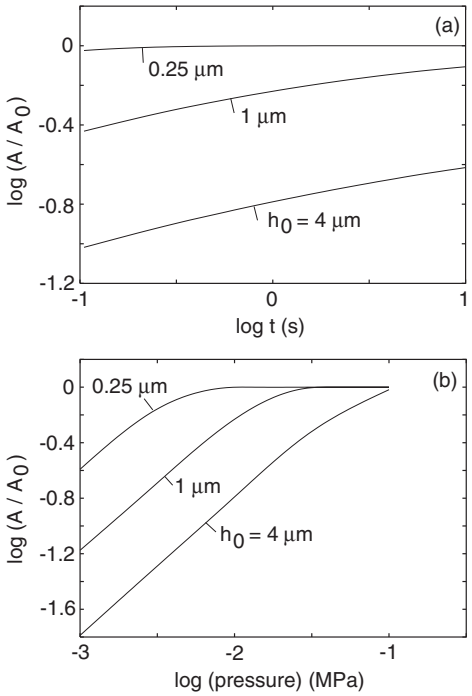

Figure 28. (a) The calculated logarithm of the relative contact area at the magnification $\zeta = 10^5$, as a function of the logarithm of the contact time. The contact pressure $\sigma_0 = 0.01$ MPa. (b) The calculated logarithm of the relative contact area after 1 s of contact at the magnification $\zeta = 10^5$, as a function of the logarithm of the contact pressure. All calculations for PEHA-AA were at temperatures $T = 20\,^{\circ}$C on a self-affine fractal substrate with fractal dimension $D_f = 2.2$, roll-off wavevector $q_0 = 6 \times 10^4$ m^{-1} and rms roughness amplitudes $h_0 = 0.25$, 1 and 4 μm.

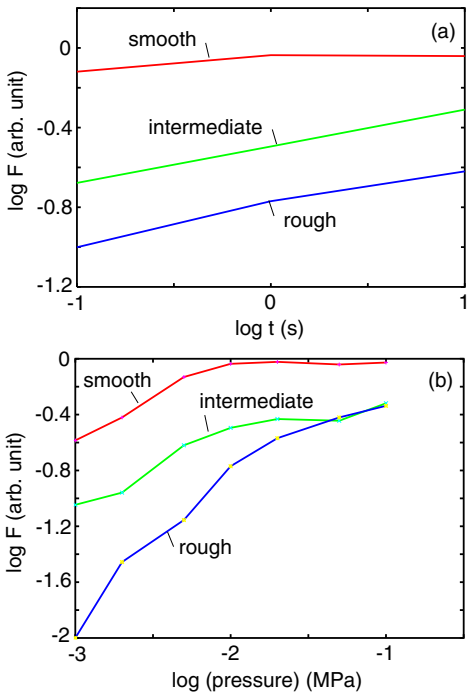

Figure 29. (a) The logarithm of the experimental pull-off force F (in arbitrary units) as a function of the logarithm of the contact time. The contact pressure $\sigma_0 = 0.01$ MPa. From [31]. (b) The logarithm of the experimental pull-off force (with the same units as in (a)) after 1 s of contact, as a function of the logarithm of the contact pressure, for three different surface roughnesses. From [31].

Viscoelastic materials such as rubber are often used for sealing. Here we consider the tyre–rim sealing. We are interested in the air (or gas) flow from inside the tyre to the outside via the roughness induced non-contact area (pore channels) in the rubber–steel rim area. The rim is made from steel. We assume that the steel surface is a self-affine fractal with the fractal dimension $D_f = 2.2$. The surface root mean square roughness is assumed to be 1 μm, a fairly typical value for polished steel surfaces.

Figure 30 shows the calculated (using equation (15)) relative contact area after one second of contact for tread tyre rubber (a) and for rim tyre rubber (b), and the root mean square roughness (c), as a function of the magnification (the time dependence of the contact area was presented in figure 27). Here we defined

$$h_{\rm rms}(\zeta) = \left(2\pi \int_{\zeta q_0}^{q_1} {\rm d}q\, q C(q)\right)^{1/2}.$$

The results are for nominal pressures $\sigma_0 = 0.1, 0.2$ and 0.4 MPa, at the temperature $T = 60\,^{\circ}$C.

We consider now a rubber–steel interface with increasing magnification. At the lowest magnification $\zeta = 1$ complete contact occurs at the interface; see figure 31 (left). When we

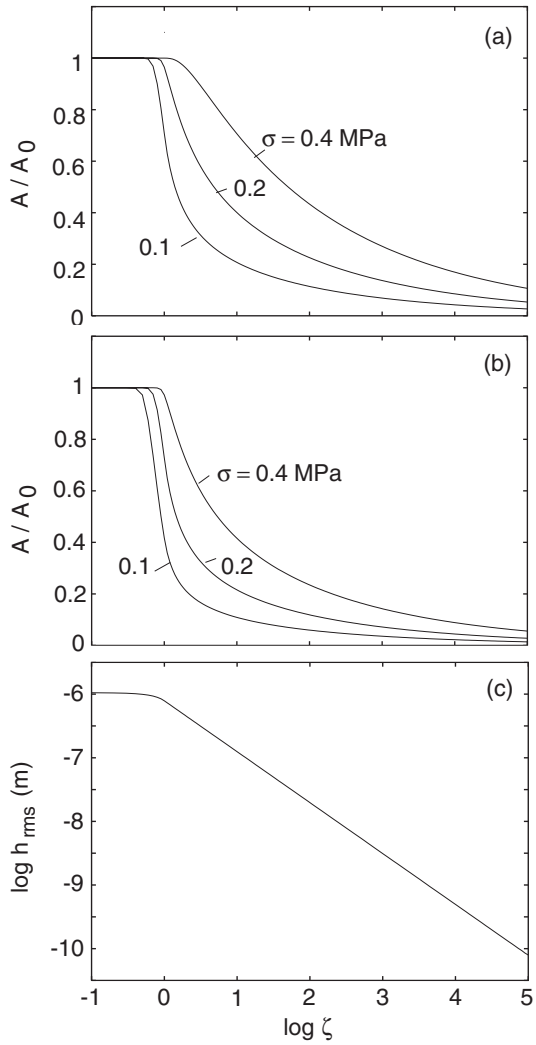

Figure 30. The calculated relative contact area after one second of contact for tread tyre rubber (a) and for rim tyre rubber (b), and the root mean square roughness as a function of magnification (c), where $\zeta = 1$ corresponds to a wavelength $\lambda_0 = 2\pi/q_0 \approx 100\ \mu$m. The highest magnification $\zeta = 10^5$ corresponds to $\lambda \approx 1$ nm. Results are shown for nominal pressures $\sigma_0 = 0.1$, 0.2 and 0.4 MPa, at the temperature $T = 60\,^{\circ}$C, for a substrate surface with rms roughness $h_0 = 1\ \mu$m and roll-off wavevector $q_0 = 6 \times 10^4\ \mathrm{m}^{-1}$.

increase the magnification we observe non-contact areas or islands. Magnification is now increased until the non-contact areas percolate, i.e., until a channel of non-contact surface area, extending from the high pressure internal region of the tyre to the outside (atmospheric pressure region), is first observed. As the magnification is increased further, a new non-contact region will be observed, but the separation between the surfaces in these new non-contact areas will be smaller than along the percolation channel. Since the gas flow $\dot{N}$ (number of molecules per unit time) through a rectangular pore of height h depends as $\dot{N} \sim h^3$ we will assume that most gas will leak through the percolation channel.

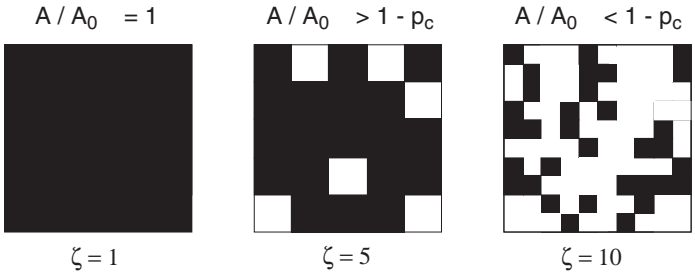

Figure 31. When the interface between the solids is studied at low magnification, there appears to be complete contact (black area) between the solids. When the magnification is increased it is observed that only partial contact occurs. At high enough magnification the non-contact (white) surface area will percolate and one (or several) air flow channels will be visible at the interface.

Assuming that the contact regions at any magnification are approximately randomly distributed in the apparent contact region, we expect from percolation theory that the non-contact region will percolate when $A/A_0 \approx 1-p_c$, where p_c is the site percolation number [33]. For a hexagonal lattice, which is the most plausible lattice structure in the present case, one has $p_c \approx 0.7$, while for a square lattice (as in figure 31) percolation occurs at $p_c \approx 0.6$. Thus, the exact value of the percolation threshold does not depend sensitively on the symmetry of the unit cell.

We assume that the main gas leakage comes from gas flow through the percolation channel. The narrowest passage in this channel can be considered as a rectangular pore of height h, and of width and length λ, where λ is determined by the magnification ζ_c at the point where $A/A_0 = 1 - p_c \approx 0.3$. The height h of the pore is (approximately) determined by the rms roughness at the magnification ζ_c. In the present case, if the tyre gas pressure is in the range 0.2–0.3 MPa, from figure 30(b) we get $\zeta_c \approx 10$ and from figure 30(c), $h \approx 0.1$ μm.

We divide the tyre–rim contact area into m square areas, $B \times B$, where B is the width of the tyre–rim contact area (we expect B to be of order a few centimetres). The number of squares is $m = 2\pi R/B$, where R is the radius of the tyre at the rim. We expect $m \approx 100$.

Let us study the gas flow through a rectangular pore of height h and width and length (in the flow direction) λ. We assume stationary and laminar flow, and that $h \ll \lambda$. In this case, if $N_1(t)$ denotes the number of gas phase molecules in the tyre (which is proportional to the pressure P_1 in the tyre), the basic equations of hydrodynamics give, for the typical case $P_1 \gg P_0$,

$$\dot{N}_1 \approx -\frac{m P_1^2 h^3}{24 \mu k_B T}.$$

Here we have implicitly assumed that the full pressure drop $P_1 - P_0$ occurs over the pore. Thus, the time it takes for the pressure in the tyre to drop with $\approx$4% is

$$\Delta t = \frac{\mu}{P_1}\frac{V_1}{h^3}$$

where $V_1 = V_0/m$ (where V_0 is the air volume in the tyre) is the volume of air in an angular section of the tyre of width B. With $B \approx 3$ cm we get $V_1 \approx 3 \times 10^{-4}$ m^3 and using the viscosity of air $\mu \approx 17 \times 10^{-6}$ N s m^{-2} gives $\Delta t \approx 1$ year. This is an upper limit of the leakage time, since when the interfacial contact area is studied at higher magnification new pore channels through which the air can leak will be detected.

It is interesting to note that the adhesional interaction between the rubber surface and the steel rim is likely to have negligible influence on the leakage rate. Adhesion will affect the

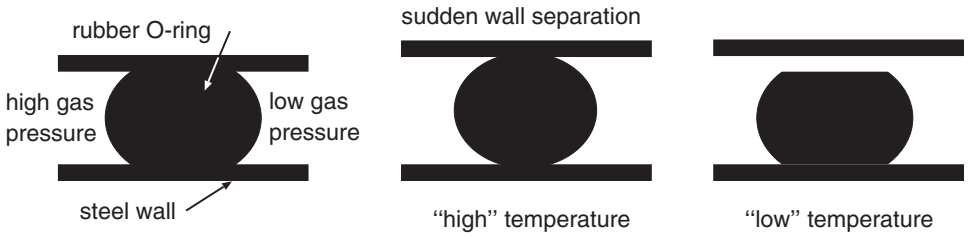

Figure 32. A rubber O-ring seal. Left: the rubber ring is compressed between two solid walls. Middle: if the rubber temperature is high enough, when the separation between the walls is suddenly increased the rubber will quickly expand and still provide tight sealing. Right: at low temperature the rubber relaxation times are very long and the rubber will not be able to expand quickly enough. Here the sealing fails.

(apparent) contact area only at very high magnification [27], but most of the gas leakage occurs via the much larger air flow channels which can be observed at low magnification.

When a rubber block is squeezed with a constant force against a rough substrate, the area of real contact will increase continuously, and the (average) space between the surfaces will decrease continuously with increasing time. This is due to stress relaxation in the rubber and was illustrated in figure 27. In some applications, such as tyre air sealing, this relaxation effect is clearly beneficial as it will reduce the size of the interfacial air flow channels. However, in other applications, such as O-ring seals, stress relaxation may result in catastrophic events. One prominent example was that of the Challenger catastrophe [34]. In this case rubber O-rings were used to seal the hot gas inside the solid fuel rockets and prevent it from leaking through the horizontal joints used to hold the different vertical rocket wall segments together. O-rings normally operate with about $\sim$15% compression to ensure a tight seal; see figure 32 (left). However, the high pressure inside the rocket resulted in expansion of the steel tube, and led to a (small) separation of the wall surfaces at the joints. At high enough rubber temperature this is not a problem since the compressed rubber O-ring would then quickly expand and seal tightly even if the space between the walls increased; see figure 32 (middle). However, the temperature at the time of the Challenger launch was exceptionally low (around 0 °C), which resulted in very long rubber relaxation times, so the compressed form of the O-ring seals was 'frozen in'; see figure 32 (right). This effect was demonstrated by Feynman in a famous experiment where part of an O-ring was kept in a deformed state in a glass with ice water. Thus, as the space between the surfaces increased, one or several O-rings failed to seal off the hot gas, resulting in leakage of hot fuel gas through the rocket joints, and finally the catastrophic failure of the rocket.

Stress relaxation is an important aspects of O-ring seals even when no separation of the solid walls occurs at the seal [35]. When a seal is under constant compression (fixed solid walls), the initial stress decays with increasing time, roughly proportional to the logarithm of the time of contact. Thus the peak compressive stress may eventually drop below the system (gas or liquid) pressure, and the seal leaks. Clearly, stress relaxation effects must be taken into account when determining the type of O-ring to be used in a particular application.

6. Rubber friction

First-principles calculations of frictional forces for realistic systems are generally impossible. The reason is that friction is usually an interfacial property, often determined by the last few uncontrolled monolayers of atoms or molecules at the interface. An extreme illustration of this

is offered by diamond. The friction between two clean diamond surfaces in ultrahigh vacuum is huge because of the strong interaction between the surface dangling bonds. However, when the dangling bonds are saturated by a hydrogen monolayer (as they generally are in real life conditions), friction becomes extremely low [36]. Since most surfaces of practical use are covered by several monolayers of contaminant molecules of unknown composition, a quantitative prediction of sliding friction coefficients is out of the question. An exception to this may be rubber friction on rough surfaces, which is the subject that we address here.

Rubber friction on smooth substrates, e.g., on smooth glass surfaces, has two contributions, namely an adhesive (surface) and a hysteretic (bulk) contribution [37, 38]. The adhesive contribution results from the attractive binding forces between the rubber surface and the substrate. Surface forces are often dominated by weak attractive van der Waals interactions. For very smooth substrates, because of the low elastic moduli of rubber-like materials, even when the applied squeezing force is very gentle this weak attraction may result in a nearly complete contact at the interface [22, 39], leading to the large sliding friction force usually observed [40]. For rough surfaces, on the other hand, the adhesive contribution to rubber friction will be much smaller because of the small contact area. The actual contact area between a tyre and the road surface, for example, is typically only $\sim$1% of the nominal footprint contact area [20, 41, 42]. Under these conditions the bulk (hysteretic) friction mechanism is believed to prevail [20, 42]. For example, the exquisite sensitivity of tyre–road friction to temperature just reflects the strong temperature dependence of the viscoelastic bulk properties of rubber.

Here we discuss how rubber friction depends on the surface roughness power spectra, and we consider the influence of wear induced polishing and of water on the road track on rubber friction.

6.1. Basic theory of rubber friction

The main contribution to rubber friction when a rubber block is sliding on a rough substrate, such as in the case of a tyre on a road surface, is due to the viscoelastic energy dissipation in the surface region of the rubber as a result of the pulsating forces acting on the rubber surface from the substrate asperities; see figure 33. Recently one of us developed a theory which accurately describes this energy dissipation process, and which predicts the velocity dependence (and, in a more general case, the time history dependence) of the rubber friction coefficient [20, 41]. The results depend only on the (complex) viscoelastic modulus $E(\omega)$ of the rubber, and on the substrate surface roughness power spectra $C(q)$. Neglecting the flash temperature effect (the term *flash temperature* refers to a local and sharp temperature rise occurring in the tyre–road asperity contact regions during slip), the kinetic friction coefficient at velocity v is determined by [20]

$$\mu_{\rm k} = \frac{1}{2}\int_{q_0^*}^{q_1^*} {\rm d}q\, q^3 C(q) P(q) \int_0^{2\pi} {\rm d}\phi\, \cos\phi\, {\rm Im}\, \frac{E(qv\,\cos\phi)}{(1-\nu^2)\sigma},$$

where

$$P(q) = {\rm erf}\left(\frac{1}{2\sqrt{G}}\right),$$

with

$$G(q) = \frac{1}{8}\int_{q_0^*}^{q} {\rm d}q\, q^3 C(q) \int_0^{2\pi} {\rm d}\phi \left|\frac{E(qv\,\cos\phi)}{(1-\nu^2)\sigma}\right|^2,$$

where σ is the mean perpendicular pressure (the load divided by the nominal contact area), and ν is the Poisson ratio which equals 0.5 for rubber-like materials.

