Chapter 6
Friction, Wear, and Self-Lubrication

The second part of this book covers self-lubrication and various issues related to the reduction of friction. Friction and wear during sliding or rolling of solid surfaces are universal phenomena and they reflect the tendencies of energy to dissipate and material to deteriorate, which are consequences of the Second law of thermodynamics. In general, solid surfaces in relative motion require lubrication, which dramatically reduces the extent of friction and wear. The situation when no external lubrication is required is called self-lubrication. There are many mechanisms of self-lubrication ranging from coatings to embedding lubricant into the matrix of a composite material, to self-organized in situ tribofilms, and to biological surfaces.

The study of friction, wear, and lubrication, as well as other phenomena related to the contact of solid surfaces in relative motion is referred to as Tribology. In this chapter, we introduce the general concepts of Tribology including the principles mechanical contact of rough surfaces, friction, wear, and lubrication. After that we discuss current advances in self-lubricating materials. Friction-induced self-organization and tribology of metal matrix composites are discussed in the consequent chapters.

6.1 Friction and Wear as Manifestations of the Second Law of Thermodynamics

Friction is resistance to the relative lateral motion of solid surfaces, fluid layers, or material elements in contact. Friction is a universal phenomenon and it reflects the general trend of energy to dissipate. In the ideal case of the contact of absolutely rigid bodies and conservative forces acting among them, there would be no energy dissipation. However, any nonideality, such as surface roughness and deformation tends to lead to dissipation. Therefore, friction has a fundamental nature and reflects the general trend of energy to be dissipated, as expressed in the Second law

M. Nosonovsky and P.K. Rohatgi, *Biomimetics in Materials Science: Self-Healing, Self-Lubricating, and Self-Cleaning Materials*, Springer Series in Materials Science 152, DOI 10.1007/978-1-4614-0926-7_6,

Table 6.1 Dissipation and friction mechanisms corresponding to different hierarchy levels (based on Nosonovsky and Bhushan 2008a)

Ideal situation (reversible process)	Real situation (irreversible process)	Mechanism of dissipation leading to friction	Friction mechanism	Hierarchy level
Nonadhesive surfaces	Chemical interaction between surfaces is possible	Breaking chemical adhesive bonds	Adhesion	Molecule
Conservative adhesive forces	Conservative (van der Waals) forces and nonconservative (chemical) bonds	Breaking chemical adhesive bonds	Adhesion	Molecule
Rigid material	Deformable (elastic and plastic) material	Radiation of elastic waves (phonons)	Adhesion	Surface
Smooth surface	Rough surface	Plowing, ratchet mechanism, cobblestone mechanism	Deformation, ratchet, cobblestone mechanisms	Asperity
Homogeneous surface	Inhomogeneous surface	Energy dissipation due to inhomogeneity	Adhesion	Surface

of thermodynamics. Despites the apparent simplicity of the dry friction and its relation to the Second law of thermodynamics, there is no simple thermodynamic theory of friction. The reason is that there are many mechanisms of dry friction, and these mechanisms have to be considered separately. Instead, there are so-called empirical laws or rules of dry friction, which were formulated several hundred years ago by Leonardo da Vinci (1452–1519), Guillaume Amontons (1663–1705), and Charles-Augustin de Coulomb (1736–1806) and are commonly referred to as the "Amontons–Coulomb laws" or just "the Coulomb law" (sometimes also "the Amontons law"). Frictional mechanisms are summarized in Table 6.1.

Similarly to friction, wear reflects the tendency of matter to deteriorate irreversibly, which is another consequence of the Second law of thermodynamics. Again, there is no single quantitative "law" of wear, since there are several physical and chemical processes which lead to the surface deterioration. Instead, there are empirical laws or rules, which relate the wear rate (the volume of worn material per second) with the normal load at the interface. Most models of friction and wear are based on the models of mechanical contact of rough surfaces.

The third big area of Tribology, in addition to the friction and wear, is the lubrication. Lubrication is applied to the interface in order to reduce friction and wear and thus the lubrication cannot be studied separately from friction and wear. Since our interest is in self-lubrication or the ability of the interface to achieve low friction and wear without the external supply of lubrication, in this chapter, we will have to discuss first theories of contact of rough surfaces, then modern theories of friction and wear, and only after that we will be able to review modern approaches to self-lubrication.

6.2 Contact of Rough Solid Surfaces

All natural and artificial solid surfaces contain irregularities irrespective of the method of their formation. No machining method can produce a molecularly smooth surface using conventional materials. Even the smoothest surfaces, obtained by cleavage of some crystals (such as graphite or mica), contain irregularities, heights of which exceed interatomic distances.

Engineered surfaces typically have several types of deviation from the prescribed form: the waviness, random roughness, and flow. The waviness may result from machine vibration or chatter during machining as well as the heat treatment or warping strains. It includes irregularities with a relatively long (many microns) wavelength. Roughness is formed by fluctuation of the surface of short wavelengths, characterized by asperities (local maxima of the surface height) and valleys (local minima of the surface height). Flows are unintentional, unexpected, and unwanted interruptions in the texture.

In addition to surface irregularities, the engineering solid surface itself consists of several zones or layers, such as the chemisorbed layer (0.3 nm), physisorbed layer (0.3–3 nm), chemically reacted layer (10–100 nm), etc. In the chemisorbed layer, the solid surface bonds to the adsorption species through covalent bonds with an actual sharing of electrons. In the physisorbed layer, there are no chemical bonds between the substrate and the adsorbent, and only van der Waals force is involved. The van der Waals force is relatively weak (under 10 kJ mol^{-1}) and long range (nanometers) as opposed to the strong (40–400 kJ mol^{-1}) and short-range (comparable with the interatomic distance of about 0.3 nm) chemical bonding. Typical adsorbents are oxygen, water vapor, or hydrocarbons from the environment, which can condense at the surface. While chemisorbed layer is usually a monolayer, the physisorbed layer may include several layers of the molecules. The chemically reacted is significantly thicker and involves many layers of molecules. The typical example of the chemically reacted layer is the oxide layer at the surface of a metallic substrate (Bhushan 2002; Nosonovsky and Bhushan 2008a).

When two rough solid surfaces come into a mechanical contact, the real (or "true") area of contact is usually small in comparison with the nominal (or "apparent") area of contact, because the contact takes place only at the tops of the asperities. In most cases, only the highest asperities participate in the contact. This makes the dependency of the real area of contact, A_r, and the total number of contact spots, N, upon the roughness parameters during the elastic contact, an almost linear function.

There are several quantitative parameters, which are commonly used to characterize random solid surface roughness, that is, a random derivation from the nominal (prescribed) shape. These are the amplitude (or height) parameters and the spatial (related to the length of asperities) parameters (Thomas 1982; Bhushan 2002). The most commonly used amplitude parameter is the root mean square (RMS) or the standard deviation from the center-line average. For a 2D roughness

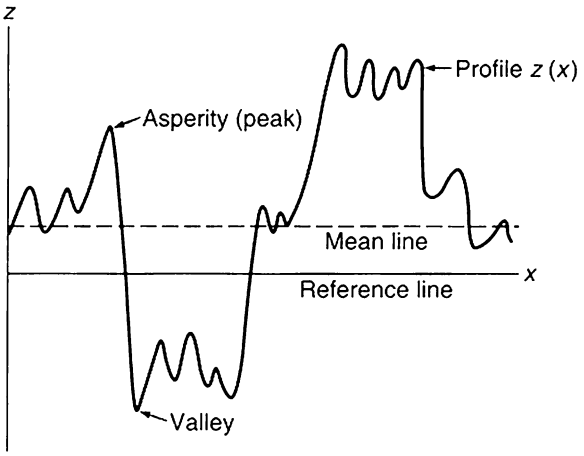

Fig. 6.1 Schematics of a rough surface profile (based on Nosonovsky and Bhushan 2008a)

profile $z(x)$, the center-line average is defined as the arithmetic mean of the absolute value of the vertical deviation from the mean line of the profile (Fig. 6.1).

$$R_{\mathrm{a}} = \frac{1}{L}\int_0^L |z - m| \mathrm{d}x, \tag{6.1}$$

where L is the sampling length and m is the mean.

$$m = \frac{1}{L}\int_0^L z\,\mathrm{d}x. \tag{6.2}$$

The square RMS is given by

$$\sigma^2 = \frac{1}{L}\int_0^L (z - m)^2 \mathrm{d}x. \tag{6.3}$$

Since different rough surface profiles can have same RMS, additional parameters are required to characterize details of surface profile. The cumulative probability distribution function, $P(h)$ associated with the random variable $z(h)$, is defined as the probability of the event that $z(x) < h$, and is written as

$$P(h) = \mathrm{Probability}(z < h). \tag{6.4}$$

It is common to describe the probability structure of random data in terms of the slope of the distribution function, known as the probability density function (PDF) and given by the derivative

$$p(z) = \frac{\mathrm{d}P(z)}{\mathrm{d}z}. \tag{6.5}$$

The integral of the PDF is equal to $P(z)$ and the total area under the PDF must be unity (Bhushan 2002).