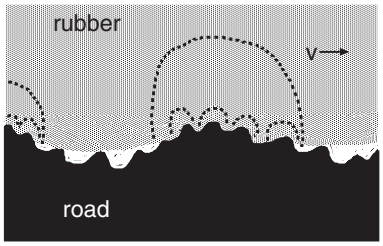

Figure 33. Hysteretic friction of a rubber block on a rough road surface. The road asperities exert pulsating forces on the sliding rubber block, leading to energy dissipation in the rubber via the rubber internal friction. Most of the energy dissipation occurs in the volume elements bounded by the dashed curves. The rubber viscoelastic deformations in the large volume elements are induced by the large road asperities, while the smaller dissipative regions result from the smaller asperities distributed on top of the large asperities. In general, in calculating the rubber friction, the viscoelastic energy dissipation induced by all the asperity sizes must be included, and the local temperature increase (flash temperature) in the rubber resulting from the energy dissipation should also be taken into account in the analysis.

The theory takes into account the substrate roughness in the range $q_0^* < q < q_1^*$, where q_0^* is the smallest relevant wavevector of order $2\pi/L$, where (in the case of a tyre) L is the lateral size of a tread block, and where q_1^* may have different origins (see below). Since q_0^* for a tyre tread block is smaller than the roll-off wavevector q_0 of the power spectra of most road surfaces (see figure 15), rubber friction is very insensitive to the exact value of q_0^*.

The large wavevector cut-off q_1^* may be related to road contamination, or may be an intrinsic property of the tyre rubber. For example, if the road surface is covered by small contamination particles (diameter D), then $q_1^* \approx 2\pi/D$. In this case, the physical picture is that when the tyre rubber surface is covered by hard particles of linear size D, the rubber will not be able to penetrate into surface roughness 'cavities' with diameter (or wavelength) smaller than D, and such short range roughness will therefore not contribute to the rubber friction. For perfectly clean road surfaces we believe instead that the cut-off q_1^* is related to the tyre rubber properties. Thus, the high local (flash) temperatures during braking, and the high local stresses which occur in the tyre rubber–road asperity contact regions, may result in a thin (typically of order a few micrometres) surface layer of rubber with modified properties (a 'dead' layer), which would contribute very little to the observed rubber friction. Since the stresses and temperatures which develop in the asperity contact regions depend somewhat on the type of road (via the surface roughness power spectra $C(q)$), the thickness of this 'dead' layer may vary from one road surface to another, and some run-in time period will be necessary for a new 'dead' layer to form when a car switches from one road surface to another. Such 'run-in' effects are well known experimentally.

In the theory that one of us has developed, the thickness of the dead layer is determined by studying (via computer simulations) the temperatures and stresses which develop on the surfaces of the tyre tread blocks during ABS breaking.

Since this is not the proper place for a full presentation of all details of the theoretical calculations of tyre–road friction, we shall simply present some numerical results to illustrate how the rubber friction depends on the surface roughness power spectra. Figure 34(a) shows the kinetic friction coefficient as a function of the sliding velocity calculated for a rubber block sliding on a self-affine fractal surface with the Hurst exponents $H = 0.8$, roll-off wavevector $q_0 = 1000\ \mathrm{m}^{-1}$ and large wavevector cut-off $q_1^* = 320.0 \times q_0$, typical for road surfaces. We show results both with and without the flash temperature effect.

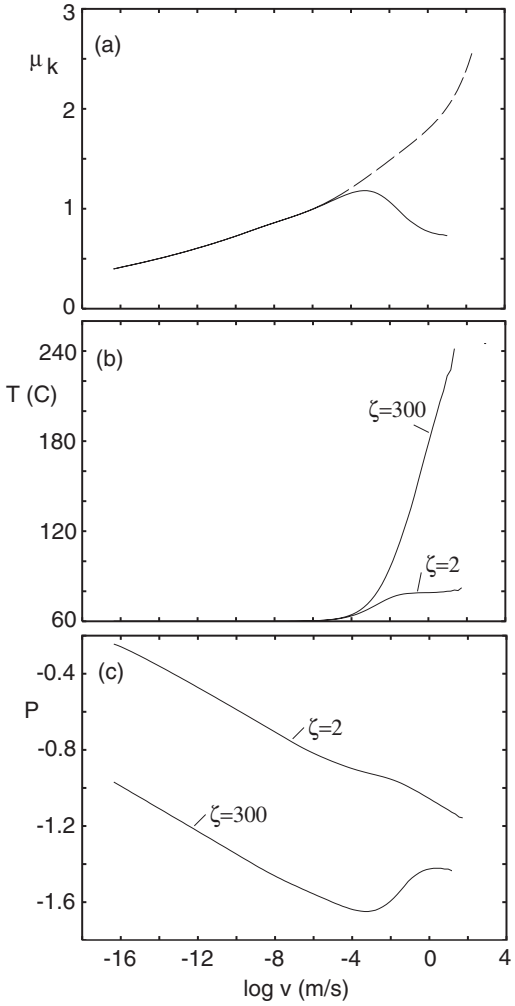

Figure 34. Calculated friction coefficients for a rubber block sliding on a self-affine fractal surface with $H = 0.8$, roll-off wavevector $q_0 = 1000\ \mathrm{m}^{-1}$, largest wavevector $q_1^* = q_0 \times 320.0$ (corresponding to the magnification $\zeta = 320$), rms roughness $\sigma = 1$ mm; the background temperature assumed was $T_0 = 60\,^{\circ}\mathrm{C}$. (a) The kinetic friction coefficient as a function of the logarithm of the sliding velocity. The solid and dashed curves are with and without flash temperature effects, respectively. (b) The flash temperature calculated at two different magnifications. The highest magnification $\zeta = 300$ corresponds to a temperature very close to the surface while the low magnification curve corresponds to a temperature deeper into the rubber. (c) The relative contact area $P(\zeta) = A(\zeta)/A_0$ calculated for two different magnifications.

For low sliding velocities, the kinetic sliding friction depends very weakly on the sliding velocity, and only at extremely low velocities is friction strongly reduced. Thus the present calculation (as well as experiments) shows that even sliding velocities as small as 10^{-10} m s^{-1} may give rise to a relative large kinetic friction. The physical reason for this is the very wide distribution of relaxation times in most rubber materials, extending to very long times.

For typical tyre rubber compounds, calculations (neglecting temperature effects) as well as measurements at low sliding velocities and different temperatures (subsequently shifted to

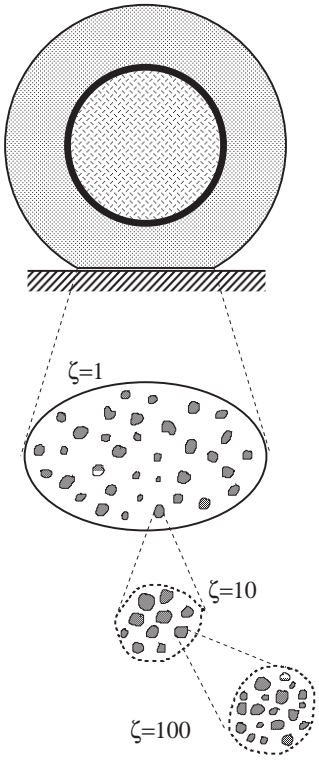

Figure 35. A tyre in contact with a road surface does not make perfect contact with the road over the whole footprint area. At the lowest magnification the asperity contact regions have diameters of order ∼0.3–1 cm. When the magnification is increased these macrocontact regions break up into smaller contact regions separated by non-contact areas. At a magnification of order ∼100–1000 the tyre rubber–road contact area is just a few per cent of the nominal footprint contact area.

a common temperature utilizing the frequency–temperature Williams–Landel–Ferry (WLF) transform [43]) indicate that the maximum friction coefficient $\mu_k(v)$ will typically occur at very high velocities, of order 1000 m s^{-1} (see the dashed curve in figure 34(a)). However, direct experiments without the WLF transform show a maximum in μ_k at much lower velocities. This is a result of the flash temperature; see the solid curve in figure 34(a). Thus when the flash temperature effect is taken into account the maximum of $\mu_k(v)$ shifts to much lower sliding velocities in the typical range 0.1–1 cm s^{-1}, in agreement with experiment.

Figures 34(b) and (c) show the flash temperature and the relative contact area at two different magnifications. For velocities $v < 10^{-4}$ m s^{-1} the contact area decreases with increasing sliding velocity. This is a result of the increasing frequencies of the deformations of the tyre surface with increasing sliding velocity. However, for $v > 10^{-3}$ m s^{-1} the contact area *increases*. This is a result of the increased temperature of the rubber in the contact areas shifting the viscoelastic dissipation maximum to higher frequencies and hence making the rubber elastically softer at any given frequency. It is also of interest to consider the relative contact area. For sliding velocities of order 1 cm s^{-1} the tyre–road contact area at a magnification of order 100–1000 is just a few per cent of the nominal footprint contact area. This is illustrated in figure 35 which shows the tyre–road contact region at different magnifications.

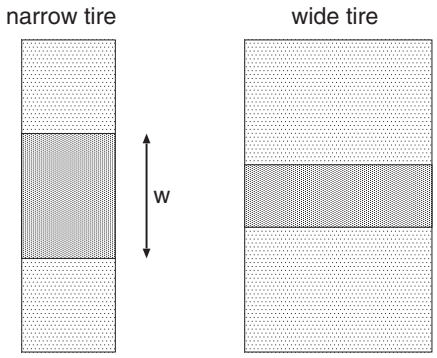

Figure 36. The nominal tyre–road contact area is approximately the same for a wide and a narrow tyre. Hence the length w of the footprint area in the rolling direction will be smaller for the wide tyre.

The flash temperature effect described above is of extreme practical importance. If the excessive build-up of flash temperatures could be avoided or reduced, the effective tyre–road friction coefficient would clearly increase. One way to reduce the flash temperature is to use wide tyres instead of narrow tyres[10]. This can be understood as follows. In a first approximation the nominal tyre–road contact area is the same for narrow and wide tyres. In fact, this holds exactly when a tyre can be described as a thin elastic (torus shaped) membrane (which is a good approximation for airplane tyres but less so for car tyres) and the tyre air pressure and the load are the same for both tyres[11]. Thus, wide tyres will have a footprint area which is shorter in the longitudinal rolling direction than a narrow tyre; see figure 36. In order to fully build up the flash temperature, a tread block must slide a distance of order the average diameter D of the macroasperity contact regions. The macroasperity contact regions are the contact between the rubber and the largest road asperities involving the longest wavelength roughness components characterized by the roll-off wavelength $\lambda_0 = 2\pi/q_0$ in the surface roughness power spectra (see figure 15). In a typical case $\lambda_0 \approx 1$ cm, and the average diameter of the macroasperity contact regions $D \approx 0.5$ cm. If a tread block in the footprint contact region slips by less than the distance D, the local temperature increase will be below that expected during stationary sliding at the same slip velocity. If v denotes the average slip velocity of a tread block then the slip distance $d \approx vt$, where the time the tread block spends in the footprint area $t = w/v_R$, v_R being the tyre rolling velocity and w the length of the footprint area. Thus, $d \approx (v/v_R)w$ and the slip velocity necessary for the slip to be of order D is $v \approx (D/w)v_R$. Hence, for a wide tyre (small w) the tread block slip velocity can reach higher values than for a narrow tyre without building up the flash temperature effect fully. Since the kinetic friction coefficient in the absence of the flash temperature effect increases monotonically with increasing slip velocities (up to very high slip velocities, perhaps 1000 m s^{-1} or more; see the dashed curve in

[10] The discovery that wide tyres give higher effective friction than narrow tyres was made by Jim Hall and co-workers in the middle 1960s for racing tyres. The story is beautifully described in [44], where Hall stated: 'in a matter of less than a year we went from those narrow little Dunlop Green Sports to a tyre that was almost double the width. And our single-speed torque converter just could not handle all that grip. We immediately went to a two-speed transmission so we could get more torque multiplication off the slow speed corners. So that's what happened with me on tyres. In the middle 1960s, probably 1964 or 1965, Firestone took us from less than 6 inches of tread to width up to 12 inches'. Paul Haney gives in his book an explanation for why wide tyres give higher effective friction than narrow tyres. However, this 'explanation', like many others given on the Internet, is incorrect.

[11] For the membrane tyre the nominal pressure in the footprint contact area is given by the air pressure P in the tyre, so the nominal contact areas $A_0 = F_N/P$ are the same for both tyres if the loads F_N are the same.

figure 34) it follows that wide tyres will give higher effective friction than narrow tyres. The discussion above is only of semi-quantitative validity since in reality the tread blocks will not slip with a constant velocity in the footprint area, but with a non-uniform velocity.

Another way to reduce the flash temperature could be to use a filler material which exhibits a first-order phase transition somewhere close to (but above) the background tyre–rubber temperature, which typically may be around 60 °C. If the phase transition (e.g., a structural, magnetic or ferroelectric transition) could absorb a sufficiently large latent heat, it would strongly reduce the flash temperature. The filler materials used today (mainly carbon black and silica) do not exhibit any such phase transition.

The theoretical results presented above are in accordance with experimental observations [45–48]. Thus, flash temperature effects have been observed during dry ABS braking tests, where the temperature distribution on the tyre surface can be measured with high speed IR cameras [45, 49]. In one experiment the temperature increase in the hot spots on the tread block surfaces on the length scale of the measurement was $\sim$25 °C above the background temperature. The decrease of the kinetic friction force with increasing sliding velocity for $v > 0.1$–1 cm s^{-1} (as a result of the flash temperature) has been reported by several research groups [45–48] and is in good agreement with the theory. A decreasing kinetic friction coefficient may result in stick–slip motion; thus one mechanism of rubber stick–slip is due to the flash temperature.

The theory also predicts that when a rubber block is sliding on a rough substrate the friction force is nearly proportional to the normal load, i.e., the friction coefficient is independent of the load. This is again in agreement with experiment; see, e.g., [46]. We note that for tyres the effective friction coefficient usually decreases with increasing load. This does not actually reflect a load dependence of the fundamental rubber–substrate friction coefficient, but results rather as an effect of the dependence of the size and the pressure distribution in the footprint area on the load, which affects the motion of the tread blocks [50].

The theory sketched above has been used successfully to predict the ABS breaking performance of tyres [45, 48, 49, 51]. In one study six tyres were produced with variation only in the tread compound [45, 49]. The tread rubber compounds were varied so big variations in ABS rating were expected. The observed braking distance was found to correlate very well with the theoretical predicted effective friction coefficient, with a high correlation coefficient $R^2 = 0.97$. This result shows that it is possible to estimate the performance of a tyre using physically sound frictional models which focus not only on the viscoelastic properties of the rubber compound, but also on taking explicitly into account additional important information such as the surface topography at all length scales as well as ambient and flash temperature conditions.

6.2. Rubber friction and the influence of polishing

Let us now discuss the role of road polishing by sliding tyres[12]. It is known that a relative new asphalt roadway has a 'sharper' surface with a higher coefficient of friction compared to an older well travelled asphalt roadway. This is a consequence of the wear action exerted by vehicle tyres, which results in polishing the exposing aggregate surface, and the state of polish is one of the main factors affecting the resistance to skidding. (Note: the fact that the polishing of the surface asperities *reduces* the friction is in itself an indication that the adhesional contribution to rubber friction on road surfaces is unimportant, since the latter should increase when the surface becomes smoother.) Resistance to this polishing action is

[12] See, e.g., http://www.betterroads.com/articles/oct03b.htm for experimental information from road engineers about rubber friction and road surfaces, in particular about road polishing.

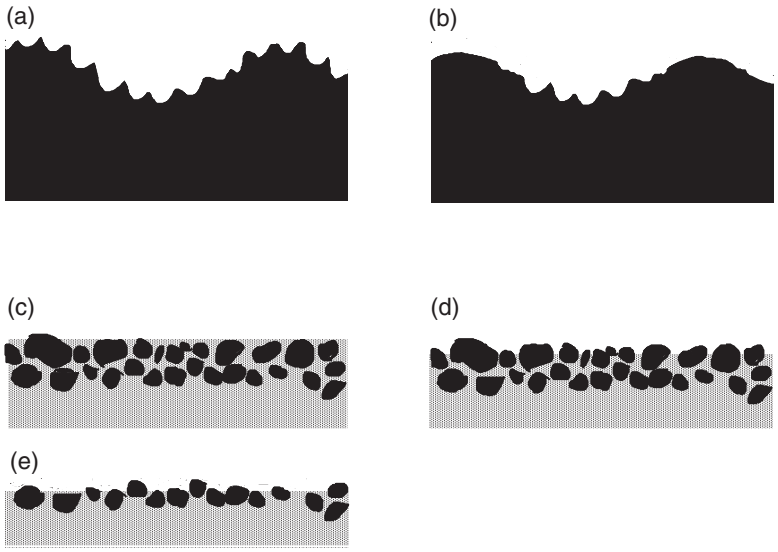

Figure 37. ((a), (b)) Rolling and braking on cobblestone road surfaces causes the surface asperities to become polished. ((c)–(e)) On asphalt or concrete road surfaces smaller hard stone particles never become highly polished as they are continuously exposed and/or removed by wear of the softer surrounding matrix.

determined principally by the inherent quality of the aggregate itself. Rocks composed of minerals of widely different hardness, and rocks that wear by the pulling out of mineral grains from a relative soft matrix, have high resistance to polishing; see figures 37(c)–(e). Conversely, rocks consisting of minerals having nearly the same hardness wear uniformly and tend to have lower resistance to polishing; see figures 37(a), (b). Thus, sandstones have a good resistance to polishing, whereas the limestone and flint groups present the lowest resistance[13]. Among other groups, basalt, granite and quartzite yield intermediate results. Microroughness is gradually polished away by the action of heavy traffic so, over time, the skidding resistance of a road will fall to an 'equilibrium' level that depends upon the type of aggregate used in the surfacing. For concrete and asphalt surfaces that have too low friction, diamond grinding will bring the skid resistance up. Grinding increases the surface microstructure by dislodging polished sand particles in the mortar matrix.

Experience of road engineers shows that rubber friction on dry clean roads depends on microroughness with wavelength mainly in the range $1\ \mu\text{m} < \lambda < 1000\ \mu\text{m}$, in accordance with our calculations. Longer wavelength roughness $0.1\ \text{cm} < \lambda < 5\ \text{cm}$ is important on wet road surfaces especially to prevent hydroplaning for velocities above 60 km h^{-1}. High dry friction on asphalt or concrete road surfaces is obtained when hard minerals (such as feldspar and quartz) and aggregates with good microstructure, such as sandstone and slag

[13] The influence of polishing of road surfaces is studied in a standardized laboratory test where fragments (of diameter ∼1 cm) of the stone (or other material) to be tested are glued onto a flat surface and exposed to accelerated (as compared to polishing on the road) polishing using standardized conditions. The rubber friction coefficient obtained after a fixed time period of polishing is then measured. The result will depend on the rubber used in the friction test (and the laboratory conditions, e.g., the temperature), and only relative friction values (for different polished surfaces) are analysed. In one such set of measurements (see http://www.wainwright.co.uk/technical.htm) it was found that gritstone and sandstone have the best resistance against polishing (giving friction coefficients after polishing $\mu = 0.68$ and 0.67, respectively) followed by basalt and granite ($\mu = 0.56$ and 0.55 respectively), while limestone gave the worst result ($\mu = 0.37$).

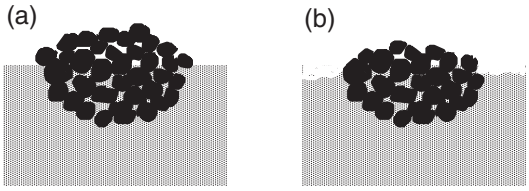

Figure 38. Asphalt or concrete roads with aggregates of small (micrometre) particles, e.g., fragments of sandstone or slag, have a high resistance against polishing if the tyre–road interaction mainly results in a slow removal of the microparticles in the aggregates, preserving an aggregate particle surface that is (on the microscale) constantly rough. (a) and (b) show an aggregate before and after wear (schematically).

(usually, air-cooled blast furnace slag from iron production), are embedded in a softer matrix, e.g., tar (asphalt) or mortar (concrete). Minerals with rough grains, or mixtures of minerals with different textures, will resist polishing and maintain high frictional properties. Special surface coatings with very high skid resistance have been developed. One such coating consists of epoxy adhesive binding a synthetic aggregate to the road surfaces[14]. The aggregate, manufactured from steel slag, has extremely high polish resistance, and is claimed to result in tyre–road friction about 40% higher than on normal road surfaces. In temperate climates such as Sweden's, the surface microroughness varies cyclically, with skidding resistance at its lowest during the summer and autumn, recovering to some extent during the winter. This is due in part to weather induced erosion of the road surface (e.g., freezing water in the road cavities and frost heave may break off road surface fragments), and partly due to the strong wear induced surface roughening which results from the 'heavy' equipment used to clean the road surfaces of snow and ice.

Summarizing, it is clear that the best road surfaces, with high resistance against polishing, consist of aggregates of very small (micrometre) and hard particles, e.g., fragments of sandstones, where the wear occurs by removal of the individual microparticles or clusters of microparticles, rather than by polishing the particles; see figure 38. However, the binding of the particles in the aggregates must be so strong that the wear of the road surface is slow, i.e., the rate of wear by removal of particles should be only slightly larger than the rate of wear by polishing of the particles. The optimal case is when the wear process results in particle aggregates of various sizes being removed from the road surface so that the road surface remains rough at all the length scales above the size of the smallest particles.

6.3. Rubber friction on wet road surfaces

For rubber friction on *wet* rough substrates at low sliding velocities it is known that the friction typically drops by as much as 20–30% relative to the corresponding dry case [52, 53]. Owing to the small contact area, this cannot be the result of a water induced change of adhesion. On the other hand, as will be discussed below, the friction decrease cannot be blamed on a purely hydrodynamical effect either. That leaves finally the possibility that water might change precisely the bulk, hysteretic friction. We proposed recently [54] that this is indeed the case. Water pools that form in the wet rough substrate are *sealed off* by the rubber, as sketched in figure 39, and that will effectively smooth the substrate surface. Smoothing reduces the viscoelastic deformation from the surface asperities, and thus reduces rubber friction.

[14] See http://www.new-technologies.org/ECT/Civil/italgrip.htm for high friction coatings for road surfaces.

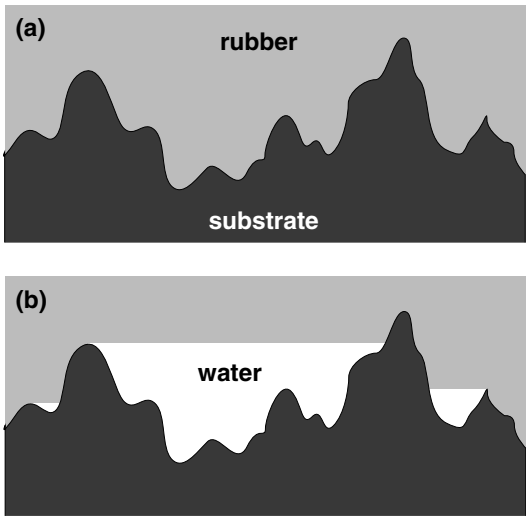

Figure 39. A rubber block sliding on a rough hard substrate. (a) On a dry substrate the rubber penetrates a large valley and explores the short wavelength roughness within. The pulsating rubber deformations induced by the short wavelength roughness contribute to the friction force. (b) On a wet substrate the valley turns into a water pool. Sealing of the pool now prevents the rubber from entering the valley. By removing the valley contribution to the frictional force, this *sealing effect* of rubber reduces the overall sliding friction.

As discussed before, rubber friction from the viscoelastic deformation of the substrate asperities is determined by the complex frequency-dependent bulk viscoelastic modulus $E(\omega)$ of rubber and by the substrate surface roughness power spectrum $C(q)$.

The upper curve in figure 41 shows the power spectrum extracted via equation (1) from the measured height profile $h(\boldsymbol{x})$. The log–log scale shows that for $q > 1600\ \mathrm{m}^{-1}$, $C(q)$ drops as a power law, as expected for a self-affine fractal surface. The fractal dimension of this surface is determined by the slope of the curve in figure 41 and is about $D_\mathrm{f} = 2.2$. The root mean square roughness can be obtained directly from the height profile, $h_\mathrm{rms} \approx 0.3$ mm.

Consider now a tyre rolling and sliding on a wet road surface. At low velocities (say $v < 60\ \mathrm{km\ h}^{-1}$) there will be negligible hydrodynamic water build-up between the tyre and the road surface [54]. In essence, if $v < (\sigma/\rho)^{1/2}$, where σ is the perpendicular stress in the tyre–road contact area and ρ the water mass density, there is sufficient time for the water to be squeezed out of the contact regions between the tyre and the road surface, *except* for water trapped in road cavities. The water pools will be sealed off by the road–rubber contact at the upper boundaries of the cavities (see figure 39). Thus, we can focus on the smoothing effect on the road profile caused by the sealing effect.

Starting from a dry substrate profile $h(\boldsymbol{x})$ we can numerically build a new wet surface height profile $h'(\boldsymbol{x})$ as shown in figure 40(b). The algorithm assumes every valley to be filled with water up to the maximum level where the water will remain confined, i.e., up to the lowest point of the edge surrounding the pool. Any extra water added to the profile of figure 40(b) will flow straight out of the square area. Once the size of the area considered is at least as large as λ_0 this construction becomes unique, with no free parameter.

From the water-smoothed height profile $h'(\boldsymbol{x})$ we obtain a modified power spectrum $C'(q)$ shown by the lower curve in figure 41. While the fractal power-law decay and the roll-off wavevector are essentially the same as for the dry surface, the reduction in the power spectrum reflects the effective water induced smoothing of the rough substrate.

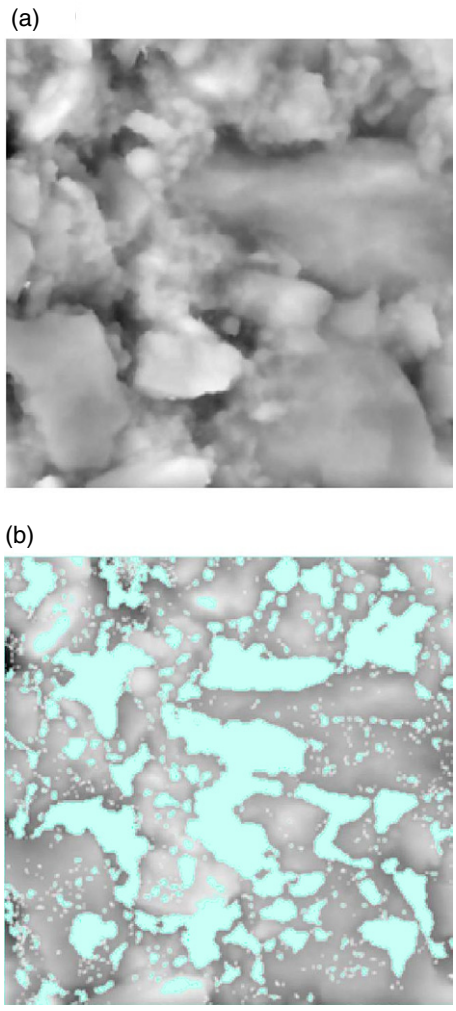

Figure 40. (a) The optically observed height profile of a dry asphalt road (1.5 cm × 1.5 cm area), darker areas corresponding to deeper regions. (b) The calculated wet profile for the same area, with water pools (light grey (blue)).

The sealed-off water in the pools (see figure 39) removes the contact with the interior of the valley, which smooths the effective substrate roughness profile. Our basic assumption is therefore that when rubber slides on the wet rough surface, the friction force will be determined by the modified power spectrum $C'(q) < C(q)$.

Let us now examine, on the basis of this model, numerical results for tyre friction on dry and wet substrates, calculated using the hysteretic friction theory presented in [20]. The hysteretic friction coefficient at velocity v is determined from knowledge of the rubber viscoelastic modulus $E(\omega)$ and of the surface roughness spectrum $C(q)$.

We present results for the friction of a standard tread compound, sliding on the asphalt road just characterized. We used the measured rubber complex viscoelastic modulus (not shown) along with the power spectra presented in figure 41 for the dry and for the wet road surfaces.

Figure 42 shows the rubber–asphalt kinetic friction coefficient calculated for the dry surface (here including also the flash temperature) at $T = 60\,^{\circ}\mathrm{C}$ as a typical tyre temperature while rolling on a dry road and for the wet surface at four different temperatures, namely $T = 30$, 40, 50 and $60\,^{\circ}\mathrm{C}$ (on a wet road the tyre temperature is typically about $30\,^{\circ}\mathrm{C}$, and

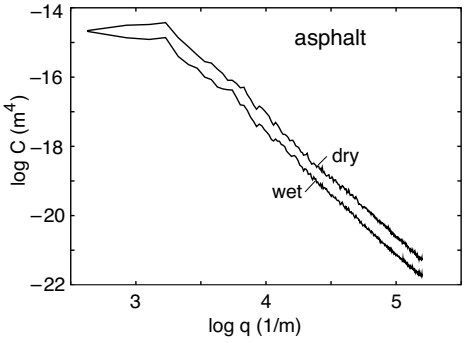

Figure 41. Surface roughness power spectra $C(q)$ (above) extracted from the measured height profile for a dry asphalt road surface and (below) calculated assuming sealing of all pools in the same surface when wet, as in figure 40. Note the logarithmic scales.

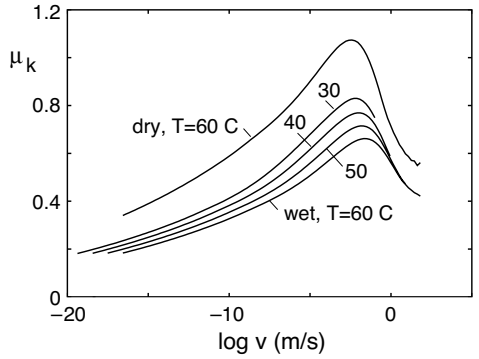

Figure 42. The kinetic friction coefficient as a function of the logarithm of the sliding velocity, calculated for a standard tread compound and an asphalt substrate with the roughness spectra of figure 41.

generally lower than on the dry surface). When dry and wet frictions are compared, the calculation shows first of all a water induced friction decrease of ∼30% at $T = 60\,^{\circ}$C. The water induced decrease becomes somewhat less (∼20%) if the wet substrate temperature is (realistically) reduced to $T = 30\,^{\circ}$C. We have also calculated μ-slip curves, and they show a similar reduction in the friction for wet road surfaces. These figures are in excellent agreement with the known reduction of low speed rubber friction on road surfaces [52, 53]. We note in addition that the decrease in friction with increasing temperature shown in figure 42 is very commonly observed for rubber. It results from the shift in the viscoelastic spectrum to higher frequencies with increasing temperature, making the rubber more elastic and less viscous, in turn reducing the rubber friction.

The above picture in our view catches an important novel effect of water on rubber friction. Yet it is certainly open to refinements in various ways. First, dry friction of tyres is not pure sliding but also involves some stick–slip [55]. This effect is included in the calculation of μ-slip curves, but the observed reduction in the effective friction is similar to that in figure 42. Second, after enough time all sealing regions leak. This will be particularly true in the present case because the upper boundary of a water filled pool, which is in contact with the rubber, still has roughness on many length scales. So one cannot expect the rubber to make equally perfect contact everywhere, and there will be narrow channels through which the water slowly leaks out of the pools. As a result, for sufficiently low sliding velocities the negative water influence on rubber friction may revert to negligible. Experiments have indeed shown that for extremely low velocities $v < 0.7$ m s^{-1} the difference in μ_{k} between dry and wet surfaces is very small [55]. We should also stress that the effects addressed here clearly apply only to moderately wet substrates and for rolling or sliding velocities $v < 60$ km h^{-1}. For flooded surfaces and $v > 60$ km h^{-1} aquaplaning may occur, which originates instead from the inertia of the water. Finally, for rubber friction on relative smooth wet surfaces, where the adhesional interaction is important, the so-called dewetting transition may be important [56–58].

6.4. Lubricated rubber O-ring seals

Surface roughness also has a big influence on lubricated rubber O-ring seals [35]. Tests have shown that the longer a lubricated seal sits idle, the higher its static, or start-up friction

coefficient μ_s. Eventually, for smooth surfaces, the friction coefficient reaches a maximum almost as high as for an unlubricated seal. This increase in μ_s with time is caused by the squeeze-out of most of the lubricant from the contact area [59]. However, the static friction can be reduced by optimizing the surface roughness and lubricant viscosity. Thus, in one experiment [35] with surfaces with the rms roughness amplitude $\sigma \approx 0.4$ μm tiny pockets of lubricant were found trapped at the interface, making it available at start-up. Too smooth a finish leaves no pockets of trapped lubricant, while too rough surfaces can cause high wear.

7. Adhesion

In this section we discuss adhesion between rough surfaces. We point out that even when the force for separating two solids vanishes, there may still be a finite contact area (at zero load) between two solids as a result of the adhesional interaction between the solids. We also present two applications where surface roughness influences adhesion in a fundamental way. We first consider an industrial application related to one step in the fabrication of rubber sheets. For this system a recently developed viscoelastic contact mechanics theory, combined with surface topography measurements using the atomic force microscope, was able to explain the origin of a long standing technological problem.

The second application is a study of adhesion relevant to biological systems, e.g., flies, crickets and lizards, where the adhesive microstructure consists of arrays of thin fibres and plates. The effective elastic modulus of the fibre–plate arrays can be very small, which is of fundamental importance for adhesion on smooth and rough substrates. This application illustrates how Nature, through the process of natural selection, has been able to produce elastically very soft, but still wear-resistant layers, which can make good contact and exhibit strong adhesion even to surfaces with roughness on all length scales down to the atomic dimension.

7.1. Adhesion between rough surfaces

A theory of adhesion between an elastic solid and a hard randomly rough substrate must take into account that partial contact may occur between the solids on all length scales. For the case where the substrate surface is a self-affine fractal, theory shows that when the fractal dimension is close to 2, complete contact typically occurs in the macroasperity contact areas (the contact regions observed when the system is studied at a magnification corresponding to the roll-off wavelength $\lambda_0 = 2\pi/q_0$ of the surface power spectra; see figure 6), while when the fractal dimension is larger than 2.5, the area of (apparent) contact decreases continuously when the magnification is increased. An important result is that even when the surface roughness is so high that no adhesion can be detected in a pull-off experiment, the area of real contact (when adhesion is included) may still be several times larger than when the adhesion is neglected. Since it is the area of real contact which determines the sliding friction force, *the adhesion interaction may strongly affect the friction force even when no adhesive force can be detected in a pull-off experiment*.