In many practical cases, the random data tend to have the so-called Gaussian or normal distribution with the PDF given by

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-m)^2}{2\sigma^2}\right), \tag{6.6}$$

where m is the mean and σ is the standard deviation. For convenience, the Gaussian function is often plotted in terms of the normalized variable $z^* = (z - m)/\sigma$ as

$$p(z^*) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^{*2}}{2}\right) \tag{6.7}$$

(Bhushan 2002). The Gaussian distribution is found in nature and technical applications when the random quantity is a sum of many random factors acting independently of each other. When an engineering surface is formed, many random factors contribute into the roughness, and thus in many cases roughness height is governed by the Gaussian distribution. Such surfaces are called Gaussian surfaces (Nosonovsky and Bhushan 2008a).

In order to represent spatial distribution of random roughness, the autocorrelation function, which is defined as

$$C(\tau) = \lim_{L\to\infty} \frac{1}{\sigma^2 L} \int_0^L [z(x) - m][z(x+\tau) - m]\,\mathrm{d}x. \tag{6.8}$$

The autocorrelation function characterizes the correlation between two measurements taken at the distance τ apart, $z(x)$, and $z(x + \tau)$. It is obtained by comparing the function $z(x)$ with a replica of itself shifted for the distance τ. The function $C(\tau)$ approaches zero if there is no statistical correlation between values of z separated by the distance τ; in the opposite case $C(\tau)$ is different from zero. Many engineering surfaces are found to have an exponential autocorrelation function

$$C(\tau) = \exp\left(\frac{-\tau}{\beta}\right), \tag{6.9}$$

where β is the parameter called the correlation length or the length over which the autocorrelation function drops to a small fraction of its original value. At the distance β, the autocorrelation function falls to the $1/e$. In many cases, the value $\beta^* = 2.3\beta$ is used for the correlation length, at which the function falls to 10% of its original value (Bhushan 2002).

For a Gaussian surface with the exponential autocorrelation function, σ and β^* are two parameters of the length dimension which conveniently characterize the roughness. While σ is the height parameter, which characterizes the height of a typical roughness detail (asperity), β^* is the length parameter, which characterizes the length of the detail. The average absolute value of the slope is proportional to the ratio σ/β^*, whereas the average curvature is proportional to β^*/σ^2. For a Gaussian surface, σ is related to the RMS as $\sigma = \left(\sqrt{\pi/2}\right)R_a$ (Bhushan 2002). These two parameters, σ and β^*, are convenient for characterization of many random surfaces. Note that a Gaussian surface has only one inherent length scale parameter, β^*, and one vertical length scale parameter, σ, and thus it cannot describe the multiscale roughness.

When two rough surfaces come into a mechanical contact, the real area of contact is small in comparison with the nominal area of contact, because the contact takes place only at the tops of the asperities. For two rough surfaces in contact, an equivalent rough surface can be defined of which the values of the local heights, slopes, and local curvature are added to each other. The composite standard deviation of profile heights is related to those of the two rough surfaces, σ_1 and σ_2 as

$$\sigma^2 = {\sigma_1}^2 + {\sigma_2}^2. \tag{6.10}$$

The composite correlation length is related to those of the two rough surfaces, β_1^* and β_2^* as

$$\frac{1}{\beta^*} = \frac{1}{\beta_1^*} + \frac{1}{\beta_2^*}. \tag{6.11}$$

Using the composite rough parameters allows to effectively reduce the contact problem of two rough surfaces to the contact of a composite rough surface with a flat surface (Bhushan 1999, 2002).

Two parameters of interest during the elastic and plastic contact of two rough surfaces are the real area of contact, A_r, and the total number of contact spots, N. In most cases, only the highest asperities participate in the contact. This allows to linearize the dependence of A_r and N upon the roughness parameters during the elastic contact as

$$A_r \propto \frac{W\beta^*}{\sigma E}, \tag{6.12}$$

$$N \propto \frac{W}{\sigma\beta^*}, \tag{6.13}$$

where W is the normal load force and E is the composite elastic modulus. Qualitatively, the higher the asperities, the larger is σ and smaller is A_r, the wider the asperities, the larger is β^* and smaller is A_r. The larger and wider the asperities, the smaller is A_r

(Bhushan and Nosonovsky 2003, 2004a, b). For plastic contact, N, which depends upon the contact topography and thus is independent on whether the contact is elastic or plastic, is still given by (6.13) for a given separation between the surfaces (Bhushan and Nosonovsky 2004a), whereas the real contact area is found by dividing the load by the hardness

$$A_{\mathrm{r}} \propto \frac{W}{H}. \tag{6.14}$$

Note that for elastic contact, the linear dependence of A_{r} on W (6.12) is the consequence of the fact the load W is moderate and therefore only highest asperities participate in the contact. For very high loads (or elastic materials such as rubber), the real area of contact can become comparable with the nominal area of contact and thus (6.12) and (6.13) will not stand (Nosonovsky and Bhushan 2008a). The linear dependence of the real area of contact (for both elastic and plastic contact) on the normal load is the factor responsible for the linear proportionality of the friction force to the normal load, as it will be explained below.

6.3 Dry Friction and Its Laws

Empirical observations of dry friction between solid surfaces have resulted in the laws of friction. These laws are sometimes attributed to Leonardo da Vinci. However, they became known in the scientific community after they were formulated by Guillaume Amontons and Charles-Augustin Coulomb. The three Amontons (or Amontons–Coulomb) laws state the following:

1. The friction force, F, is directly proportional to the applied normal load, W.

$$F = \mu W \tag{6.15}$$

   where μ is a constant coefficient referred to as the coefficient of dry friction. In an alternative formulation, the coefficient of friction is independent of the normal load, W.
2. The friction force, F, is independent of the nominal (or apparent) area of contact, or, μ is independent of A.
3. The friction force, F, is independent of the sliding velocity V, or μ is independent of V.

Thus, the there laws state that the coefficient of friction is independent of W, A, and V. Sometimes the first two laws are attributed to Amontons, whereas the third one is attributed to Coulomb. In the original formulation, the authors of the laws admitted that the friction force can slightly depend on the sliding velocity; in particular, it tends to increase for small but increasing velocities, remain constant for higher velocities, and decrease for even higher velocities. Summarizing the

three laws, the coefficient of friction is independent of the normal load, nominal area of contact, and sliding velocity.

The three Amontons–Coulomb laws of friction should not be viewed as logically independent "axioms." Quite oppositely, the first and the second laws can be treated as a logical consequence of each other. If the friction force F is independent of the nominal area of contact A, dividing the contact region into two halves (each having the area of $A/2$) will result in the normal load $W/2$ supported by each half. Since the total force is equal to sum of forces acting at each part of the contact region, one concludes that the friction force $F/2$ act at each half and thus the friction force is proportional to the normal load.

Speaking more formally, suppose that the friction force depends on the normal load and nominal area of contact as $F = f(W, A)$, so that changing the scale of W and A by a and b times results in the change on friction with the power exponents of n and m correspondingly

$$f(\alpha W, \beta A) = \alpha^n \beta^m f(W, A). \tag{6.16}$$

Since force is an additive function, we have also

$$f(\alpha W, \alpha A) = \alpha f(W, A). \tag{6.17}$$

Setting $\alpha = \beta$ and combining (6.16) and (6.17) yields $n + m = 1$. Therefore, the validity of the first law ($n = 1$) implies also the validity of the second law ($m = 0$) and vice versa (Nosonovsky and Bhushan 2008a).