The influence of surface roughness on the adhesion between rubber (or any other elastic solid) and a hard substrates has been studied in a classic paper by Fuller and Tabor [60] (see also [61–67]). They found that even a relatively small surface roughness can completely remove the adhesion. In order to understand the experimental data they developed a very simple model based on the assumption of surface roughness on a single length scale. In this model the rough surface is modelled by asperities, all possessing the same radius of curvature, and with heights following a Gaussian distribution. The overall contact force was obtained

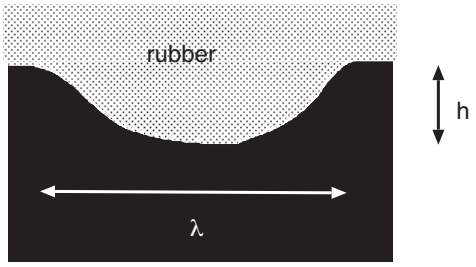

Figure 43. A rubber surface is 'pulled' into a cavity of the hard solid substrate by the rubber–substrate adhesional interaction. The elastic energy stored in the deformation field is of order $E\lambda h^2$.

by applying the contact theory of Johnson *et al* [68] to each individual asperity. The theory predicted that the pull-off force, expressed as a fraction of the maximum value, depends upon a single parameter, which may be regarded as representing the statistically averaged competition between the compressive forces exerted by the higher asperities trying to prise the surfaces apart and the adhesive forces between the lower asperities trying to hold the surfaces together. This picture of adhesion developed by Tabor and Fuller would be fine *if* real surfaces had roughness on a single length scale as assumed in their study. However, with roughness occurring on many different length scales, a qualitatively new picture emerges [22], where, e.g., the adhesion force may even vanish (or at least be strongly reduced) if the rough surface can be described as a self-affine fractal with fractal dimension $D_f > 2.5$. In fact even for surfaces with roughness on a single length scale, the formalism used by Fuller and Tabor is only valid at 'high' surface roughness, where the area of real contact (and the adhesion force) is very small. The theory that will be presented below is particularly accurate for 'small' surface roughness, where the area of real contact equals the nominal contact area.

7.1.1. Qualitative discussion. Let us estimate the energy necessary in order to deform a rubber block so that the rubber fills out a substrate cavity of height h and width λ. The elastic energy stored in the deformation field in the rubber is given by

$$U_{el} \approx \tfrac{1}{2} \int d^3x\, \sigma\epsilon$$

where the stress $\sigma \approx E\epsilon$ and E is the elastic modulus. The deformation field is mainly localized in a volume $\sim\lambda^3$ (see figure 43) where the strain $\epsilon \approx h/\lambda$. Thus we get $U_{el} \approx \lambda^3 E(h/\lambda)^2 = E\lambda h^2$.

Let us now consider the role of the rubber–substrate adhesion interaction. If the elastic energy $U_{el} \approx E\lambda h^2$ stored in the deformed rubber is smaller than the gain in adhesion energy $U_{ad} \approx \Delta\gamma\lambda^2$, where $\Delta\gamma = \gamma_1 + \gamma_2 - \gamma_{12}$ is the change of surface free energy (per unit area) upon contact due to the rubber–substrate interaction (which usually is mainly of the van der Waals type), then (even in the absence of an external load F_N) the rubber will deform *spontaneously* to fill out the substrate cavities. The condition $U_{el} = U_{ad}$ gives $h/\lambda \approx (\Delta\gamma/E\lambda)^{1/2}$. For example, for very rough surfaces with $h/\lambda \approx 1$ and with parameters typical for rubber, $E = 1$ MPa and $\Delta\gamma = 3$ meV Å^{-2}, the adhesion interaction will be able to deform the rubber and completely fill out the cavities if $\lambda < 0.1\ \mu$m. For very smooth surfaces $h/\lambda \sim 0.01$ or smaller. In that case the rubber will be able to follow the surface roughness profile up to the length scale $\lambda \sim 1$ mm or longer.

The argument given above shows that for elastic solids with surface roughness on a *single length scale* λ, the competition between adhesion and elastic deformation is characterized by the parameter $\theta = Eh^2/\lambda\delta \approx U_{\rm el}/U_{\rm ad}$, where h is the amplitude of the surface roughness and $\delta = 4(1-\nu^2)\Delta\gamma/E$ is the so-called *adhesion length*, ν being the Poisson ratio of the rubber. The parameter θ is the ratio between the elastic energy and the surface energy stored at the interface, assuming that complete contact occurs. When $\theta \gg 1$ only partial contact occurs, where the elastic solids make contact only close to the top of the highest asperities, while complete contact occurs when $\theta \ll 1$.

7.1.2. Pull-off force. Consider a rubber ball (radius R_0) in adhesive contact with a perfectly smooth and hard substrate. The elastic deformation of the rubber can be determined by minimizing the total energy which is the sum of the (positive) elastic energy stored in the deformation field in the rubber ball and the (negative) energy of binding between the ball and the substrate at the contact interface. The energy minimization gives the pull-off force [68, 69]

$$F_{\rm c} = (3\pi/2)R_0\Delta\gamma. \tag{16}$$

Consider now the same problems as above, but assume that the substrate surface has a roughness described by the function $z = h(\boldsymbol{x})$. Let us further assume a surface roughness power spectrum with a roll-off wavelength $\lambda_0 = 2\pi/q_0$ (see figure 6) smaller than the diameter of the nominal contact area between the two solids. In this case we can still use the result (16), but with $\Delta\gamma$ replaced by $\gamma_{\rm eff}$. The effective interfacial energy $\gamma_{\rm eff}$ is the change in the interfacial free energy when the elastic solid is brought in contact with the rough substrate. $\gamma_{\rm eff}(\zeta)$ depends on the magnification ζ, and the interfacial energy which enters in the rubber ball pull-off experiment is the macroscopic interfacial energy, i.e., $\gamma_{\rm eff}(\zeta)$ for $\zeta = 1$. If A_0 is the nominal contact area and A_1 the true atomic contact area, then

$$A_0\gamma_{\rm eff}(1) = A_1\Delta\gamma - U_{\rm el} \tag{17}$$

where $U_{\rm el}$ is the elastic energy stored at the interface as a result of the elastic deformations necessary in order to bring the solids into atomic contact in the area A_1.

7.1.3. Stress probability distribution. The theory in [22] is based on the contact mechanics formalism described in section 4.1. Thus, we focus on the stress probability distribution function $P(\sigma, \zeta)$ which satisfies equation (5):

$$\frac{\partial P}{\partial \zeta} = f(\zeta)\frac{\partial^2 P}{\partial \sigma^2}.$$

We assume that detachment occurs when the local stress on the length scale L/ζ reaches $-\sigma_{\rm a}(\zeta)$. Thus, the following boundary condition applies to the present case:

$$P(-\sigma_{\rm a}(\zeta), \zeta) = 0.$$

This boundary condition replaces the condition $P(0, \zeta) = 0$ valid in the absence of adhesion (see section 4.1).

Let us consider the system on the characteristic length scale $\lambda = L/\zeta$. The quantity $\sigma_{\rm a}(\zeta)$ is the stress necessary to induce a detached area of width λ. This stress can be obtained from the theory of cracks, where, for a penny shaped crack of diameter λ,

$$\sigma_{\rm a} = \left[\frac{\pi\gamma_{\rm eff}(\zeta)E}{(1-\nu^2)\lambda}\right]^{1/2} = \left[\frac{\gamma_{\rm eff}(\zeta)Eq}{2(1-\nu^2)}\right]^{1/2} \tag{18}$$

where $q = 2\pi/\lambda = \zeta q_L$. In [22] we derived two equations for $\gamma_{\rm eff}(\zeta)$ and $P(\zeta)$ which determine how these quantities depend on the magnification ζ; those equations are the basis for the numerical results presented below.

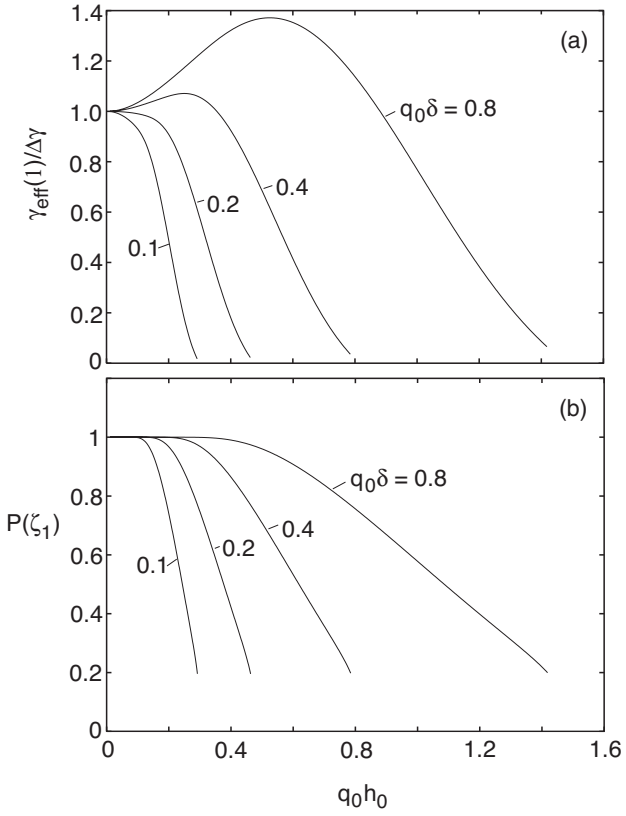

Figure 44. (a) The macroscopic interfacial energy for adhesion of rubber to a fractal surface with $H = 0.8$ and $q_1/q_0 = \zeta_1 = 100$, as a function of q_0h_0. (b) The normalized area of real contact, $P(\zeta_1) = A(\zeta_1)/A_0$, as a function of q_0h_0. For $q_0\delta = 0.1$, 0.2, 0.4 and 0.8 as indicated.

7.1.4. Numerical results. Figure 44 shows (a) the effective interfacial energy $\gamma_{\rm eff}(\zeta)$ ($\zeta = 1$) and (b) the normalized area of real contact, $P(\zeta_1) = A(\zeta_1)/A_0$, as a function of q_0h_0. Results are shown for $q_0\delta = 0.1, 0.2, 0.4$ and 0.8. We will refer to $\gamma_{\rm eff}(1)$ at the magnification $\zeta = 1$ as the *macroscopic* interfacial free energy which can be deduced from, e.g., the pull-off force for a ball according to equation (16). Note that for $q_0\delta = 0.4$ and 0.8 the macroscopic interfacial energy first increases with increasing amplitude h_0 of the surface roughness and then decreases. The increase in $\gamma_{\rm eff}$ arises from the increase in the surface area. As shown in figure 44(b), for small h_0 the two solids are in complete contact and, as expected, the complete contact remains to higher h_0 as $\delta \sim \Delta\gamma/E$ increases. Note also that the contact area is non-zero even when $\gamma_{\rm eff}(1)$ is virtually zero: the fact that $\gamma_{\rm eff}(1)$ nearly vanishes does not imply that the contact area vanishes (even in the absence of an external load), but rather that the (positive) elastic energy stored at the interface exactly balances the (negative) adhesion energy from the area of real contact. *The stored elastic energy at the interface is returned when removing the block, and when* $\gamma_{\rm eff}(1) \approx 0$ *it is just large enough to break the block–substrate bonding.*

7.1.5. Experimental manifestations of adhesion. Unfortunately, the surface roughness power spectrum has not been measured for any of the real surfaces for which adhesion has been studied in detail. Instead only the roughness amplitude (centreline average) and the radius of curvature

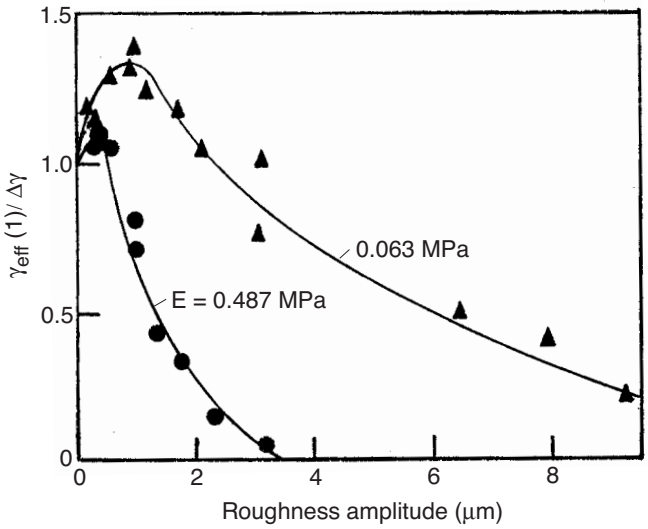

Figure 45. The macroscopic interfacial energy (obtained from the pull-off force) for a smooth rubber surface (ball) in contact with a Perspex surface as a function of the roughness (centreline average) of the Perspex. Results are shown for a 'soft' rubber ($E = 0.063$ MPa) and a 'hard' rubber ($E = 0.487$ MPa). From [63].

of the largest surface asperities were determined. Nevertheless, the experimental data of Fuller, Tabor, Briggs, Briscoe and Roberts [60, 63, 64] are in good qualitative agreement with our theoretical results. In figure 45 we show the macroscopic interfacial energy for 'hard' and 'soft' rubber in contact with Perspex, as a function the substrate (Perspex) roughness amplitude as obtained by Briggs and Briscoe [63]. It is not possible to compare these results quantitatively with the theory developed above since the power spectrum $C(q)$ was not measured for the Perspex substrate. Even if the surfaces were self-affine fractal as assumed above, not only would the surface roughness amplitude change from one surface to another, but also so will the long distance cut-off length λ_0 and hence also the ratio $\zeta_1 = q_1/q_0$. In the experiments reported in [63] the Perspex surfaces where roughened by blasting with fine particles. The roughness could be varied through the choice of the particles and the air pressure.

One practical problem in comparing the theory to experimental data is that most rubber materials have a wide distribution of relaxation times, extending to extremely long times. This effect is well known in the context of rubber friction (see section 6.1), where measurements of the complex elastic modulus show an extremely wide distribution of relaxation times, resulting in a large sliding friction even at very low sliding velocities, $v < 10^{-8}$ m s^{-1}.

The effect of the stored elastic energy on adhesion has recently been studied using a polyvinylsiloxane rubber block squeezed against a smooth glass surface for a fixed time period before measuring the pull-off force [70]. The square symbols in figure 46 show the pull-off force as a function of the squeezing force. For squeezing forces $F_N > 850$ mN the pull-off force decreases. This may be explained by a drastic increase of the elastic energy stored in the rubber because of the strong deformation of the rubber—see figure 47 (top)—some of which remains even when the load is removed as a result of the rubber–glass friction at the interface. This energy, freed during the process of unloading, will help to break the adhesive bonds at the interface. This effect is even stronger when the surface is structured. Thus, the triangles in the figure show the pull-off force when the rubber surface is covered by a regular array of rubber

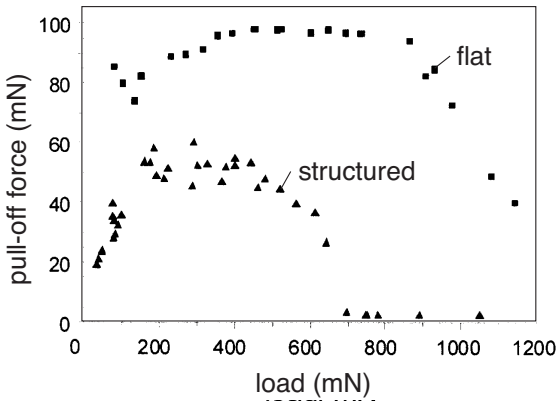

Figure 46. The pull-off force as a function of the squeeze force or load, for silicon rubber in contact with a smooth glass surface. From [70].

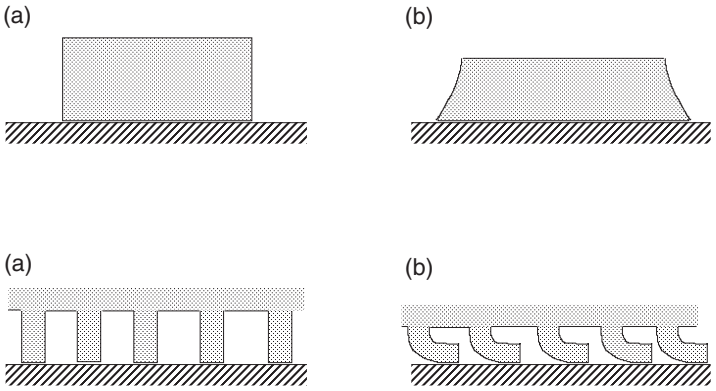

Figure 47. Elastic deformation of a rubber block with a smooth surface (top) and a structured surface (bottom). (a) shows the initial state before applying a squeezing force and (b) the new state (without load) after applying (and then removing) a very large squeezing force. In state (b) a large elastic energy is stored in the rubber which is 'given back' during pull-off resulting in a nearly vanishing pull-off force.

cylindrical asperities. In this case the pull-off force drops to nearly zero for $F_N > 700$ mN. Visual inspection shows that in this case the cylindrical asperities at high load bend and make contact with the glass on one side of the cylinder surface; see figure 47 (bottom). This again stores a large amount of elastic energy at the interface which is given back during pull-off, reducing the pull-off force to nearly zero.

7.1.6. The role of plastic yield in adhesion. When the local stress in the asperity contact regions of two solids becomes high enough, at least one of the solids yields plastically. This will tend to increase the effective adhesion (or pull-off force) for the following three reasons. First, the area of real contact between the solids will increase as compared to the case where the deformations are purely elastic. Secondly, the amount of stored elastic energy in the contact regions (to be given back at pull-off) will be reduced because of the lowered elastic deformations. Finally, for many materials plastic yield will strengthen the junctions [71].

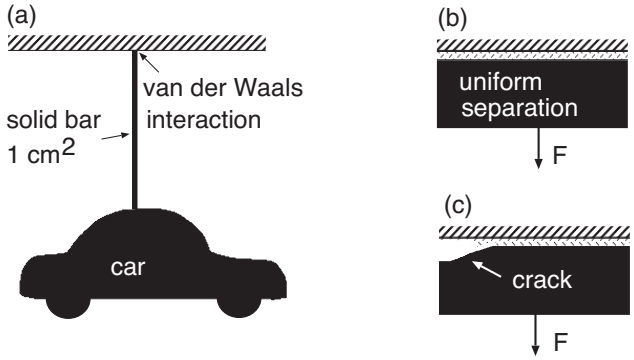

Figure 48. Even the weakest force in Nature which is of relevance in condensed matter physics, namely the van der Waals force, is relatively strong on a macroscopic scale. Thus, for example, if the bond breaking were to occur uniformly over the contact area as in (b), a contact area of order 1 cm^2 could already sustain the weight of a car (i.e., a force of order 10^4 N) (see (a)). However, on a macroscopic scale the bond breaking does not usually occur uniformly over the contact area, but by crack propagation—see (c)—which drastically reduces the pull-off force. Secondly, interfacial surface roughness drastically reduces the pull-off force. Finally, the stored elastic energy will largely balance the adhesion energy (see the text).

For example, most metals are protected by thin oxide layers, and as long as these are intact the main interaction between the surfaces in the contact areas may be of van der Waals and electrostatic origin. However, when plastic yield occurs it may break up the oxide films resulting in direct metal–metal contact and the formation of 'cold welded' junctions. When this occurs, because of the high ductility of many metals, during pull-off 'long' metallic bridges may be formed between the solids with the result that instead of having junctions popping one after another during pull-off, a large number of adhesive junctions may simultaneously impede the surface separation during pull-off, leading to a large pull-off force. However, experiments have shown [72] that just squeezing before pull-off will in general only result in very few cold welded junctions, while squeezing *and* sliding will break up the oxide film, resulting in the formation of many more cold welded contact regions, and will hence result in a much larger pull-off force.

7.2. The adhesion paradox

The biggest 'mystery' related to adhesion is not why it is sometimes observed but rather why it is usually not observed. Even the weakest force in Nature of relevance in condensed matter physics, namely the van der Waals force, is relatively strong on a macroscopic scale. For example, even a contact area of order 1 cm^2 could sustain the weight of a car (i.e., a force of order 10^4 N) (see figure 48(a)) even if only the van der Waals interaction operated at the interface. (Here we assume that the bond breaking occurs uniformly over the contact area, as illustrated in figure 48(b).) However, this is never observed in practice and this fact is referred to as the *adhesion paradox*.

There are several reasons that adhesion is usually not observed between macroscopic bodies. For example, on a macroscopic scale the bond breaking usually does not occur uniformly as in figure 48(b), but occurs by crack propagation; see figure 48(c). The local stress at the crack tip is much higher than the average stress acting in the contact area, and this drastically reduces the pull-off force. Another reason, already addressed in section 7.1, is

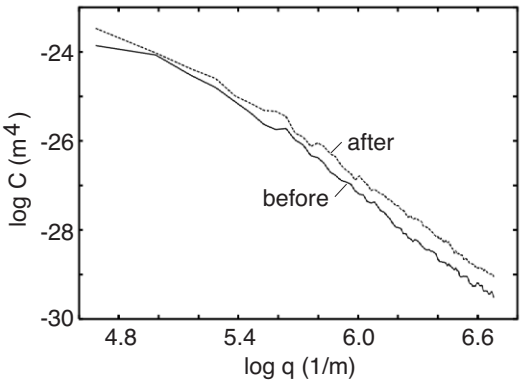

Figure 49. The surface roughness power spectra of a steel cylinder before and after lapping with corundum paper. The lapping increases the surface roughness.

the influence of surface roughness. Thus, for elastically hard surfaces the true atomic contact between the solids at the interface is usually much smaller than the nominal contact area. In addition, the elastic energy stored in the solids in the vicinity of the contact regions is given back during pull-off and helps to break the interfacial bonds between the solids (see section 7.1).

It is interesting to note that for very small solid objects, typically of order 100 μm or smaller, the bond breaking may occur uniformly over the contact area (no crack propagation), so adhesion between smooth surfaces of small objects, e.g., in micromechanical applications (microelectromechanical systems, MEMS), may be much stronger than for macroscopic bodies, and this fact must be taken into account when designing MEMS [73, 74].

7.3. Adhesion in rubber technology

Most rubber compounds of commercial use contain a large fraction (around 30%) of filler particles which are usually mixtures of carbon and silica particles. The silica particles are very small (below micrometre size) and very hard. As a result, when rubber is sliding on a substrate, even if the latter is very hard, e.g., stone or steel, the substrate will become polished by the rubber. We have discussed this effect above for road surfaces (see section 6.2). Here we describe another recent observation of the same effect with important practical implications.

One stage in the production of rubber for technological applications involves mixing in open mills, where unvulcanized rubber (with filler particles) is fed between rotating steel cylinders. The slip of the (silica particle containing) rubber relative to the steel walls during the mixing process results in a continuous 'polishing' (or wear) of the steel surfaces. This will slowly increase the rubber–steel contact area, and finally the rubber may adhere to the steel surfaces, which is of course unwanted.

Using atomic force microscopy (AFM) we measured the surface topography of the steel cylinders, and found that during production the steel surfaces are continuously being polished by the rubber containing silica particles. This leads to a continuously increased contact area between the steel and the rubber sheet, and to an increased adhesive steel–rubber interaction. Using a recently developed viscoelastic contact mechanics theory (see section 4.3) we further found that the increase in the contact area can be very large.

Also shown by AFM measurements was that lapping of the steel surface with corundum paper *increases* the steel surface roughness (see figure 49), which is easy to understand when one recognizes that the corundum particles have on average larger diameter than the silica

Figure 50. A lizard adhering to a nearly vertical surface.

particles in the rubber. Thus we conclude that lapping or 'polishing' may increase the roughness of the steel cylinders, and hence reduce the adhesive interaction to such a level that it becomes unimportant.

7.4. Adhesion in biology

How can a fly or a cricket walk on a glass window, or a lizard (see figure 50) move on a stone or concrete wall? These fundamental questions have interested children and scientists for many years, and recently very important experimental work has been performed that gave a deeper insight into these questions [75, 76]. Here we focus mainly on dry adhesion, which seems to be relevant for lizards, and we discuss the influence of surface roughness on the adhesion between a lizard or a gecko toe and a rough hard substrate.

It has been demonstrated that a foot of a gecko can adhere to a substrate with a force $\sim$10 N (corresponding to the weight of 1 kg!) The typical weight of a Tokay gecko lizard (*Gecko gecko*) is approximately 40 g, meaning that only 1% of the maximum adhering force of its feet is required to support the whole weight of the gecko. This raises the question of why geckos are apparently so overbuilt. However, a gecko must be able to adhere to very rough surfaces, and we will show below that the adhesion to rough surfaces can be reduced significantly [77, 78].

The adhesion between two bodies in contact results almost entirely from the regions where the surfaces are separated by one nanometre or less. For hard solids this area of real contact is extremely small. Furthermore, for elastically hard solids a large elastic energy is stored in the solids in the vicinity of the contact regions and during separation of the solids this energy is released and will hence reduce the separation force as explained earlier. Thus, strong adhesion is only possible if at least one of the solids is elastically very soft, or if there is an elastically very soft layer at the interface between the solids. Since the lizard skin comprises a relatively stiff material (keratin), with an elastic modulus of the order of 1 GPa, i.e., about 1000 times higher than that of tyre rubber, it is not immediately obvious why the lizard can adhere to very rough stone walls.

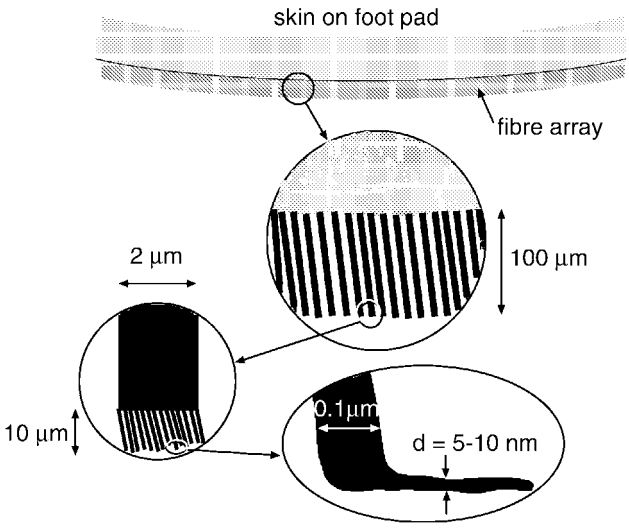

Figure 51. A schematic picture of the lizard adhesive system. The skin of the lizard is covered by a dense layer of thin fibres or hairs (setae; length: $\approx$100 μm and thickness of fibre: $\sim$4 μm). Each of these fibres branches into about 1000 thinner fibres (length: $\sim$10 μm and width: $\sim$0.1 μm). Each of the thin fibres ends with a thin (5–10 nm) leaf-like plate (spatula).

As it turns out, during millions of years of evolution and natural selection, an extremely soft elastic layer has appeared on the lizard pad surface. This layer is built in a hierarchical manner from fibres and plates (figures 51 and 52), that reflect the hierarchical nature of most natural surfaces (to which the lizard must be able to adhere), which have roughness on all length scales, from the macroscopic scale (e.g., the size of the lizard toe pad) down to the atomic scale. The skin of the lizard pad, which consists of a $\sim$100 μm thick keratin layer, is covered by a dense layer of fibres or hair (length: $\approx$100 μm and thickness: $\sim$4 μm). Each of these fibres branches into about 1000 thinner fibres (length: $\sim$10 μm and width: $\sim$0.1 μm), and each terminal fibre ends with a thin (5–10 nm) leaf-like plate (spatula). This construction makes the lizard adhesive system elastically very soft on all relevant length scales (from millimetres to nanometres).

The skin of the gecko toe pad is able to deform and follow the substrate roughness profile on length scales much longer than the thickness $d \approx 100$ μm of the elastic keratin film, say beyond $\sim$1000 μm. At shorter length scales the keratin film, because of its high elastic modulus (of order 1 GPa), can be considered rigid and flat. Elastic deformation of the pad surface on length scales shorter than $\sim$1000 μm involves the compliant setae fibre array system, with fibres of thickness $\sim$4 μm. In [77] one of us has shown that if the surface roughness root mean square amplitude, measured over a patch $D \times D$ with $D \approx 1000$ μm, is smaller than a characteristic length (the adhesion length) (see [77]), then the fibre array system is able to deform (without storing much elastic energy) to follow the surface roughness in the wavelength range $10\,\mu\text{m} < \lambda < 1000\,\mu\text{m}$. However, if the setae fibre tips were instead blunt and compact, they would not be able to penetrate into the surface 'cavities' with diameters less than a few microns. Thus, negligible atomic contact would occur between the surfaces, and adhesion would be negligible. For this reason, there is an array of $\sim$1000 thinner fibres (diameter of order $\sim$0.1 μm) at the tip of each long (thick) fibre. These fibres are able to penetrate into the surface roughness cavities down to length scales of a few tenths of a micrometre; see figure 54. Again if the thin fibres had blunt and compact tips made of the same 'hard' keratin as the rest

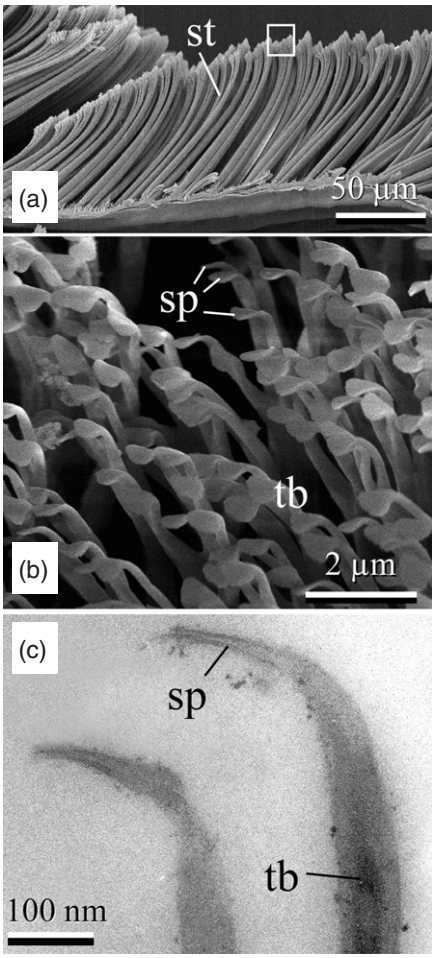

Figure 52. Details of the attachment system of the Tokay gecko (*Gecko gecko*). (a) A scanning electron microscopy (SEM) micrograph of setae (st) located on thin keratin film. (b) An enlargement (SEM micrograph) of the area surrounded by the white rectangle in (a), showing terminal branches (tb) of setae with the spatula (sp). (c) A transmission electron microscopy micrograph of an ultrathin section of two terminal branches (tb) with spatulae (sp). From [78].

of the fibre, one would still obtain very low adhesion, since much elastic energy would be needed to deform the surfaces of the thin fibres to make atomic contact with the substrate. But in fact the tops of the thin fibres end with thin leaf-like plates, which can easily bend (without storing much elastic energy) to follow the surface roughness profile [78, 79]. The calculations presented in [78] show that, for rough surfaces with the fractal dimension $D_f > 2.3$, very low spatula–substrate adhesion may occur in most cases. However, natural surfaces tend to have a fractal dimension of order 2.2, and adhesion may be appreciable even for very rough surfaces in these cases. Experiments to test the theoretical results are in progress.

There is an interesting aspect of fibre adhesion which differs from that involving compact solids. Fibre adhesion depends only on the height probability distribution P_h and not on the power spectra $C(q)$. The reason is that the elastic energy stored in a bent fibre depends only on the separation between the surfaces at the interface which is determined by the surface

Figure 53. Spider-Man. Somewhere in our Universe there may be a planet with 'animals' as large as humans moving on vertical walls .... 'SPIDER-MAN'. Motion Picture © 2002 Columbia Pictures Industries, Inc. Spider-Man Character ® & © 2002 Marvel Characters, Inc. All Rights Reserved. Courtesy of Columbia Pictures.

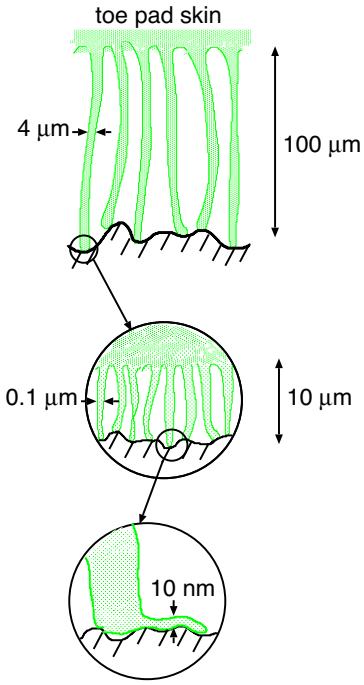

Figure 54. The lizard fibre and plate adhesive system can be easily elastically deformed to bind to a rough substrate, even when the substrate has roughness on all length scales from, say, ∼1 mm to ∼1 nm.

roughness height distribution P_h. Thus, the elastic energy stored in a bent fibre does not depend directly on $C(q)$. On the other hand, for compact solids the elastic energy is determined only by $C(q)$.

Lizards are the heaviest living objects on this planet that are able to adhere to, for example, a rough vertical stone wall. Since the surface area of a body increases less than the volume or mass as the size of the body increases, the adhesive system in large living bodies, such as lizards, must be much more effective (per unit attachment area) than in smaller living objects such as flies and beetles. This implies that lizards have the most effective adhesive systems found in the biological evolution for the purpose of locomotion. This is confirmed by electron microscopy studies. Thus, the spatulas are thinner in lizards than in beetles. Also the diameter of the terminal branches is smaller. This implies that less elastic energy per unit surface area will be stored in the lizard adhesive system, resulting in a stronger adhesion for lizards than for beetles.