The third empirical law of friction, which states that the friction force does not depend on the real area of contact, is logically independent of the first two laws. Despite that, it is possible to show, however, that velocity-dependence of dry friction is also related to the dependence on the size of contact. Consider the contact of a plastic rough surface with a rigid flat. An important characteristic of such contact is the average size of individual asperity contacts, a, which depends upon the separation distance between the two bodies. The simplified models of contact predict that a is independent of the normal load and the size of contact. For example, assuming that the contact shape is circular, the average area of individual asperity contact is πa^2. On the other hand, the individual contact area is given by the ratio of the real area of contact to the number of contact $\pi a^2 = A_r/N$. From (6.13) and (6.14), a is independent of W. The reasoning behind that is that when the load is increased, and the separation between the contacting surfaces increases, the size of individual asperity contacts grows; however, more new small contacts are created, so the average contact size does not change. However, in reality, the average size of contact depends on the separation between two bodies, for large separation the contact size is small (Fig. 6.2). The separation can change due to the change of the load, nominal area of contact, and sliding velocity (longer existing contacts tend to provide lower separation due to the creep and visco-plastic deformation), which makes the load-, contact size-, and velocity-dependence of friction.

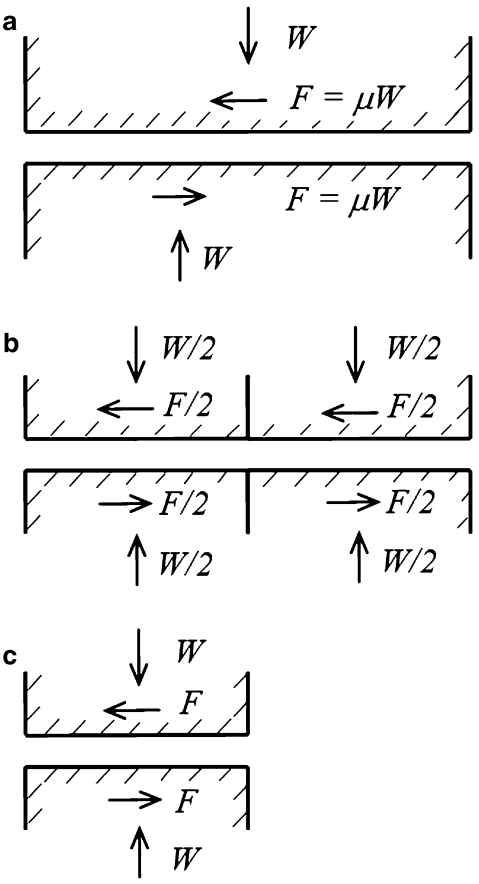

Fig. 6.2 Illustration of coupling of the laws, which state that the coefficient of friction is independent of the normal load and apparent area of contact. (**a**) The nominal area of contact A_a between two bodies supports the normal load of W and results in friction force $F = \mu W$. (**b**) Parts of the nominal area of contact, A_a/c (shown for $c = 2$) support the normal load of W/c and results in the friction force $F/c = \mu W/c$. (**c**) For friction force linearly dependent on the normal load, an increase of the load at the small contact area up to W, results in the friction force equal to F and thus independent of the area of contact (Nosonovsky 2007c)

It is usually stressed in the tribological literature that the three Amontins–Coulomb empirical laws of friction are only approximations and there are many situations when these laws are not valid. For this reason, some tribologists prefer to speak about the "rules of friction" rather than laws, reserving the word "law" for fundamental laws of nature, such as the Newton's laws of mechanics. On the other hand, the Amontins–Coulomb laws are valid for an amazingly broad range of material combinations, friction mechanisms, and loads ranging from nanonewtons to thousands of tons.

6.4 Theories Explaining Dry Friction

Despite the simplicity of its empirical laws, friction is a very complex phenomenon and it can involve various mechanisms of different physical nature, apparently unrelated, acting independently or simultaneously.

6.4.1 Adhesive Friction

Adhesion is the most common and best studied mechanism of dry friction, which occurs at a wide range of length scales and conditions. When two surfaces are brought into contact, adhesion or bonding across the interface can occur, and a normal force, called the adhesion force, is required to pull apart the two solids (Bowden and Tabor 1950). The word "adhesion" is a general term for such a force of any physical nature, which may include chemical covalent forces, electrical van der Waals interactions, electrostatic forces, and capillary forces. Since the typical range of the adhesion force (with the exception of the capillary force) is in nanometers, the role of the adhesion is important at the nanoscale. For chemically nonactive surfaces, there are two types of interatomic adhesive forces: the strong (chemical) forces, such as covalent, ionic, metallic bonds, whose rupture corresponds to large absorption of energy (around 400 kJ mol^{-1}), and weak forces, such as hydrogen bonds, van der Waals forces (few kJ mol^{-1}) (Maugis 1999). Weak conservative forces act at larger ranges of distance, whereas strong bonds act at short distances.

For macrofriction of nonadhesive surfaces, Bowden and Tabor (1950) suggested that the friction force F is directly proportional to the real area of contact A_r and shear strength at the interface τ_f

$$F = \tau_f A_r. \tag{6.18}$$

Every nominally flat surface in reality has roughness. The real area of contact is only a small fraction of the nominal area of contact because the contacts take place only at the summits of the asperities (Fig. 6.3a). Various statistical models of contact of rough surfaces show that A_r is almost directly proportional to the applied normal load W, for elastic and plastic surfaces, which explains the empirically observed linear proportionality of F and W (the so-called Coulomb–Amontons' rule), assuming constant τ_f (Greenwood and Williamson 1966). The physical nature of the surface shear strength τ_f, however, remains a subject of discussion. For the pure interfacial friction, τ_f may be viewed as the shear component of the adhesive force, which is required to move surfaces relative to each other. Note that the van der Waals adhesion force is conservative and by itself it does not provide a mechanism needed for energy dissipation. It was suggested recently (Maeda et al. 2002; Szoszkiewicz et al. 2005; Zeng et al. 2006; Ruths and Israelachvili 2007) that nanofriction is not related to the adhesion per se, but to the adhesion *hysteresis*. The energy needed to separate two surfaces is always greater than the energy gained by bringing them together (Fig. 6.4). As a result, the energy is dissipated during the separation process. The adhesion hysteresis or surface energy hysteresis can arise even between perfectly smooth and chemically homogeneous surfaces supported by perfectly elastic materials. The adhesion hysteresis exists due to surface roughness and inhomogeneity (Maeda et al. 2002). The van der Waals force itself is conservative and does not provide a mechanism of energy dissipation. However, the

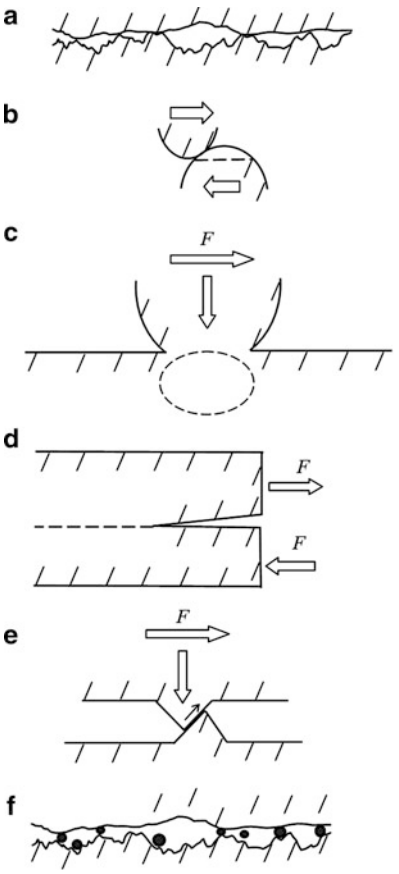

Fig. 6.3 Fundamental mechanisms of friction (**a**) adhesion between rough surfaces, (**b**) plowing, (**c**) the plastic yield, (**d**) the similarity of a mode II crack propagation and friction, (**e**) the ratchet mechanism, (**f**) the third-body mechanism (Nosonovsky and Bhushan 2008a)

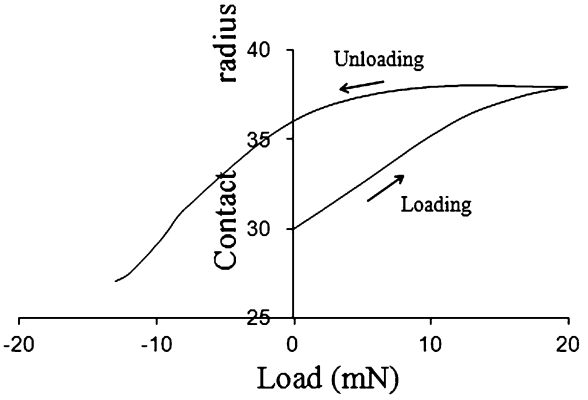

Fig. 6.4 Adhesion hysteresis. Adhesion force is different when surfaces are approaching contact and when separating for polystyrene (based on Maeda et al. 2002)

adhesion hysteresis due to surface heterogeneity and chemical reactions leads to dissipation (Maeda et al. 2002; Szoszkiewicz et al. 2005; Zeng et al. 2006; Ruths and Israelachvili 2007). Both sliding and rolling friction involve the creation and consequent destruction of the solid–solid interface. During such a loading–unloading cycle, the amount of energy ΔW is dissipated per unit area.

Since the underlying physical reason of the adhesion hysteresis is in surface roughness and chemical heterogeneity, there is a natural way to obtain the hysteresis of a conservative van der Waals force by assuming that the surface is not perfectly rigid, that is, deformable. There are a number of contact models which combine the elastic deformation and adhesion (Johnson 1998), however, these theories do not address the issue of adhesion hysteresis.

Consider a rigid cylinder of radius R and length L rolling along a solid surface with the van der Waals attractive adhesion force between them. From the energy balance, when the cylinder passes the distance d, the amount of dissipated energy $\Delta W A_\mathrm{r}$ is equal to the work of the friction force F at the distance d, and therefore, the friction force is given by (Nosonovsky 2007b)

$$F = \frac{A_\mathrm{r} \Delta W}{d}. \tag{6.19}$$

For a multiasperity contact, the real area of contact, A_r, is only a small fraction of the nominal contact area, which is equal to the surface area covered by the cylinder, Ld. During frictional sliding of a solid cylinder against a flat surface, the solid–solid interface is created and destroyed in a similar manner to rolling. Based on the adhesion hysteresis approach, the frictional force during sliding is also given by (6.19) and all considerations presented in the preceding section are valid also for the sliding friction.

Summarizing, the adhesive friction provides the mechanism of energy dissipation due to breaking strong adhesive bonds between the contacting surfaces and due to the adhesion hysteresis. In order for adhesive friction to exist, either irreversible adhesion bonds should form, or the contacting bodies should be deformable and thus nonideally rigid. Adhesive friction mechanism involves weak short-range adhesive force and strong long-range bulk forces.

6.4.2 *Deformation of Asperities*

Another important mechanism of friction is the deformation of interlocking asperities (Fig. 6.3b). Like adhesion, which may be reversible (weak) and irreversible (strong), deformation may be elastic (i.e., reversible) and plastic (irreversible plowing of asperities). For elastic deformation, certain amount of energy is dissipated during the loading–unloading cycle due to radiation of elastic waves and viscoelasticity, so an elastic deformation hysteresis exists, similar to the adhesion hysteresis. The value of deformational friction force is usually higher

than that of adhesive friction and depends on the yield strength and hardness, which trigger a transition to the plastic deformation and plowing. The transition from adhesive to deformational fiction mechanism depends on load and yield strength of materials and usually results in a significant increase of the friction force.

We discuss in the preceding section the effect of adhesion hysteresis on friction of a cylinder sliding upon a flat surface. We found that the contribution of adhesion hysteresis into the sliding friction is equal to that of rolling friction. However, it is well known from the experiments that sliding friction is usually greater than the rolling friction. This is because plowing of asperities takes place during sliding. Even smooth surfaces have nanoasperities, and their interlocking can result in plowing and plastic deformation of the material. Usually, asperities of softer material are deformed by asperities of a harder material. The shear strength during plowing is often assumed to be proportional to the average absolute value of the surface slope (Bhushan 1999, 2002). It is therefore assumed that in addition to the adhesion hysteresis term, there is another component, H_p, which is responsible for friction due to surface roughness and plowing (Nosonovsky 2007b).

$$F = A_r\left(\frac{\Delta W}{d + H_p}\right). \tag{6.20}$$

The plowing term may be assumed to be proportional to the average absolute value of the surface slope. Note that the normal load is not included into (6.20) directly; however, A_r depends upon the normal load. The right-hand side of (6.20) involves two terms: a term that is proportional to adhesion hysteresis and a term that is proportional to roughness. Nosonovsky (2007b) pointed out that Eq. 6.20, which governs energy dissipation during the solid–solid friction, is similar to the equations that govern energy dissipation during solid–liquid friction.

Due to the surface roughness, deformation occurs only at small parts of the nominal contact area, and the friction force is proportional to the real area of contact involving plowing, as given by (3.2). Due to the small size of the real area of contact compared with the nominal area of contact, the plastically deformed regions constitute only a small part of the bulk volume of the contacting bodies.

6.4.3 Plastic Yield

Chang et al. (1987) proposed a model of friction based upon plastic yield, which was later modified by Kogut and Etzion (2004). They considered a single-asperity contact of a rigid asperity with an elastic–plastic material. With an increasing normal load, the maximum shear strength grows and the onset of yielding is possible. The maximum shear strength occurs at a certain depth in the bulk of the body (Fig. 6.3c). When the load is further increased and the tangential load is applied,

the plastic zone grows and reaches the interface. This corresponds to the onset of sliding. Kogut and Etzion (2004) calculated the tangential load at the onset of sliding as a function of the normal load using the finite elements analysis and found a nonlinear dependence between the shear and tangential forces. This mechanism involves plasticity and implies structural vulnerability of the interface compared to the bulk of the contacting bodies.

6.4.4 *Fracture*

For a brittle material, asperities can break forming wear debris. Therefore, fracture also can contribute into friction. There is also an analogy between mode II crack propagation and sliding of an asperity (Rice 1991; Gerde and Marder 2001; Kessler 2001, Fig. 6.3d). When an asperity slides, the bonds are breaking at the rear, while new bonds are being created at the front end. Thus, the rear edge of asperity can be viewed as a tip of a propagating mode II crack, while the front edge can be viewed as a closing crack. Gliding dislocations, emitted from the crack tip, can also lead to the microslip or local relative motion of the two bodies (Bhushan and Nosonovsky 2003; Nosonovsky and Bhushan 2005a, b). Calculations have been conducted to relate the stress intensity factors with friction parameters (Rice 1991; Gerde and Marder 2001; Kessler 2001). Crack and dislocation propagation along the interface implies that the interface is weak compared to the bulk of the body.

6.4.5 *Ratchet and Cobblestone Mechanisms*

Interlocking of asperities may result in one asperity climbing upon the other, leading to the so-called ratchet mechanism (Bhushan 1999, 2002). In this case, in order to maintain sliding, a horizontal force should be applied, which is proportional to the slope of the asperity (Fig. 6.3e). At the atomic scale, a similar situation exists when an asperity slides upon a molecularly smooth surface and passes through the tops of molecules and valleys between them. This sliding mechanism is called "cobblestone mechanism" (Israelachvili 1992). This mechanism implies that the strong bonds are acting in the bulk of the body, whereas interface bonds are weak.

6.4.6 *"Third-Body" Mechanism*

During the contact of two solid bodies, wear and contamination particles can be trapped at the interface between the bodies (Fig. 6.3f). Along with liquid which condensates at the interface, they form the so-called "third body" which plays a significant role in friction. The trapped particles can significantly increase the

coefficient of friction due to plowing. Some particles can also roll and thus serve as rolling bearings, leading to reduced coefficient of friction. However, in most engineering situations, only 10% of the particles roll (Bhushan 2002) and thus the third-body mechanism leads to an increase of the coefficient of friction. At the atomic scale, adsorbed mobile molecules can constitute the "third body" and lead to significant friction increase (He et al. 2003). The third body has much weaker bonds to the surface, than those in the bulk of the body.

In summary, there are several mechanisms of dry friction. They all are associated with a certain type of heterogeneity or nonideality, including surface roughness, chemical heterogeneity, contamination, and irreversible forces. All these mechanisms are also characterized by the interface forces being small compared to the bulk force. In the consequent chapters, we discuss linearity of friction as a result of presence of a small parameter, nonlinearity of friction, related to heterogeneity and hierarchical structure and multiscale nature of the frictional mechanisms.