The construction of synthetic adhesives based on fibre and plate arrays might constitute an attractive alternative to the usual adhesives based on thin polymer films (see figure 53 for one 'application'). Some pioneering experiments have indeed shown enhanced adhesion for fibre array systems, but no synthetic systems with the hierarchical nature found in biological systems have so far been produced. In addition, if the fibres are too long and thin, or the fibre material is (elastically) too soft, the attractive fibre–fibre interaction will result in elastic instabilities, leading to fibre bundles or fibre 'condensation' into compact layers. This effect

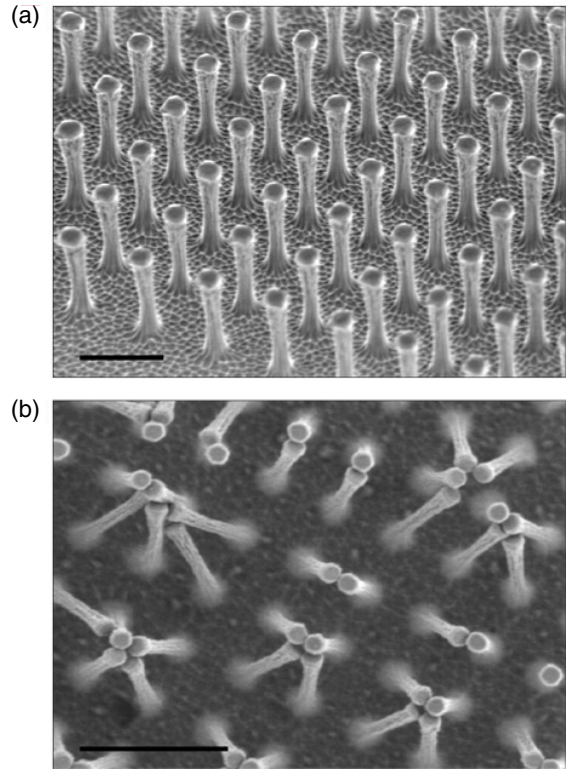

Figure 55. When the fibres are too thin and long, or made of materials that are elastically too soft (as in the present case of rubber shown here), the standing-up fibre array system in (a) is unstable, and after contact with a substrate the fibres end up in the bundled-up state (b). Adapted from [80]. Reproduced with permission.

was in fact observed in the first synthetic fibre adhesion system (which used fibres made from rubber which is 1000 elastically softer than keratin, used in most biological applications)—see figure 55—and was at the same time predicted theoretically [77]. It turns out that the fibre arrays in lizards are close to this instability, but, unsurprisingly, on the correct side of it, and no fibre bundling or condensation occurs in these biological systems.

7.5. The role of liquids in adhesion between rough solid surfaces

As explained in section 7.1, surface roughness reduces the adhesion between clean surfaces. First, it lowers the area of real contact. Since the adhesion interaction comes almost entirely from the area where the solids make atomic contact, it is clear that the surface roughness can drastically reduce the adhesion. Secondly, elastic deformation energy is stored in the vicinity of the asperity contact regions. During pull-off, the elastic energy is 'returned' to the system, usually resulting in a drastic reduction in the effective adhesion and the pull-off force.

Most surfaces have at least nanoscale roughness, and hard solids in the normal atmosphere have at least a monolayer of liquid-like 'contamination' molecules, e.g., water and hydrocarbons. Small amounts of a wetting lubricant or contamination liquids between rough solid walls may drastically enhance the adhesion. Thus, for surfaces with nanoscale roughness, a monolayer of wetting liquid may result in the formation of a large number of

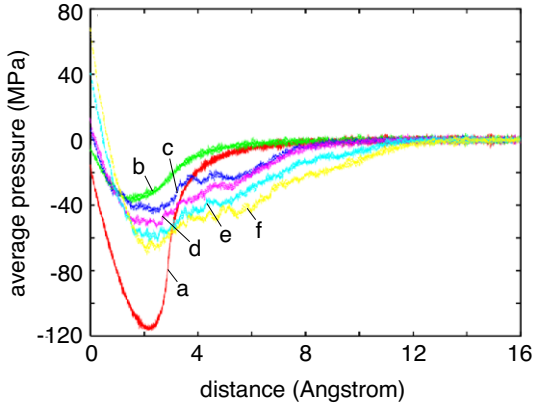

Figure 56. Simulated adhesion between two solid elastic walls, one of which has nanoscale roughness, and with a variable amount of wetting lubricant in between. The variation of the average pressure during retraction develops as a block moves a distance of 16 Å away from the substrate. Octane, C_8H_{18}, was used as the lubricant. The pull-off (retraction) velocity was $v_z = 1$ m s^{-1}. (a) For the flat substrate without lubricant. (b) For the corrugated substrate without lubricant. Curves (c)–(f) show results for the corrugated substrate with about $1/8$, $1/4$, $1/2$ and 1 monolayer of octane in the contact region, respectively. For clarity, the curve for the flat substrate (a) is displaced to the right, by 2 Å.

nanobridges between the solids, which increases the pull-off force. This effect is well known experimentally. For example, the adhesion force which can be detected between gauge blocks (steel blocks with very smooth surfaces) is due to the formation of many very small capillary bridges made of water or organic contamination. For thicker lubrication or contamination films the effective adhesion will be more long ranged but the pull-off force may be smaller. The thickness of the lubricant or contamination layer for which the pull-off force is maximal will in general depend on the nature of the surface roughness, but is likely to be of the order of the root mean square roughness amplitude. In fact, it is interesting and important to find out at exactly what liquid thickness the pull-off force is maximal.

Some insects such as flies and crickets inject a thin layer of a wetting liquid in the contact region between the insect attachment surfaces and the (rough) substrate. The optimum amount of injected liquid will depend on the nature of the substrate roughness, and it is likely that the insect can regulate the amount of injected liquid via a feedback system involving the insect nerve system.

Here we consider the adhesion between two solid elastic walls with nanoscale roughness, lubricated by octane [59, 66, 67]. We consider two types of substrate (bottom surface)—flat and nanocorrugated (corrugation amplitude: 1 nm and wavelength of the corrugation in x and y directions: 4 nm)—and varied the lubricant coverage from $\sim 1/8$ to ~ 4 monolayers of octane. The upper surface (the block) is assumed to be atomically smooth but with a uniform cylindrical curvature with a radius of curvature $R \approx 100$ nm (see figure 58 below). The simulation results presented here were obtained using standard molecular dynamics calculations [66].

Figure 56 shows the variation of the average pressure during retraction as the block moves a distance of 16 Å away from the substrate. The pull-off (retraction) velocity was $v_z = 1$ m s^{-1}. We varied the lubricant coverage from 0 to 1 monolayer in the contact region. The pull-off force is maximal when the adsorbate coverage is of the order of one monolayer (curve (f)). However, the pull-off force is still smaller than for a *flat* substrate without lubricant (curve (a)). As a function of the octane lubricant coverage and for the corrugated substrate, the pull-off

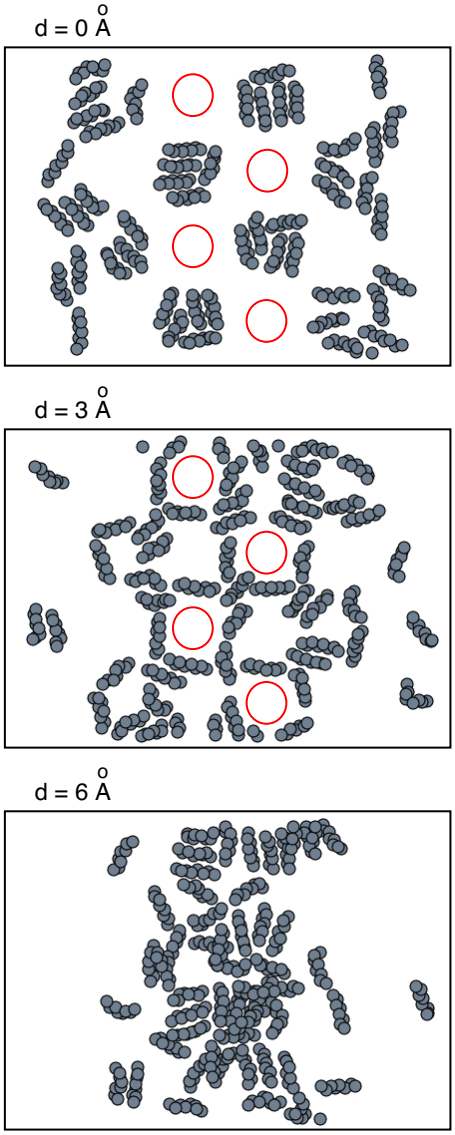

Figure 57. Snapshot pictures (for three different block positions $d = 0$, 3 and 6 Å) of the lubricant layer during retraction. We only show the lubricant molecules in the central part of the contact area between the block and the substrate surfaces (top view, surfaces parallel to the plane of the image). For a corrugated substrate with about 1/4 monolayer of octane in the contact region. The circles indicate the positions of several asperity tops from the corrugated substrate surface.

force first increases as the coverage increases from zero to ∼1 monolayer, and then decreases as the coverage is increased beyond monolayer coverage (not shown).

At low octane coverage, the octane molecules located in the substrate corrugation wells during squeezing are pulled out of the wells during pull-off, forming a network of nanocapillary bridges around the substrate nanoasperities, thus increasing adhesion between the two surfaces; see figures 57 and 58. For greater lubricant coverages a single capillary bridge is formed.

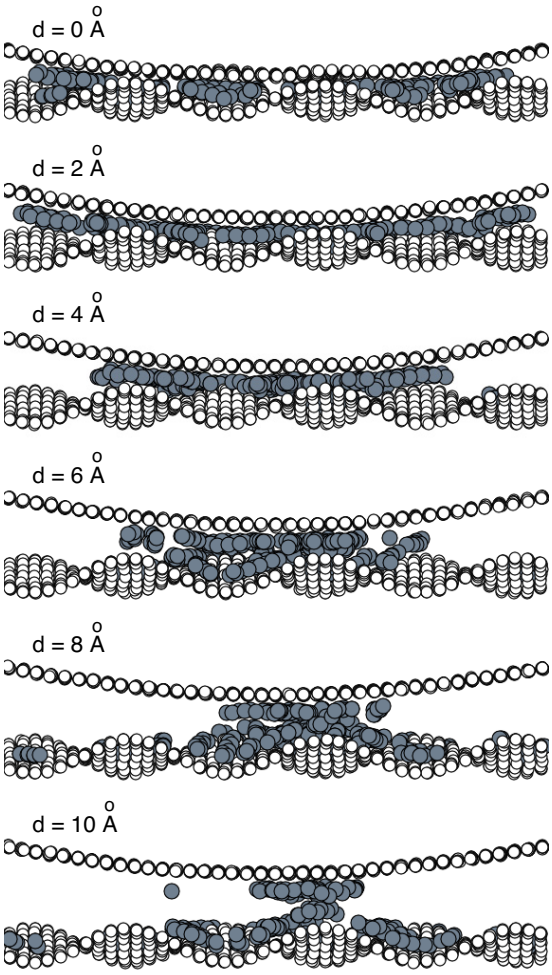

Figure 58. Snapshot pictures (for six different block positions) during retraction. The snapshot pictures show the side view of the central 108 Å × 50 Å section (in the xy-plane) of the contact area. Octane, C_8H_{18}, was used as the lubricant. The pull-off (retraction) velocity was $v_z = 1$ m s^{-1}. For a corrugated substrate with about 1/4 monolayer of octane in the contact region.

Let us discuss the nature of the adhesion for the corrugated substrate, with about 1/4 monolayer of octane in the contact region. Figure 57 shows snapshot pictures of the lubricant layer during retraction, as the block moves away from the substrate for three different block positions $d = 0$, 3 and 6 Å. Only the central part of the contact between the block and the substrate is shown, top view, after removing the block and substrate atoms. At the beginning ($d = 0$ Å) octane molecules are located in the substrate corrugation wells, or cavities with direct metal–metal contact between the block and the tops of the substrate nanoasperities (see figure 58). During retraction ($d = 3$ Å) the octane molecules are pulled out of the wells forming an almost symmetric network of nanobridges around the asperity tops, increasing the adhesion between the two surfaces. This configuration corresponds to the maximal adhesion force; see curve (d) in figure 56. Thus maximal adhesion is achieved via the formation of many small capillary nanobridges, involving just a few molecules for each bridge (see figure 58).

Further retraction ($d = 6$ Å) results in the collapse of the nanobridges and the formation of a single 'large' capillary bridge in the centre of the contact region.

8. Summary and outlook

Surface roughness has a huge influence on many common phenomena. It is the main reason that macroscopic bodies usually do not adhere to each other with any measurable strength. For example, if the floor and the soles of shoes were atomically smooth and clean, it would be impossible to walk. The (near) absence of adhesion in most situations is crucial for the function of many synthetic and natural constructions.

The surface to volume ratio of solid objects increases as the lateral size of the object decreases. The role of surface roughness therefore becomes more important as the size of the object decreases. The present drive toward the miniaturization of mechanical devices, e.g., microelectromechanical systems, requires a better understanding of the role of surface roughness in, e.g., contact mechanics and adhesion.

Surface roughness is also of great importance for the function of many biological systems. Flies, bugs, crickets and lizards have developed very soft layers on their feet which allow them to attach and move on both very smooth and rough vertical solid walls, e.g. stone walls or leaves. Another example is non-wetting coatings on plant surfaces based on surface roughness on many different length scales (the so-called lotus effect; see footnote 6).

The roughness of surfaces can nowadays be studied rather straightforwardly using standard equipment based on optical methods and on cantilever methods, e.g., the atomic force microscope (AFM). These methods cover the whole length scale from atomic dimension to macroscopic distances. Thus, the AFM can probe the surface profile from ∼1 nm to 100 μm and optical methods from ∼1 μm to kilometres. For randomly rough surfaces, the most important quantity which can be deduced from the measured height profile is the surface roughness power spectrum. We have shown throughout this paper how the roughness power spectrum determines the contact mechanics and adhesion for solid objects in direct contact. It also governs rubber friction on rough substrates, e.g., friction of tyres on a road surfaces, and influences other phenomena of technological importance, e.g., the roughness induced leaking of sealings.

Thus study and characterization of surface roughness is important not only for understanding many natural and biological phenomena, but also for many technological processes. The present drive toward miniaturization and the design of optimal systems by transfer of ideas from biology to materials science (bionics; see footnote 6)[15], is likely to accelerate the interest in and efforts to study and predict the influence of surface roughness on many phenomena.

Acknowledgments

We thank M Benz, K Rosenberg and J Israelachvili for providing unpublished experimental information about the surface topography of polymer films. The work at SISSA was sponsored through INFM PRA NANORUB, through MIUR COFIN 2003 and COFIN 2004, as well as FIRB RBAU01LX5H and FIRB RBAU017S8R, and by Regione Friuli Venezia Giulia. BP thanks the EC for a 'Smart QuasiCrystals' grant under the EC Programme 'Promoting Competitive and Sustainable GROWTH'.

[15] See http://finix.mpi-stuttgart.mpg.de/arzt/mf_mpg_abteilungen_e/mf_mpg_abteilungen_arzt/bio/topics.html for information involving surface roughness in biological 'applications'.

Appendix A. More about surface roughness

Assume that a randomly rough surface is described by the height profile $z = h(\boldsymbol{x})$, where $\boldsymbol{x} = (x, y)$ is a 2D vector in the surface plane $z = 0$, conveniently chosen so that $\langle h \rangle = 0$ and so that $\langle (h - \langle h \rangle)^2 \rangle$ is minimal. The statistical properties of randomly rough surfaces are completely specified when all the correlation functions

$$\langle h(\boldsymbol{x}_1)h(\boldsymbol{x}_2) \rangle, \qquad \langle h(\boldsymbol{x}_1)h(\boldsymbol{x}_2)h(\boldsymbol{x}_3) \rangle, \ldots,$$

are known. Here $\langle \cdots \rangle$ stands for ensemble averaging. In most cases one assumes that the correlation functions involving an odd number of h-functions vanish, while the correlation functions involving an even number of h-functions can be decomposed into a product of pair correlation functions; e.g., if we write $h(\boldsymbol{x}_1) = h_1$ and so on,

$$\langle h_1 h_2 h_3 h_4 \rangle = \langle h_1 h_2 \rangle \langle h_3 h_4 \rangle + \langle h_1 h_3 \rangle \langle h_2 h_4 \rangle + \langle h_1 h_4 \rangle \langle h_2 h_3 \rangle.$$

In this case the surface is completely specified by the surface roughness power spectrum $C(q)$, which is the Fourier transform of the pair correlation function $\langle h(\boldsymbol{x})h(\boldsymbol{0}) \rangle$, and the surface height distribution P_{h} is Gaussian (see below).

It is possible to generate surface roughness profiles, which are very similar to experimentally observed surface profiles, as follows. The surface height over a $L \times L$ square area can be expressed through its Fourier series:

$$h(\boldsymbol{x}) = \sum_{\boldsymbol{q}} B(\boldsymbol{q}) \mathrm{e}^{\mathrm{i}[\boldsymbol{q} \cdot \boldsymbol{x} + \phi(\boldsymbol{q})]} \tag{A.1}$$

where $\boldsymbol{q}$ spans all the vectors whose components are whole multiples of $2\pi/L$. Since $h(\boldsymbol{x})$ is real, $B(-\boldsymbol{q}) = B(\boldsymbol{q})$ and $\phi(-\boldsymbol{q}) = -\phi(\boldsymbol{q})$. If $\phi(\boldsymbol{q})$ are independent random variables, uniformly distributed in the interval $[0, 2\pi[$, then one can easily show that higher order correlation functions can be decomposed into a product of pair correlations in the way described above. Here we will demonstrate that in this case the height probability distribution P_{h} is always Gaussian, while any power spectrum can be arbitrarily imposed by choosing properly the amplitudes $B(\boldsymbol{q})$.