6.4.7 Origins of the Linearity of Friction

Empirical observations regarding dry friction are summarized in the so-called Coulomb–Amontons' rule, which states that the friction force F is linearly proportional to the normal load W

$$F = \mu W, \tag{6.21}$$

where μ is a constant for any pair of contacting materials, called the coefficient of friction. The coefficient of friction is almost independent of the normal load, nominal size of contact, and sliding velocity. Although there is no underlying physical principle, which would require the friction force to be linearly proportional to the normal load, (6.21) is valid for a remarkably large range of conditions and regimes of friction, from macro- to nanoscale, for loads ranging from meganewtons to nanonewtons and for various material combinations. Two main physical explanations of the linearity of friction have been suggested, based on the friction force proportionality to the real area of contact between the two bodies and to the average slope of a rough surface.

As it has been explained above, the real area of contact usually constitutes a small fraction of the nominal area of contact. Thus, for metals at loads typical for technical applications, the real area of contact constitutes less than 1% of the nominal area of contact. Various statistical models of contacting rough surfaces have been proposed, following the first publication by Greenwood and Williamson (1966). These models conclude, using the numerical computations, that for typical roughness distributions, such as the Gaussian roughness, for both elastic and plastic materials, the real area of contact is almost linearly proportional to the load (Adams and Nosonovsky 2000).

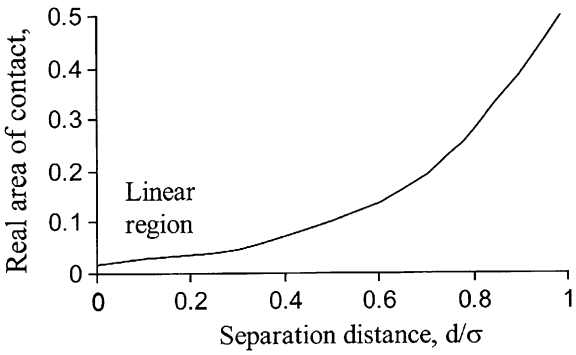

Fig. 6.5 The number of contacts and contact area as a function of separation between the contacting bodies (based on Onions and Archard 1973)

For the elastic contact of a smooth surface and a rough surface with the correlation length β^* and standard deviation of profile height σ, the real area of contact is given by (6.12). Note that σ is the vertical and β^* is the horizontal roughness parameters with the dimension of length. The smoother is the surface (higher the ratio β^*/σ), the larger is A_r. Physically, the almost linear dependence of the real area of contact upon the normal load in this case is a result of the small extent of contact, in other words, it is the consequence of the fact that the real area of contact is a small fraction of the nominal area of contact. With increasing load, as the fraction of the real area of contact grows, or for very elastic materials, such as the rubber, the dependence is significantly nonlinear. However, for small real area of contact, with increasing load the area of contact for every individual asperity grows, but the number of asperity contacts also grows, so the average contact area per asperity remains almost constant (Fig. 6.5).

For plastic contact, the real area of contact is independent of roughness parameters and given by the ratio of the normal load to the hardness of a softer material H_s (6.14). Hardness is usually defined in indentation experiments as force divided by the indentation area, so (3.11) naturally follows from this definition. In many cases, it may be assumed that the hardness is proportional to the yield strength. Whether the contact is elastic or plastic may depends upon the roughness parameters, elastic modulus, and hardness. Interestingly, Greenwood and Williamson (1966) showed whether the contact is elastic or plastic does not depend upon the load, but solely upon the so-called plasticity index $\psi = \left(\sqrt{\sigma/R_p}\right)E^*/H$, where σ is the standard deviation of peak heights and R_p is mean asperity peak radius. Combining the Bowden and Tabor's model (6.18) with the conclusions of the statistical models of contact of rough surface ((6.12)–(6.14)), the friction force due to adhesion is proportional to the real area of contact and adhesive shear strength τ_a which yields a linear dependence of F upon W.

A different explanation of the linear dependence of A_r on W is given by so-called fractal models of rough surfaces. These models assume that a rough surface or profile possesses the mathematical properties of self-similar, or fractal, surfaces and curves. When the resolution of a measuring equipment is magnified, more and more further details are observed, which are similar to the details observed at higher scale of resolution. Such a curve or surface, which reproduces all its properties at

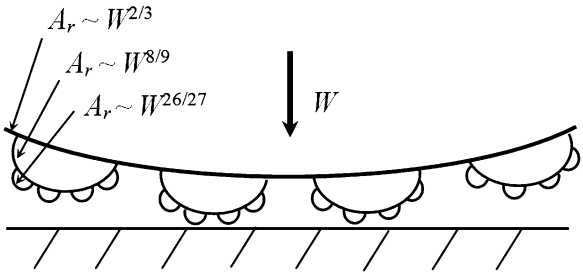

Fig. 6.6 A multiscale rough elastic surface in contact with a flat surface (Nosonovsky and Bhushan 2008a)

different length scales, is called self-similar. If some properties are reproduced, it is called a self-affine curve or surface. Self-similar and self-affine objects were introduced into the mathematics yet in 1930s; however, their application to various physical and engineering problems became popular in the consequent decades after they were popularized in the 1970s by B. Mandelbrot who actually coined the term "fractals."

Long before the term "fractal" was invented by mathematicians, Archard (1957) studied multiscale roughness with small asperities on top of bigger asperities, with even smaller asperities on top of those, and so on (Fig. 6.6). According to the Hertzian model, for the contact of an elastic sphere of radius R with an elastic flat with the contact radius a, and the contact area $A = \pi a^2$ are related the normal load as

$$A_{\mathrm{r}} = \pi\left(\frac{3RW}{4E^*}\right)^{2/3}. \tag{6.22}$$

The pressure distribution as the function of the distance from the center of the contact spot, r, is given by

$$p = \left(\frac{6WE^{*2}}{\pi^3 R^2}\right)^{1/3}\left(\sqrt{1-\left(\frac{r}{a}\right)^2}\right). \tag{6.23}$$

Let us now assume that the big spherical asperity is covered uniformly by many asperities with a much smaller radius, and these asperities form the contact. For an asperity located at the distance r from the center, the load is proportional to the stress given by (6.24). The area of contact of this small asperity is still given by (6.22) with using the corresponding load. The dependence of total contact area upon W is then given by integration of the individual contact areas by r as (Archard 1957)

$$\begin{aligned} A_{\mathrm{r}} &\propto \int_0^a \left[W^{(1/3)}\left(\sqrt{1-\frac{r^2}{a^2}}\right)\right]^{2/3} 2\pi r\,\mathrm{d}r, \\ &\propto \int_0^{\pi} \left[W^{(1/3)}\cos\phi\right] \propto^{2/3} 2\pi(a\sin\phi)a\cos\phi\,\mathrm{d}\phi, \\ &\propto W^{(2/9)}a^2 \propto W^{(2/9)}W^{(2/3)} \propto W^{(8/9)}. \end{aligned} \tag{6.24}$$

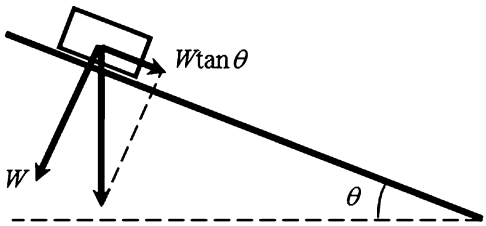

Fig. 6.7 Slope-controlled friction. For a body moving without acceleration upon an inclined surface with slope θ, the shear force, $W \tan \theta$, is proportional to the normal load, W (Nosonovsky and Bhushan 2008a)

In the above derivation, the variable change $r = a \sin \phi$ and (3.6) were used. The integral of the trigonometric functions can be easily calculated, however, its value is not important for us, because it is independent of a and W.

If the small asperities are covered by the "third-order" asperities of even smaller radius, the total area of contact can be calculated in a similar way as

$$A_r \propto \int_0^a \left[W^{(1/3)} \left(\sqrt{1 - \frac{r^2}{a^2}} \right) \right]^{2/3} 2\pi r dr \propto W^{(8/9)} a^2 \propto W^{(26/27)}. \tag{6.25}$$

Continuing this iterative procedure of building small asperities on top of larger asperities, we find that

$$A_r \propto \lim_{n \to \infty} \left(W^{\frac{3^n - 1}{3^n}} \right) = W, \tag{6.26}$$

where n is the number of orders of asperities, leading to an almost linear dependence of A_r upon W with increasing n. Later more sophisticated fractal surface models were introduced, which lead to similar results (Majumdar and Bhushan 1991).

Thus, both statistical and fractal roughness, for elastic and plastic contact, combined with the adhesive friction law (6.18) results in an almost linear dependence of the friction force upon the normal load.

A completely different explanation of the linearity of friction is based on the assumption that during sliding asperities climb upon each other (the ratchet mechanism) (Fig. 6.7). From the balance of forces, the horizontal force, which is required to initiate motion, is given by the normal load multiplied by the slope of the asperities.

$$F = W \tan \theta, \tag{6.27}$$

where θ is the slope angle of the asperities. Comparing (6.21) and (6.26), it may be concluded that for a rough surface, the coefficient of fiction is equal to the average absolute value of its slope, $\mu = |\tan \theta|$. The sign of the absolute value is required because asperities can climb only if the slope is positive. Similar to the ratchet mechanism is the cobblestone mechanism, which is typical for the atomic friction.

Among other attempts to explain the linearity of the friction force with respect to the load, two modeling approaches are worth to mention. Sokoloff (2006) suggested that the origin of the friction force is in the hardcore atomic repulsion. The vertical

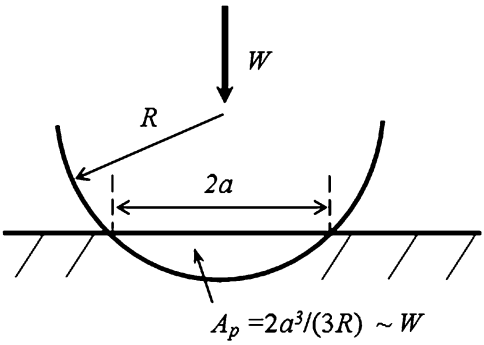

Fig. 6.8 "Elastic plowing": the transsectional area of the asperity is linearly proportional to the Hertzian normal load (Nosonovsky and Bhushan 2008a)

component of the vector of the repulsion force, which contributes to the normal load, is proportional to the horizontal component of the same vector, which contributes into friction, because the vector has a certain average orientation. In a sense, this is still the same slope-controlled mechanism, however, considered at the atomic level.

Ying and Hsu (2005) suggested another interesting macroscale explanation of the linearity of friction. They noticed that using the Herzt theory (6.22) for a spherical asperity of radius R, slightly indented into a substrate, the contact radius, a, is proportional to the power 1/3 of the penetration h (Fig. 6.8).

$$a \propto W^{1/3}. \tag{6.28}$$

When such an asperity plows the substrate, the cross-sectional plowing area (or projection of the indented part of the sphere upon a vertical plane) A_p is given by a cubic function of a and thus is proportional to the normal load

$$A_\mathrm{p} = \frac{2a^3}{3R} \propto W. \tag{6.29}$$

Thus, in the case of "elastic plowing," the plowing force, which is proportional to A_p, is linearly proportional to the normal load.

We have found that the mechanisms of friction are quite divers and apparently unrelated to each other (Table 6.2). Several physical mechanisms result in a linear dependence of the friction force upon the normal load. Mathematically, a linear dependence between the two parameters usually exists, when the domain of a changing parameter is small, and thus a more complicated dependency can be approximated within this domain as a linear function. For example, if the dependency of the friction force upon the normal load is given by

$$F = f(W) \approx f(0) + f'(0)W + \frac{f''(0)}{2}W^2 = \mu W + \frac{f''(0)}{2}W^2, \tag{6.30}$$

Table 6.2 Mechanisms of friction and linear dependence of the friction force upon the normal load (Nosonovsky and Bhushan 2008a)

	Mechanism	Friction force and real area of contact as functions of the normal load
Area-controlled	Elastic hierarchical (Archard)	$F = \tau_a A_r \propto W^{(3^n-1)/3^n}$
	Elastic statistical	$F = \tau_a A_r \propto \frac{\beta^*}{E^* \sigma} W$
	Plastic	$F = \tau_a A_r = \frac{W}{H_s}$
Slope-controlled	Ratchet	$F = W \tan \theta$
Other	Elastic plowing	$F = \tau_a A_p = \frac{2a^3}{3R} \propto W$

The dependency can be linearized as $F = \mu W$ if $W \ll 2\mu/f''(0)$. In other words, the ratio of the load W to a corresponding parameter of the system, given by (3.22) (with the dimension of force), is small. That parameter may correspond to the bulk strength of the body (Nosonovsky and Bhushan 2008a).

6.5 Wear

Wear is defined as material removal and deterioration during the contact of solid surfaces in relative motion. There are several mechanisms which lead to wear. First is the adhesive wear due to the adhesion between the contacting surfaces which can lead to the removal and transfer of particles of a material and displacement of wear debris from one surface to the other. Adhesive wear occurs when two bodies slide over or pressed into each other, which promote material transfer. This involves the plastic deformation of very small fragments within the surface layers.

Second is the abrasive wear due to plowing, cutting, and fragmentation of asperities. Abrasive wear is the loss of material due to hard particles or hard protuberances that are forced against and move along a solid surface. It occurs when a hard rough surface slides across a softer surface. Plowing is the displacement of the material to the side, away from the wear particles, resulting in the formation of grooves that do not involve direct material removal. The displaced material forms ridges adjacent to grooves, which may be removed by subsequent passage of abrasive particles. Cutting is material separated from the surface in the form of primary debris, or microchips, with little or no material displaced to the sides of the grooves. This mechanism closely resembles conventional machining. Fragmentation is material separation from a surface by a cutting process and the indenting abrasive causes localized fracture of the wear material.

Other types of wear are surface fatigue, fretting wear, erosive wear, and cavitation wear. Surface fatigue is weakening of the surface due to cyclic loading. Fretting or fretting fatigue is a repeated cyclical rubbing of the surface. Erosive wear is caused by the impact of solid particles or fluid on the surface. Cavitation wear is due to the contact with fluid.

Fracture can be another mode of wear. During the contact of two asperities with friction and shear loading, the maximum shear strength is usually achieved beneath the surface, rather than at the surface. As a result, plastic yield and flow starts at the subsurface zone and, due to repeated loading, a thin "skin-like" layer of material delaminates.

The quantitative characteristic of wear is the wear volume or volume of worn material, w. The rate of wear is measured in wear volume per unit time. The so-called empirical Archard wear law (or rule) relates the wear rate with the sliding velocity, V, the applied normal load, W, and the hardness, H, of a softer material among the two contacting materials

$$\dot{w} = k\frac{WV}{H}. \tag{6.31}$$

This empirical law of wear was formulated by Archard for the abrasive wear and it states that the wear rate is linearly proportional to the sliding velocity (or the wear volume is linearly proportional to the sliding distance), the ratio of the normal load to the hardness of the softer material, and the coefficient k referred to as "the wear coefficient," which is a characteristic of tribological system somewhat similar to the coefficient of friction.

6.6 Lubrication

Lubrication is interposing a substance called lubricant between the surfaces to carry or to help carry the load between the sliding surfaces. It is employed to reduce friction and wear. The lubricant can be a solid (e.g., graphite, MoS_2), a solid–liquid dispersion, a liquid, a liquid–liquid dispersion, or a gas. In the most common case, the lubricant is fluid capable of bearing the pressure between the surfaces. Adequate lubrication allows smooth continuous operation of equipment, with only mild wear, and without excessive stresses or seizures at bearings.

As the load increases on the contacting surfaces, three lubrication regimes can be observed:

1. Fluid film lubrication regime in which the load is fully supported by the lubricant within the space or gap between the parts in motion relative to one another (the lubricated conjunction), and solid–solid contact is avoided. Two types of the fluid film lubrication are the hydrostatic lubrication, when an external pressure is applied to the lubricant in the bearing, and the hydrodynamic lubrication, when the motion of the contacting surfaces is used to pump lubricant to maintain the lubricating film.
2. Elastohydrodynamic lubrication regime in which the contacting surfaces are separated in general, however the interaction between the asperities can occur and an elastic deformation on the contacting surface enlarges the load-bearing

area whereby the viscous resistance of the lubricant becomes capable of supporting the load.
3. Boundary lubrication is the regime in which the bodies come into closer contact at their asperities and the load is carried by the surface asperities rather than by the lubricant.

Lubricant may also cool the contact areas and remove wear products. While carrying out this function, the lubricant is constantly replaced from the contact areas either by the relative movement (hydrodynamics) or by externally induced forces.