Let us consider a randomly rough surface described by equation (A.1). Provided that the phases $\phi(\boldsymbol{q})$ are uniformly distributed and independent, the statistical properties of the surface are translationally invariant; therefore,

$$\langle h(\boldsymbol{x}_1)h(\boldsymbol{x}_2) \rangle = C(\boldsymbol{x}_1 - \boldsymbol{x}_2).$$

The surface roughness power spectrum is defined by

$$C(\boldsymbol{q}) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2 x \, C(\boldsymbol{x}) \mathrm{e}^{-\mathrm{i}\boldsymbol{q} \cdot \boldsymbol{x}} \tag{A.2}$$

On substituting (A.1) in (A.2) and using

$$\langle \mathrm{e}^{\mathrm{i}\phi(\boldsymbol{q}')} \mathrm{e}^{\mathrm{i}\phi(\boldsymbol{q}'')} \rangle = \delta_{\boldsymbol{q}', -\boldsymbol{q}''}$$

it follows that

$$C(\boldsymbol{q}) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2 x \sum_{\boldsymbol{q}'} |B(\boldsymbol{q}')|^2 \mathrm{e}^{\mathrm{i}(\boldsymbol{q} - \boldsymbol{q}') \cdot \boldsymbol{x}} = \sum_{\boldsymbol{q}'} |B(\boldsymbol{q}')|^2 \delta(\boldsymbol{q} - \boldsymbol{q}').$$

When the sampling of the $\boldsymbol{q}$-space is dense enough we can approach the continuous limit by making the replacement

$$\sum_{\boldsymbol{q}} \to \frac{A}{(2\pi)^2} \int \mathrm{d}^2 q,$$

where A is the nominal surface area. This gives

$$C(\boldsymbol{q}) = \frac{A}{(2\pi)^2}|B(\boldsymbol{q})|^2.$$

Thus, if we choose

$$B(\boldsymbol{q}) = (2\pi/L)[C(\boldsymbol{q})]^{1/2},$$

where $L = A^{1/2}$, then the surface roughness profile (A.1) has the surface roughness power density $C(\boldsymbol{q})$. We can also guarantee that the statistical properties of the rough surface are isotropic by imposing $B(\boldsymbol{q}) = B(q)$; then $C(\boldsymbol{q}) = C(q)$ is a function of the magnitude $q = |\boldsymbol{q}|$, but not of the direction of $\boldsymbol{q}$. (The condition above shows the proportionality relation between the Fourier transform of the height correlation and the amplitude of the corresponding Fourier component of the surface profile. This result, known as the Wiener–Khintchine theorem, is further discussed in appendix C.)

Let us prove that surfaces whose correlation functions $\langle h_1 h_2 \dots h_n \rangle$ vanish for odd n while for even n they can be decomposed into a product of pair correlation functions have Gaussian height probability distributions. First note that the height probability distribution

$$\begin{aligned} P_{\mathrm{h}} = \langle \delta[h - h(\boldsymbol{x})] \rangle &= \frac{1}{2\pi} \int \mathrm{d}\alpha \, \langle \mathrm{e}^{\mathrm{i}\alpha[h - h(\boldsymbol{x})]} \rangle \\ &= \frac{1}{2\pi} \int \mathrm{d}\alpha \, \mathrm{e}^{\mathrm{i}\alpha h} \langle \mathrm{e}^{-\mathrm{i}\alpha h(\boldsymbol{x})} \rangle. \end{aligned} \tag{A.3}$$

But

$$\langle \mathrm{e}^{-\mathrm{i}\alpha h(\boldsymbol{x})} \rangle = \sum_n \frac{(-\mathrm{i}\alpha)^n}{n!} \langle [h(\boldsymbol{x})]^n \rangle.$$

However, $\langle [h(\boldsymbol{x})]^n \rangle$ vanishes for odd n. Thus, writing $n = 2m$ ($m = 0, 1, \dots$) gives

$$\langle \mathrm{e}^{-\mathrm{i}\alpha h(\boldsymbol{x})} \rangle = \sum_m \frac{(-\mathrm{i}\alpha)^{2m}}{(2m)!} \langle [h(\boldsymbol{x})]^{2m} \rangle. \tag{A.4}$$

We will now make use of the following equation:

$$\langle [h(\boldsymbol{x})]^{2m} \rangle = \frac{(2m)!}{m! 2^m} \langle h^2 \rangle^m. \tag{A.5}$$

The prefactor $(2m)!/m!2^m$ is easy to understand: when decomposing $\langle h_1 h_2 \dots h_{2m} \rangle$ there are $(2m)!$ possible orderings of $h_1, h_2, \dots, h_{2m}$. However, many of these products of pair correlation functions are identical. Thus, since in the pair correlation function $\langle h_i h_j \rangle$ the order of h_i and h_j is irrelevant, we must divide $(2m)!$ by 2^m. Furthermore, permutation of the m different pair correlation functions in a product term results in $m!$ identical terms, which is the origin of the $1/m!$ factor in (A.5). Substituting (A.5) into (A.4) gives

$$\langle \mathrm{e}^{-\mathrm{i}\alpha h(\boldsymbol{x})} \rangle = \sum_m \frac{(-\mathrm{i}\alpha)^{2m}}{m! 2^m} (\langle h^2 \rangle)^m = \mathrm{e}^{-\frac{1}{2}\alpha^2 \langle h^2 \rangle}.$$

Substituting this result into (A.3) and performing the integration over α gives

$$P_{\mathrm{h}} = \frac{1}{(2\pi)^{1/2}\sigma} \mathrm{e}^{-h^2/2\sigma^2},$$

where σ is the rms roughness amplitude, $\sigma^2 = \langle [h(\boldsymbol{x})]^2 \rangle$.

Figure A.1 shows a Gaussian random surface generated by the method discussed above, through formula (A.1). It is imposed that the power spectrum is proportional to $q^{-3.6}$ for large wavevectors, so the Hurst exponent is 0.8 and the fractal dimension is 2.2.

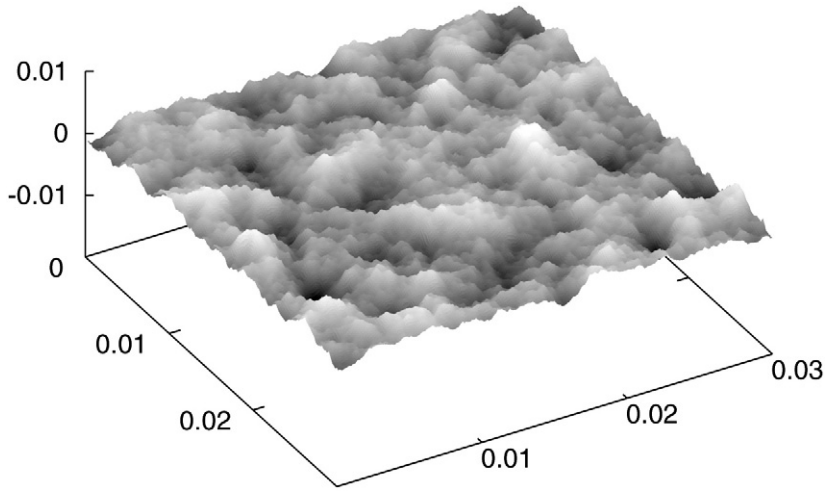

Figure A.1. A computer generated rough substrate. The surface has root mean square roughness 0.001 m, roll-off wavevector $q_0 = 1000\ \mathrm{m}^{-1}$ and Hurst exponent $H = 0.8$. Lengths in the picture are in metres.

Appendix B. The Hurst exponent and fractal dimension

A surface is said to be self-similar if it has the same statistical properties as a magnified version of itself. Thus, if $z = h(x, y)$ is the equation of a self-similar surface, then its magnified version $z = \lambda h(x/\lambda, y/\lambda)$ cannot be distinguished from the original. A self-affine surface is analogous, except that the magnification factor along the z direction differs from the in-plane magnification factor λ. In particular, the Hurst exponent H defines the scaling factor λ^H along the z direction for recovering the same statistical properties. The transformed surface $z = \lambda^H h(x/\lambda, y/\lambda)$ 'looks' exactly like the original one, $z = h(x, y)$. The Hurst exponent is in between 0 and 1, the latter being the case for a self-similar surface.

The general form of the power spectrum for an isotropic self-affine fractal surface is easy to derive. We have

$$C(q) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2x\ \langle h(\boldsymbol{x})h(\boldsymbol{0})\rangle \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}.$$

Now let us write $\boldsymbol{x} = \boldsymbol{x}'/\lambda$ so that

$$C(q) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2x'\ \lambda^{-2} \langle h(\boldsymbol{x}'/\lambda)h(\boldsymbol{0})\rangle \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}'/\lambda}.$$

But since for a self-affine fractal surface

$$\langle \lambda^H h(\boldsymbol{x}'/\lambda)\lambda^H h(\boldsymbol{0})\rangle = \langle h(\boldsymbol{x}')h(\boldsymbol{0})\rangle$$

we get

$$C(q) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2x'\ \lambda^{-2-2H} \langle h(\boldsymbol{x}')h(\boldsymbol{0})\rangle \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}'/\lambda}.$$

Thus, if we choose $\lambda = q$ and write $\hat{\boldsymbol{q}} = \boldsymbol{q}/q$ we get

$$C(q) = q^{-2(1+H)} \frac{1}{(2\pi)^2} \int \mathrm{d}^2x'\ \langle h(\boldsymbol{x}')h(\boldsymbol{0})\rangle \mathrm{e}^{-\mathrm{i}\hat{\boldsymbol{q}}\cdot\boldsymbol{x}'}.$$

Thus the power spectrum of a self-affine surface decreases as $q^{-2(H+1)}$ with increasing wavevector $q = |\boldsymbol{q}|$.

The Hurst exponent is directly related to the fractal dimension D_f through the formula $D_f = 3 - H$. The proof is straightforward. Consider a patch of the surface with extension $L \times L$ in the two in-plane directions x and y. The fractal dimension can be defined through the number of cubes of size Δ required to cover the surface completely: $N(\Delta) \propto \Delta^{-D_f}$ for $\Delta \to 0$. To cover the $L \times L$ area a number of cubes $(L/\Delta)^2$ are required; that is, the fractal dimension cannot be smaller than 2. In any $\Delta \times \Delta$ subdomain the z coordinate of the surface spans a range of values Δz proportional to Δ^H, because of the self-affine property. Actually $\Delta z = K(\Delta/L)^H$, where K is the range of values of z spanned by $h(x, y)$ over the whole $L \times L$ domain. Provided that Δ is small enough, the corresponding Δz becomes larger than Δ itself, since $H < 1$. Therefore, in any subdomain of size Δ we have to employ $K(\Delta/L)^H/\Delta \propto \Delta^{H-1}$ cubes, and the total number of cubes necessary to cover the surface is thus proportional to $\Delta^{-(3-H)}$, i.e.

$$D_f = 3 - H.$$

Appendix C. Moments of power spectra

We consider rough surfaces for which the statistical properties are isotropic and translationally invariant. Thus the surface roughness power spectrum

$$C(q) = \frac{1}{(2\pi)^2} \int d^2x \, \langle h(\boldsymbol{x} + \boldsymbol{x}')h(\boldsymbol{x}')\rangle e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \tag{C.1}$$

is independent of $\boldsymbol{x}'$ and of the orientation of the wavevector $\boldsymbol{q}$. If we express $h(\boldsymbol{x})$ through its Fourier transform

$$h(\boldsymbol{x}) = \int d^2q \, h(\boldsymbol{q}) e^{i\boldsymbol{q}\cdot\boldsymbol{x}} \tag{C.2}$$

and we substitute it into (C.1), we get

$$C(q) = \frac{1}{(2\pi)^2} \int d^2q' \int d^2q'' \int d^2x \, \langle h(\boldsymbol{q}')h(\boldsymbol{q}'')\rangle e^{i(\boldsymbol{q}'-\boldsymbol{q})\cdot\boldsymbol{x}} e^{i(\boldsymbol{q}'+\boldsymbol{q}'')\cdot\boldsymbol{x}'}.$$

Generally $h(\boldsymbol{q})$ is not an ordinary function but is a distribution; thus $\langle h(\boldsymbol{q}')h(\boldsymbol{q}'')\rangle$ can diverge. However, we can replace $h(\boldsymbol{q})$ with $h_A(\boldsymbol{q})$, the Fourier transform of $h(\boldsymbol{x})$ restricted to a square area A, i.e.,

$$h_A(\boldsymbol{q}) = \frac{1}{(2\pi)^2} \int_{-\sqrt{A}/2}^{+\sqrt{A}/2} dx_1 \int_{-\sqrt{A}/2}^{+\sqrt{A}/2} dx_2 \, h(\boldsymbol{x}) e^{-i\boldsymbol{q}\cdot\boldsymbol{x}},$$

and the equations above still hold in the limit $A \to \infty$:

$$C(q) = \frac{1}{(2\pi)^2} \int d^2q' \int d^2q'' \int d^2x \, \langle h_A(\boldsymbol{q}')h_A(\boldsymbol{q}'')\rangle e^{i(\boldsymbol{q}'-\boldsymbol{q})\cdot\boldsymbol{x}} e^{i(\boldsymbol{q}'+\boldsymbol{q}'')\cdot\boldsymbol{x}'}.$$

To simplify this equation we use the standard relation

$$\int d^2x \, e^{i\boldsymbol{q}\cdot\boldsymbol{x}} = (2\pi)^2 \delta(\boldsymbol{q}) \tag{C.3}$$

to get

$$C(q) = \int d^2q'' \, \langle h_A(\boldsymbol{q})h_A(\boldsymbol{q}'')\rangle e^{i(\boldsymbol{q}+\boldsymbol{q}'')\cdot\boldsymbol{x}'}. \tag{C.4}$$

This expression is independent of $\boldsymbol{x}'$, at least in the limit $A \to \infty$; hence $\langle h(\boldsymbol{q})h(\boldsymbol{q}'')\rangle = \lim_{A\to\infty}\langle h_A(\boldsymbol{q})h_A(\boldsymbol{q}'')\rangle$ must vanish when $\boldsymbol{q} + \boldsymbol{q}'' \neq \boldsymbol{0}$. Moreover, the isotropy implies that $\langle h(\boldsymbol{q})h(-\boldsymbol{q})\rangle$ depends on $\boldsymbol{q}$ only through its modulus q; therefore,

$$\langle h(\boldsymbol{q})h(\boldsymbol{q}')\rangle = C(q)\delta(\boldsymbol{q} + \boldsymbol{q}'). \tag{C.5}$$

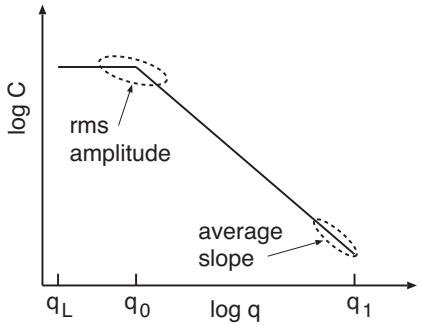

Figure C.1. Surface roughness power spectra of a surface which is self-affine fractal for $q_1 > q > q_0$. The rms roughness amplitude and the average slope (and the average curvature) are determined mainly by the encircled regions of the power spectrum.

Since (C.4) is independent of $\boldsymbol{x}'$ we may integrate over $\boldsymbol{x}'$ and divide by the surface area A to get

$$C(q) = \frac{1}{A}\int \mathrm{d}^2x' \int \mathrm{d}^2q''\, \langle h_A(\boldsymbol{q})h_A(\boldsymbol{q}'')\rangle \mathrm{e}^{\mathrm{i}(q+q'')\cdot x'}$$
$$= \frac{(2\pi)^2}{A}\,\langle h(\boldsymbol{q})h(-\boldsymbol{q})\rangle,$$

where we have again used (C.3). Since $h(\boldsymbol{x})$ is real, $h_A(-\boldsymbol{q}) = h_A^*(\boldsymbol{q})$, and we can write

$$C(q) = \frac{(2\pi)^2}{A}\langle |h_A(\boldsymbol{q})|^2\rangle. \tag{C.6}$$

Equations (C.5) and (C.6) are very important and useful equations of general validity. As an application, let us calculate the average surface slope. Using (C.2) we get

$$\langle (\nabla h)^2\rangle = \int \mathrm{d}^2q\, \mathrm{d}^2q'(-\boldsymbol{q}\cdot\boldsymbol{q}')\langle h(\boldsymbol{q})h(\boldsymbol{q}')\rangle \mathrm{e}^{\mathrm{i}(q+q')\cdot x}.$$

Using (C.5) this gives

$$\langle (\nabla h)^2\rangle = \int \mathrm{d}^2q\, q^2 C(q) = 2\pi \int \mathrm{d}q\, q^3 C(q).$$

For a self-affine fractal surface $C(q) \sim q^{-2(1+H)}$ and

$$\langle (\nabla h)^2\rangle \sim \int_{q_0}^{q_1} \mathrm{d}q\, q^{1-2H} \sim q_1^{2(1-H)} - q_0^{2(1-H)} \approx q_1^{2(1-H)}$$

if $q_1 \gg q_0$. Thus the average slope (and the average curvature) is determined by the shortest wavelength roughness components, as indicated in figure C.1. On the other hand, as shown in section 2, the rms roughness amplitude is determined mainly by the longest surface roughness wavelength components, i.e., the region around $q \sim q_0$ (see figure C.1).

Appendix D. Numerical recipes for calculating power spectra

Here we describe how the roughness power spectrum $C(q)$ can be numerically determined from the surface height $h(\boldsymbol{x})$ measured over a square area $0 < x < L$ and $0 < y < L$.