6.7 Self-Lubrication

The term "self-lubrication" has been used for more than two decades and it refers to several methods and effects that reduce friction or wear. Among these methods are the deposition of self-lubricating coatings that are either hard (to reduce wear) or with low surface energy (to reduce adhesion and friction). Besides coatings, self-lubrication can imply the development of metal-, polymer-, or ceramic-based composite self-lubricating materials, often with a matrix that provides structural integrity and reinforcement material that provides low friction and wear. The nanocomposites have become a focus of this research, as well as numerous attempts to include nanosized reinforcement, carbon nanotubes (CNT), and fullerene C_{60} molecules. Simple models assume that these large molecules and nanosized particles serve as "rolling bearings" that reduce friction; however, it is obvious now that the mechanism can be more complicated and involve self-organization. Dynamic self-organization is thought to be responsible for self-lubrication in the atomic force microscopy experiments with the atomic resolution. A different approach involves a layer of lubricant that is being formed in situ during friction due to a chemical reaction. Such a reaction can be induced in situ by mechanical contact, for example, a copper protective layer formed at a metallic frictional interface due to the selective transfer of Cu ions from a copper-containing alloy (e.g., bronze) or from a lubricant. A protective layer can be formed also due to a chemical reaction of oxidation or a reaction with water vapor. For example, a self-lubricating layer of the boric acid (H_3BO_3) is formed as a result of a reaction of water molecules with B_2O_3 coating. Another type of self-lubricating material involves lubricant embedded into the matrix, e.g., inside microcapsules that rupture during wear and release the lubricant. Surface microtexturing that provides holes and dimples which serve as reservoirs for lubricant can be viewed as another method of providing self-lubrication. In addition, we should mention that self-lubrication is observed also in many biological systems and that the term "self-lubrication" is used also in geophysics where it refers to animally low friction between tectonic plates that is observed during some earthquakes.

The design of coatings with hard and lubricious diamond-like carbon (DLC) surfaces requires a study of transitions between adhesive metal, load supporting

carbide, and wear-resistant DLC materials. Voevodin et al. (1997) investigated these transitions were investigated on the Ti–C system prepared by a hybrid of magnetron sputtering and pulsed laser deposition (PLD). Crystalline alpha-Ti, TiC, and amorphous DLC films were formed at 100°C substrate temperature by varying film chemical composition. A gradual replacement of alpha-Ti with TiC, and a two-phase region consisting of crystalline TIC and amorphous carbon (a-C) in transitions from Ti to TiC and from TiC to DLC were found. These transitions were reflected in mechanical properties investigated with nanoindentation. This provided a hard coating with a low friction surface, which also resisted brittle failure in tests with high contact loads.

Neerinck et al. (1998) used PLD to produce superhard (60–70 GPa) self-lubricating DLC with low friction and low wear rate. They obtained thin (2–3 μm) DLC-based coatings for steel substrates, which could maintain friction coefficients of about 0.1 for several million cycles of unlubricated sliding at contact pressures above 1 GPa. Their scratch resistance exeeded that of conventional ceramic (TiN, TiC) coatings.

Vilar (1999) used laser cladding for the protection of materials against wear, corrosion, and oxidation, for the deposition of self-lubricating coatings and thermal barriers, and for the refurbishing of high-cost industrial components. Laser cladding is a hard-facing process that uses a high-powered laser beam to melt the coating material and a thin layer of the substrate to form a pore- and crack-free coating 50 μm to 2 mm thick with low dilution that is perfectly bonded to the substrate. The process may be used for large area coverage by overlapping individual tracks, but it is the ability to protect smaller, localized areas that makes it unique.

Erdemir et al. (1990, 1991, 1996a, b) investigated boric-acid (H_3BO_3)-based applications for the formation and self-lubrication mechanisms of boric acid films on boric oxide coatings prepared by vacuum evaporation. In particular, they measured the coefficients of friction of a steel ball sliding on a boric-oxide-coated steel disk and a sapphire ball sliding on a boric-oxide-coated alumina disk were 0.025–0.05 at steady state, depending on load and substrate material. This low friction was correlated with the formation of a lubricious boric acid film on boric oxide coatings exposed to open air. For the mechanism of self-lubrication, the layered triclinic crystal structure of boric acid was proposed. The atoms constituting each boric acid molecule are arrayed in closely packed and strongly bonded layers that are 0.318 nm apart and held together by weak forces, such as van der Waals. The authors hypothesized that, during sliding, these layers can align themselves parallel to the direction of relative motion and, once so aligned, can slide over one another with relative ease to provide low friction.

Boric oxide tends to react with water vapor present in air to form a boric acid protective coating

$$B_2O_3 + 3H_2O \rightarrow 2H_3BO_3. \tag{6.32}$$

The protective coating, in turn, leads to reduced friction and wear. With its layered crystal structure, boric acid resembles those other solids known for their

good lubrication capabilities (e.g., MoS_2, graphite, and hexagonal boron nitride) (Erdemir et al. 1996a, b).

Many composite and nanocomposite materials were suggested for self-lubricating coatings. This includes TiC/a-C:H nanocomposite coatings (Pei et al. 2005), Ti–B–N, Ti–B–N–C, and TiN/h-BN/TiB2 multilayer coatings (Mollart et al. 1996), various MMC materials (Kerr et al. 2000), titanium nitride (TiN) coating (Akhadejdamrong et al. 2003), aluminum/SiC/graphite hybrid composites with various amount of graphite addition synthesized by the semisolid powder densification (SSPD) (Guo and Tsao 2000), plasma-sprayed cast iron splats on an aluminum alloy substrate (Morks et al. 2003), TZP-graphite self-lubricating ceramics (Liu and Xue 1996), CuO-doped yttria-stabilized tetragonal zirconia ceramics (Tocha et al. 2008), carbon–carbon composites (Chen and Ju 1995), nitride compounds (Zheng and Sun 2006), CrN–Ag self-lubricating hard coatings (Mulligan and Gall 2005), super hard self-lubricating Ti–Si–C–N nanocomposite coatings (Ma et al. 2007), microplasma oxidation on aluminum alloys in the solution of aluminate–graphite (Wu et al. 2008).

Zhang et al. (2008) fabricated porous aluminum anodic oxide films by anodizing in phosphoric acid electrolyte containing organic acid. By controlling its microstructure, a macroporous and thick alumina template were obtained. Surface self-lubricating composites were prepared by taking ultra-sonic impregnation in polytetrafluoroethylene (PTFE) latex and the relative subsequent heat treatment technology. Polcar et al. (2009) argued that transition metal dichalcogenides (TMD) have been one of the best alternatives as low friction coatings for tribological applications, particularly in dry and vacuum environments, however, they have low load-bearing capacities. To increase the load-bearing capacity of these materials, the alloying with C should be considered. They studied self-lubricating W–S–C and Mo–Se–C sputtered coatings and found self-lubricating behavior. Skarvelis and Papadimitriou (2009) used plasma transferred arc (PTA) technique to produce composite coatings based on co-melting of MoS_2, TiC, and iron ingredients, in an attempt to obtain wear-resistant layers with self-lubricating properties. Graphite and glassy carbon composites were investigated by Hokao et al. (2000). Strnad et al. (2009) developed self-lubricated MoS_2 doped Ti–Al–Cr–N coatings developed in multilayer structure.

Various ceramics are also used for self-lubricating effect. Suh et al. (2008) studied self-lubricating behavior of structural ceramic balls (ZrO_2, Al_2O_3, and SiC) sliding against the ZrO_2 disk. Blau et al. (1999) investigated self-lubricating properties of ceramic-matrix graphite composites. Lugscheider et al. (1999) studied self-lubricating properties of tungsten and vanadium oxides deposited by MSIP-PVD process. Bae et al. (1996) studied self-lubricating TiN–MoS_2 composite coatings. Mulligan and Gall (2005) studied CrN–Ag self-lubricating hard coatings.

Powder metallurgy is another area of interest. Li and Xiong (2008) prepared Nickel-based self-lubricating composites with graphite and molybdenum disulfide, as lubricant were prepared by powder metallurgy (PM) method, powder metallurgy composites (Dellacorte and Sliney 1991), and MoS_2 precursor films on aluminum (Skeldon et al. 1997).

Polymers and polymer composites and nanocomposites are also used for the self-lubricating effect. Li et al (2008) prepared polyoxymethylene (POM) composites filled with low-density polyethylene (LDPE) and rice husk flour (RHF) were prepared by injection molding. Quintelier et al. (2009) used polymer composite to develop self-lubricating coatings. Blanchet and Peng (1998) prepared self-lubricating fluorinated ethylene propylene (FEP) and PTFE composites.

With the advent of new carbon-based nanomaterials, such as the fullerene and CNT, new opportunities for tribologists emerged. Thus, fullerene C_{60} (Bhushan et al 1993) and fullerene-like WS_2 nanoparticles (Rapoport et al. 2003), Ni-based CNT (Wang et al. 2003; Scharf et al. 2009), CNT on Al_2O_3 (Tu et al. 2004), CNT-reinforced Al composites (Zhou et al. 2007) and Mg composites (Umeda et al. 2009), boric acid nanotubes, nanotips, nanorods (Li et al. 2003), composite coatings of Co plus fullerene-like WS_2 nanoparticles on stainless steel substrate (Friedman et al. 2007), Ni-based alloy matrix submicron WS_2 self-lubricant composite coatings (Wang et al. 2008), and other materials were found to demonstrate self-lubricating effect.

Alexandridou et al. (1995) developed wear-resistant MMC composite coatings and oil-containing self-lubricating metallic coatings. The latter have been produced by electrolytic codeposition of oil-containing microcapsules from Watts nickel plating baths. For this purpose, oil-containing polyterephthalamide microcapsules were synthesized based on the interfacial polymerization of an oil-soluble monomer (terephthaloyl dichloride) and a mixture of two water-soluble monomers (diethylenetriamine and 1,6-hexamethylenediamine). The influence of several synthesis parameters (e.g., type of encapsulated organic phase, monomer concentration(s), and concentration ratio of the two amine monomers) on the size distribution and morphology of the oil-containing polyamide microcapsules as well as on their electrolytic codeposition behavior is discussed. The morphological characteristics of the microcapsules were affected to a great extent by the functionality of the water-soluble amine monomer. The composition of the core material of the microcapsules showed a marked influence on their stability upon aging in the Watts nickel plating bath. The level of codeposition was influenced by the presence of additives in the nickel electrolyte and was strongly dependent on the polymerization conditions employed in the microcapsule synthesis.

Sui et al. (2009) decided to use the superhydrophobicity to combine it with self-lubrication. They synthesized carbon coating on Ti_3SiC_2 with combined superhydrophobic and self-lubricating properties by chlorination at 1,000°C followed by modification of the $CF_3(CF_2)(5)CH_2CH_2SiCl_3$ film. The porous structure as well as organic film on carbon coating endowed the surface with super-hydrophobic property. Because of chemical inertia of the carbon coating and the modifier, the super-hydrophobic surface was very stable under various environments. Carbon coating was a good solid lubricant and greatly reduced friction coefficient of Ti_3SiC_2 sliding against Si_3N_4, which was important for Ti(3)SiC(2) used as engineering material.

Materials capable for the formation of in situ tribofilms are another big class of self-lubricating materials. Al_2O_3/TiC ceramic composites with the additions of CaF_2 solid lubricants showed reduced friction and wear due to an in situ formed

self-lubricating tribofilm between the ring-block sliding couple. Deng and Cao (2007) found that two types of tribofilms are formed on the wear surface depending on the CaF_2 content. A dense tribofilm with a smooth surface associated with small friction coefficient and low wear rate was formed by the releasing and smearing of CaF_2 solid lubricants on the wear surface when with 10 vol% CaF_2 content. This dense tribofilm acted as solid lubricant film between the sliding couple, and thus significantly reduced the friction coefficient and the wear rate. Breakdown of the tribofilm on the surface associated with a large wear rate was observed on samples with 15 vol% CaF_2 content. This is due to the large degradation of mechanical properties of the composite with higher CaF_2 contents.

Aizawa et al. (2005) studied self-lubrication mechanism via the in situ formed lubricious oxide tribofilms. They note that while TiN and TiC ceramic coating films are frequently utilized as a protective coating for dies and cutting tools, these films often suffer from severe, adhesive wearing in dry forming and machining. Chlorine ion implantation assists lubricious oxide film to be in situ formed during wearing. At the presence of chlorine atoms in the inside of TiN or TiC films, in situ formation of lubricous intermediate titanium oxides with TiO and Ti_nO_{2n-1} is sustained to preserve low frictional and wearing state. The self-lubrication process works well in dry machining in order to reduce the flank wear of cutting tools even in the higher cutting speed range up to 500 m min^{-1}.

Alexeev and Jahanmir (1993) studied self-lubricating composite material as a two-phase system with the plastic deformation of self-lubricating composite materials that contain soft second-phase particles. The soft phase flows toward the sliding surface. So the properties of both the hard matrix and the soft second-phase particles, as well as the shape and size of the particles, control the processes of deformation and flow of the soft phase. The results may be used to optimize the microstructure of self-lubricating composites to obtain the best tribological performance.

Self-lubrication was found also in the atomic friction. Livshits and Shluger (1997) presented a theoretical model and conducted molecular dynamics (MD) simulation of the interaction between a crystalline sample and an AFM tip nanoasperity, combined with a semiempirical treatment of the mesoscopic van der Waals attraction between tip and surface. They demonstrated that the adsorbed cluster can adjust itself to conditions of scanning by exchanging atoms with the surface and changing its structure and argued that this dynamic "self-organization" of the surface material on the tip during scanning could be a general effect which may explain why periodic surface images are often obtained using a variety of tips and large tip loads.

Another effect closely related to atomic scale self-lubrication is the superlubricity, or the regime of motion in which friction vanishes or very nearly vanishes. Superlubricity may occur when two crystalline surfaces slide over each other in dry incommensurate contact (Dienwiebel et al. 2004). Thus, the atoms in graphite are oriented in a hexagonal manner and form an atomic asperity-and-valley landscape, which looks like an egg-crate. When the two graphite surfaces are in registry (every 60°), the friction force is high. When the two surfaces are rotated

out of registry, the friction is largely reduced. A state of ultralow friction can also be achieved when a sharp tip slides over a flat surface and the applied load is below a certain threshold.

Thermolubricity is another atomic-scale phenomenon. The thermal excitations of atoms can assist sliding by overcoming energy barriers. Jinesh et al. (2008) argued that friction at low velocities and low surface corrugations is much lower than the weak logarithmic velocity-dependence predicted by thermally activated kinetic models of atomic friction. Furthermore, friction is zero in the zero-velocity limit. The effect was also demonstrated experimentally.

Two areas other than materials science where the term "self-lubrication" is used are geophysics and biology. In geophysics, scientists suggested that self-lubricating rheological mechanisms are most capable of generating plate-like motion out of fluid flows. The basic paradigm of self-lubrication is nominally derived from the feedback between viscous heating and temperature-dependent viscosity. Bercovici (1998) proposed an idealized self-lubrication mechanism based on void (such as pore and microcrack) generation and volatile (e.g., water) ingestion. The term self-lubrication is also used in certain biomedical applications. It has been argued by Bejan and Marden (2009) that the tendency of the system to reduce lubrication is a common feature of geophysical and biological systems reflecting the tendency for self-organization.

6.8 Summary

Friction and wear are fundamental phenomena of nature reflecting the tendency of energy to dissipate and material to deteriorate as a consequence of the Second law of thermodynamics. While there are many seemingly unrelated mechanisms of friction (adhesion, deformation, ratcheting, the "third body," etc.) and wear (abrasion, adhesion, fracture, etc.), in many situations friction and wear can be described by remarkably simple linear empirical laws, such as the Coulomb's law of friction and Archard's law of wear. These laws can be seen as a linearization of some more complex bulk material constitutive dependencies due to their asymptotic expansion at the 2D frictional interface.

Lubrication is required to reduce friction and wear for most interfaces. Those materials and surfaces which do not require external lubrication are called self-lubricating. Self-lubrication can often be viewed as a self-organized process. There are several approaches to the development of self-lubricating materials. These approaches include hard coatings, e.g., with DLC, the boric acid, development of friction and wear-resistant composite and nanocomposite materials on the basis of metals, polymers, and ceramics, as well as materials capable of forming in situ protective tribofilms and other self-organized structures. In the consequent chapter, we investigate in more detail the phenomenon of friction-induces self-organization and thermodynamic conditions leading to that phenomenon.