The surface roughness power spectrum is given by

$$C(\boldsymbol{q}) = \frac{1}{(2\pi)^2}\int \mathrm{d}^2x\, \langle h(\boldsymbol{x})h(\boldsymbol{0})\rangle \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}},$$

which can also be written as (see appendix C)

$$C(\boldsymbol{q}) = \frac{(2\pi)^2}{A}\langle |h_A(\boldsymbol{q})|^2\rangle,$$

where $A = L^2$ is the surface area under study and

$$h_A(\boldsymbol{q}) = \frac{1}{(2\pi)^2}\int_A \mathrm{d}^2x\, h(\boldsymbol{x})\mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}.$$

In order to calculate $C(q)$ numerically from the measured height profile $h(\boldsymbol{x})$, we assume that the measurement points form a square mesh with the lattice constant a, and with N data points along the x and y coordinates. In other words, $h(\boldsymbol{x})$ is sampled and its values are known only at the points

$$\boldsymbol{x} = (n_x, n_y)a = \boldsymbol{n}a = \boldsymbol{x}_{\boldsymbol{n}},$$

where $n_x = 1, 2, \ldots, N$ and $n_y = 1, 2, \ldots, N$ are integers. The integral for the Fourier transform h_A can be approximated by a discrete sum:

$$h_A(\boldsymbol{q}) \approx \frac{a^2}{(2\pi)^2}\sum_{\boldsymbol{n}} h_{\boldsymbol{n}}\mathrm{e}^{-\mathrm{i}(q_x n_x a + q_y n_y a)}$$

where $h_{\boldsymbol{n}} = h(\boldsymbol{x}_{\boldsymbol{n}})$. The Fourier transform can be sampled too, by using a grid with lattice spacing $q_L = 2\pi/L = 2\pi/Na$, since the finite size of the surface limits the resolution in the $\boldsymbol{q}$-space. Thus $\boldsymbol{q} = (q_x, q_y) = (2\pi m_x/L, 2\pi m_y/L)$ where m_x and m_y are integer numbers between 0 and $N-1$ (any larger value of m_x or m_y corresponds to wavelengths shorter than a, and does not carry extra information because of the finite resolution in the real space). Now we can write

$$h_A(\boldsymbol{q}) \approx \frac{a^2}{(2\pi)^2}\sum_{\boldsymbol{n}} h_{\boldsymbol{n}}\mathrm{e}^{-\mathrm{i}2\pi N(m_x n_x + m_y n_y)} = \frac{a^2}{(2\pi)^2}H_{\boldsymbol{m}}, \tag{D.1}$$

where $H_{\boldsymbol{m}}$ is the two-dimensional *discrete Fourier transform* of $h_{\boldsymbol{n}}$. The big advantage of evaluating the power spectrum through the Fourier transformation of $h(\boldsymbol{x})$, instead of carrying on with the calculation of the correlation, is that there exists an efficient algorithm for the discrete Fourier transformation: the well known fast Fourier transform (FFT) method. Indeed the computational time scales as $N^2\log(N)$ (in two dimensions), while the convolution integral for estimating the two-dimensional correlation would imply a computational time that scales as N^4.

A crude result can be obtained by substituting (D.1) into (C.6): the power spectrum obtained so far would be identical to the one resulting from the Fourier transformation of the correlation, provided that the correlation is calculated applying periodic boundary conditions over the surface of area A. But real surfaces are portions of larger samples, where there is no match at the boundary: the periodic boundary conditions introduce discontinuities, corresponding to spurious Fourier components at the larger wavevectors of the spectrum. Therefore it is necessary to filter them out. A typical solution consists in multiplying the height profile $h_{\boldsymbol{n}}$ by a *windowing function* that goes gradually to 0 at the edge of the surface area and is almost constant around the centre; and then carrying on with the fast Fourier transformation with the adjusted data. This procedure is described in detail in [81].

Note also that from

$$\langle h(\boldsymbol{x})h(\boldsymbol{0})\rangle = \int \mathrm{d}^2q\, C(\boldsymbol{q})\mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}$$

it follows that

$$\langle h^2\rangle = \int \mathrm{d}^2q\, C(\boldsymbol{q}). \tag{D.2}$$

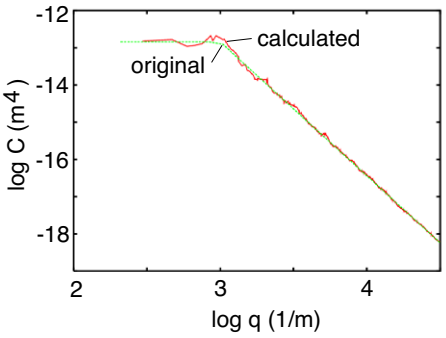

Figure D.1. Red line: the calculated power spectrum for the surface shown in figure A.1. Green line: the original power spectrum, employed to generate the surface. The surface extends over an area of 0.03 m × 0.03 m and the sampling length is $a = 0.0001$ m.

But the mean of the square of the surface roughness amplitude can also be obtained directly from the experimental data via

$$\langle h^2 \rangle = \frac{1}{N^2} \sum_{\boldsymbol{n}} (h_{\boldsymbol{n}} - \bar{h})^2. \tag{D.3}$$

It is a good check of the numerical accuracy in the evaluation of the surface roughness power spectra to show that the $\langle h^2 \rangle$ calculated from (D.2) and (D.3) gives identical results.

Let us assume that the surface roughness is isotropic. In this case $C(\boldsymbol{q})$ is independent of the direction of $\boldsymbol{q}$, i.e., $C(\boldsymbol{q}) = C(q)$. We can make use of this fact to reduce the statistical noise in $C(q)$ by performing an angular average in (D.2). Let us for simplicity write $C(\boldsymbol{q}) = C(\boldsymbol{m}q_L)$ as $C(m_x, m_y)$, where $q_L = 2\pi/L$ is the lattice spacing in the $\boldsymbol{q}$-space, $m_x = 0, 1, \ldots, N-1$, and similarly for m_y. In an analogous way we write $C(q) = C(mq_L)$ as $C(m), m = 1, 2, \ldots, N/2$. We have

$$\int \mathrm{d}^2 q\, C(\boldsymbol{q}) = 2\pi \int_0^{q_1} \mathrm{d}q\, qC(q). \tag{D.4}$$

The discretized version of (D.4) takes the form

$$q_L^2 \sum_{m_x=0}^{N/2} \sum_{m_y=0}^{N/2} \bar{C}(m_x, m_y) = 2\pi q_L^2 \sum_m mC(m) \tag{D.5}$$

where

$$\bar{C}(m_x, m_y) = C(m_x, m_y) + C(N - m_x, m_y) + C(m_x, N - m_y) + C(N - m_x, N - m_y)$$

corresponds to $C(q_x, q_y)+C(-q_x, q_y)+C(q_x, -q_y)+C(-q_x, -q_y)$ (remember that the discrete Fourier transform is periodic with period N). Equation (D.5) is obeyed if we define

$$C(m) = \frac{1}{2\pi m} \sum_{-1/2<|m'|-m<1/2} \bar{C}(m'_x, m'_y).$$

With this definition the 'sum rule'

$$\langle h^2 \rangle = 2\pi \int \mathrm{d}q\, qC(q) = 2\pi q_L^2 \sum_{m=1}^{N/2} mC(m)$$

will be exactly obeyed.

As an application, in figure D.1 we show the spectrum calculated with this method for the same surface as described in figure A.1. The agreement between the original spectrum and the calculated one provides evidence of the accuracy that can be achieved.

References

[1] Zhao Y-P, Wang L S and Yu T X 2003 *J. Adhes. Sci. Technol.* **17** 519
[2] Mastrangelo C H 1997 *Trib. Lett.* **3** 223
[3] Bico J, Marzolin C and Quere D 1999 *Europhys. Lett.* **47** 220
[4] Gui C, Elwenspoek M, Tas N and Gardeniers J G E 1999 *J. Appl. Phys.* **85** 7448
[5] Barabasi A-L and Stanley H E 1995 *Fractal Concepts in Surface Growth* (Cambridge: Cambridge University Press)
[6] Krug J 1997 *Adv. Phys.* **46** 139
[7] Feder J 1988 *Fractals* (New York: Plenum)
Berry M V and Lewis Z V 1980 *Proc. R. Soc.* A **370** 459
[8] Krim J and Palasantzas G 1995 *Int. J. Mod. Phys.* B **9** 599
[9] Nayak P R 1971 *ASME J. Lubr. Technol.* **93** 398
[10] Bouchaud E 1997 *J. Phys.: Condens. Matter* **9** 4319
[11] Experimental data obtained at Pirelli, unpublished data
[12] Benz M, Rosenberg K and Israelachvili J 2004 private communication
[13] Persson B N J 2000 *Sliding Friction: Physical Principles and Applications* 2nd edn (Heidelberg: Springer)
[14] Hertz H and Reine J 1882 *Angew. Math.* **92** 156
[15] Archard J F 1957 *Proc. R. Soc.* A **243** 190
[16] Greenwood J A 1992 *Fundamentals of Friction, Macroscopic and Microscopic Processes* ed I L Singer and H M Pollack (Dordrecht: Kluwer)
[17] Greenwood J A and Williamson J B P 1966 *Proc. R. Soc.* A **295** 300
[18] Johnson K L 1985 *Contact Mechanics* (Cambridge: Cambridge University Press)
[19] Bush A W, Gibson R D and Thomas T R 1975 *Wear* **35** 87
Bush A W, Gibson R D and Keogh G P 1976 *Mech. Res. Commun.* **3** 169
[20] Persson B N J 2001 *Phys. Rev. Lett.* **87** 116101
Persson B N J 2001 *J. Chem. Phys.* **115** 3840
[21] Persson B N J, Bucher F and Chiaia B 2002 *Phys. Rev.* B **65** 184106
[22] Persson B N J 2002 *Eur. Phys. J.* E **8** 385
Persson B N J 2002 *Phys. Rev. Lett.* **89** 245502
[23] Hyun S, Pei L, Molinari J-F and Robbins M O 2004 *Phys. Rev.* E **70** 026177
[24] Borri-Brunetto M, Chiaia B and Ciavarella M 2001 *Comput. Methods Appl. Mech. Eng.* **190** 6053
[25] Persson B N J, Albohr O, Mancosu F, Peveri V, Samoilov V N and Sivebaek I M 2003 *Wear* **254** 835
[26] Buzio R, Boragno C, Biscarini F, de Mongeot F B and Valbusa U 2003 *Nat. Mater.* **2** 233
[27] Persson B N J, Albohr O, Creton C and Peveri V 2004 *J. Chem. Phys.* **120** 8779
[28] Hui C Y, Lin Y Y and Baney J M 2000 *J. Polym. Sci.* B **38** 1485
[29] Creton C and Leibler L 1996 *J. Polym. Sci.* B **34** 545
[30] Chikina I and Gay C 2000 *Phys. Rev. Lett.* **85** 4546
[31] Hammond F H 1964 *ASTM Special Techn. Publ.* **360** 123
[32] Thomas T R 1999 *Rough Surfaces* 2nd edn (London: Imperial College Press)
[33] Stauffer D and Aharony A 1992 *Introduction to Percolation Theory* 2nd edn (London: Taylor and Francis)
[34] See, e.g. Feynman R P 2001 *What Do You Care What Other People Think?: Further Advances of a Curios Character* (New York: W.W. Norton & Company)
[35] Hertz D L 1979 *Machine Design* April 12. Available on line:
http://www.sealseastern.com/PDF/LowPsiSeals.pdf
[36] van der Oetelaar R J A and Flipse C F J 1997 *Surf. Sci.* **384** L828
[37] Moore D F 1972 *The Friction and Lubrication of Elastomers* (Oxford: Pergamon)
[38] Persson B N J 1998 *Surf. Sci.* **401** 445
[39] Galliano A, Bistac S and Schultz J 2003 *J. Colloid Interface Sci.* **265** 372
Galliano A, Bistac S and Schultz J 2003 *J. Adhes.* **79** 973
[40] Grosch K A 1974 *The Physics of Tire Traction: Theory and Experiment* ed D F Hays and A L Browne (New York: Plenum) p 143
[41] Persson B N J and Volokitin A I 2002 *Phys. Rev.* B **65** 134106
[42] Klüppel M and Heinrich G 2000 *Rubber Chem. Technol.* **73** 578

[43] Williams M L, Landel R F and Ferry J D 1955 *J. Am. Chem. Soc.* **77** 3701
[44] Haney P 2003 *The Racing and High-Performance TIRE* (Springfield, IL: TV Motorsport) (Warrendale, PA: Society of Automotive Engineering, Inc.)
[45] Westermann S, Petry F and Thielen G 2004 *Safety-Wet Grip-Traction* (Hannover: Deutsches Institut für Kautschuktechnologie eV)
[46] Lindner M, Kröger M, Popp K and Blume H 2004 *21st Int. Cong. on Theoretical and Applied Mechanics (Warsaw, Poland, Aug. 2004)* Online at http://fluid.ippt.gov.pl/ictam04/text/sessions/docs/SM2/12086/SM2_12086.pdf
[47] Heumer T, Eberhardstelner J, Liu W N and Mang H A 1999 *1st European Conf. on Constitutive Models for Rubber (Vienna, Austria, Sept. 1999)* Online at http://www.imws.tuwien.ac.at/Publications/presentations/uebersicht.html
[48] Westermann S, Petry F, Boes R and Thielen G 2004 Experimental investigations into the predictive capabilities of current physical rubber friction theories *Kautschuk Gummi Kunststoffe* at press
[49] Westermann S, Petry F and Thielen G 2004 *Tyre Technology Expo 2004 (Stuttgart)* (*Tyre Technology Annual 2004*) (Surrey: International Press Events) at press
[50] Persson B N J 2003 unpublished
[51] Tire tests performed at Pirelli, unpublished
[52] Heinrich G 2003 *Elastomerreibung und Kontactmechanik* (Hannover, Germany: Deutsches Institute für Kautschuktechnologie)
[53] Meyer W E and Walter J D 1983 *Frictional Interaction of Tire and Pavement* STP 793 (Philadelphia, PA: American Society for Testing and Materials) p 85 (ISBN: 0803102313)
[54] Persson B N J, Tartaglino U, Albohr O and Tosatti E 2004 *Nat. Mater.* at press
Persson B N J, Tartaglino U, Albohr O and Tosatti E 2004 *Phys. Rev.* B submitted
[55] Roth J 1993 *Fortschritt-Berichte VDI Reihe* vol 12, No 195 (Düsseldorf: VDI-Verlag)
[56] Brochard-Wyart F and de Gennes P G 1994 *J. Phys.: Condens. Matter* **6** A9
[57] Martin P and Brochard-Wyart F 1998 *Phys. Rev. Lett.* **80** 3296
[58] Persson B N J, Volokitin A I and Tosatti E 2003 *Eur. Phys. J.* E **11** 409
[59] Persson B N J and Mugele F 2004 *J. Phys.: Condens. Matter* **16** R295
[60] Fuller K N G and Tabor D 1975 *Proc. R. Soc.* A **345** 327
[61] Kendall K 2001 *Molecular Adhesion and it Applications* (New York: Kluwer)
Maugis D 1999 *Contact, Adhesion and Rupture of Elastic Solids* (Berlin: Springer)
[62] Kendall K 1971 *J. Phys. D: Appl. Phys.* **4** 1186
Kendall K 1973 *J. Phys. D: Appl. Phys.* **6** 1782
Kendall K 1975 *J. Phys. D: Appl. Phys.* **8** 115
See also the beautiful review article of Kendall K 1980 *Contemp. Phys.* **21** 277
[63] Briggs G A D and Briscoe B J 1977 *J. Phys. D: Appl. Phys.* **10** 2453
[64] Fuller K N G and Roberts A D 1981 *J. Phys. D: Appl. Phys.* **14** 221
[65] Zilberman S and Persson B N J 2003 *J. Chem. Phys.* **118** 6473
[66] Samoilov V N, Sievebaek I M and Persson B N J 2004 *J. Chem. Phys.* **121** 9639
[67] Israelachvili J N 1995 *Intermolecular and Surface Forces* (London: Academic)
[68] Johnson K L, Kendall K and Roberts A D 1971 *Proc. R. Soc.* A **324** 301
[69] Sperling G 1964 *PhD Thesis* Karlsruhe Technical University
[70] Peressadko A and Gorb S N 2004 *J. Adhes.* **80** 247
[71] Johnson K L 1998 *Tribol. Int.* **31** 413
Zhang L and Zhao Y-P 2004 *J. Adhes. Sci. Technol.* **18** 715
[72] McFadden C F and Gellman A J 1995 *Tribol. Lett.* **1** 201
[73] Persson B N J 2003 *Wear* **254** 832
[74] Gao H J and Yao H M 2004 *Proc. Natl Acad. Sci. USA* **101** 7851
[75] Scherge M and Gorb S 2001 *Biological Micro- and Nano-Tribology* (Berlin: Springer)
[76] Autumn K, Sitti M, Liang Y A, Peattie A M, Hansen W R, Sponberg S, Kenny T W, Fearing R, Israelachvili J N and Full R J 2002 *Proc. Natl Acad. Sci. USA* **99** 12252
Autumn K, Liang Y A, Hsieh S T, Zesch W, Chan W P, Kenny T W, Fearing R and Full R J 2000 *Nature* **405** 681
[77] Persson B N J 2003 *J. Chem. Phys.* **118** 7614
[78] Persson B N J and Gorb S 2004 *J. Chem. Phys.* **119** 11437
[79] Carbone G and Persson B N J 2004 *Phys. Rev.* B at press
[80] Geim A K, Dubonos S V, Grigorieva I V, Novoselov K S, Zhukov A A and Shapoval S Yu 2003 *Nature* **2** 461
[81] Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1992 *Numerical Recipes in Fortran* 2nd edn (Cambridge: Cambridge University Press)