# 3

# Mechanisms of Dry Friction, Their Scaling and Linear Properties

**Abstract** Various mechanisms of dry sliding friction of two solids is discussed, including adhesion and adhesion hysteresis, deformation, plastic yield, fracture, the ratchet, cobblestone and third-body mechanisms. It is discussed how all these diverse mechanisms lead to the linear Amontons–Coulomb's empirical law of friction. Various explanations of the linearity of friction are discussed (real area of contact and slope-controlled friction, etc.) and the concept of a "small parameter" responsible for the linearity is suggested.

Dry solid–solid friction is the resistance to sliding and rolling motion. Friction is a universal phenomenon which is observed in a great variety of sliding and rolling situations. Friction is also a complex phenomenon that cannot be reduced to a single mechanism, but rather is a result of a simultaneous action of various mechanisms at different hierarchy and scale levels [30, 32, 63]. In a remarkable way, all these various mechanisms result in a dissipative process, which can often be characterized by only one single parameter, the coefficient of friction that is equal to the ratio of the friction force to the normal load. In this chapter, we will discuss general scaling issues related to solid–solid dry friction, and after that we will consider various mechanisms of friction in order to investigate what they have in common and how they all result in what is observed at the macroscale as the simple process of dry friction.

In this and following chapters, we study friction as a multiscale (hierarchical) phenomenon, showing that the mechanisms of energy dissipation result from the interplay of forces at two or more scale levels. In the following sections, the fundamental mechanisms of solid–solid friction are considered involving heterogeneity, linear, nonlinear, and hierarchical effects. The main mechanisms of dry friction are adhesion, deformation of asperities (plowing), fracture and the so-called ratchet, and third-body mechanisms [30, 32, 63]. For each mechanism, we will identify the "small parameter" that is present because the forces at the interface are smaller than the forces in the bulk. In the following chapters we will show how this small parameter leads to linearity of the friction force as a function of load. We will show that two characteristic scale lengths may be identified for most of these mechanisms. Mapping of dry friction mechanisms using these characteristic scale lengths will be proposed in the next chapter. Then a scale-dependence of these mechanisms is stud-

ied, based on the presumption that inhomogeneity at each hierarchy level leads to energy dissipation. Based on this, the second part of the book discusses hierarchical biological surfaces, which are created by nature to decrease or increase solid–solid and solid–liquid adhesion and friction. We show that their hierarchy is a consequence of simultaneously acting physical mechanisms at different scale levels; thus, surface hierarchy is a consequence of the hierarchical nature of friction mechanisms. After discussing in this chapter the well-known manifestation of linearity of friction, the Amontons–Coulomb rule, we will study deviations from linearity in the next chapter. The inherent nonlinearity of friction serves as a basis for creating hierarchical mechanisms and structures.

## 3.1 Approaches to the Multiscale Nature of Friction

Dry solid–solid friction is a complex and universal phenomenon which is found at various scale sizes from the atomic scale up to the macroscale and at different levels of the hierarchy of a device, from the level of molecules up to surfaces, asperities, components, and systems. Each of these levels is characterized by a different structure and range of scales, and each may have different predominant friction mechanisms (Table 3.1). The atomic scale (on the order of 1 nm or less) is characterized by discrete atoms and quantum-mechanical interactions (chemical bonds), described by the surface energy states. The mesoscale or nanoscale (on the order from 1 nm to 0.1 μm) is characterized by dislocations, surface defects, roughness, and inhomogeneity. Mesoscale description is required in order to provide a link between the

**Table 3.1.** Dissipation and friction mechanisms corresponding to different hierarchy levels

| Ideal situation | Real situation | 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 |

atomic and continuum levels. At the mesoscale, the bulk of the body can be viewed as divided into blocks or domains, so that the quantities which are not defined at the atomic scale, such as the yield strength or the coefficient of friction, can be defined at the macroscale by averaging throughout a mesoscale block or domain.

In order to introduce the mesoscale into friction models, it is instructive to consider the approach of scale-dependent plasticity theories. The scale-dependent yield strength is introduced in this manner by strain-gradient plasticity theories [122]. These theories postulate that the yield strength, which controls the onset of plastic flow, $\sigma_Y$, depends not only upon the strain, but also upon spatial strain gradient, $\nabla\varepsilon$, as

$$\sigma_Y = \sigma_{Y0}\sqrt{1 + l\nabla\varepsilon}, \tag{3.1}$$

where $\sigma_{Y0}$ is the macroscale yield strength and $l$ is a new characteristic length parameter postulated by these theories, which is on the order of micrometer. For two geometrically proportional configurations of different sizes, the strains are the same, but the strain gradient is much greater at a smaller scale configuration [230]. Thus, for submicron-sized systems (those with a typical size greater than $l$), the value of the yield strength will be considerably greater than the macroscale value, $\sigma_{Y0}$. Physically, the yield strength depends on the strain gradient due to the presence of the so-called geometrically necessary dislocations, which are required for strain compatibility, and their density increases with decreasing scale. Figure 3.1(a) shows the randomly distributed statistically stored dislocations during shear and geometrically necessary dislocations during bending that are needed for stress compatibility. Geometrically necessary dislocations during indentation are shown in Fig. 3.1(b). However, in order to introduce this dependence into the theory of plasticity in a strict manner, it is necessary to connect the micron-scale plasticity to the dislocation theories in a multiscale framework, and this is achieved by considering mesoscale blocks (Fig. 3.1(c)) [122, 156]. Bhushan and Nosonovsky [42] showed that such scale dependence of the yield strength and of hardness leads to the scale dependence of the coefficient of friction.

Frictional sliding is a dissipative process, and it is thermodynamically irreversible resulting in an increase of entropy of a system. Friction is not a property of a surface, but rather a system response [30, 32] that results in an increase of the system's disorder and entropy. Friction force is not a fundamental force of nature because it is a result of the action of the electromagnetic and exchange forces between the atoms, which are in principle reversible. For an ideal system of perfectly rigid bodies with potential electric forces acting between the atoms, there would be no energy dissipation and therefore no friction. Real systems, however, are imperfect and involve elastically and plastically deformable as well as brittle bodies; rough, chemically active, and inhomogeneous surfaces; and reversible weak and irreversible strong adhesive bonds. These imperfections result in energy dissipation and frictional resistance to sliding.

It is well known from experiments that the values of the coefficient of friction, when measured at the micro/nanoscale, are different from those at the macroscale, and therefore friction is scale dependent [50, 154, 287]. Various approaches have

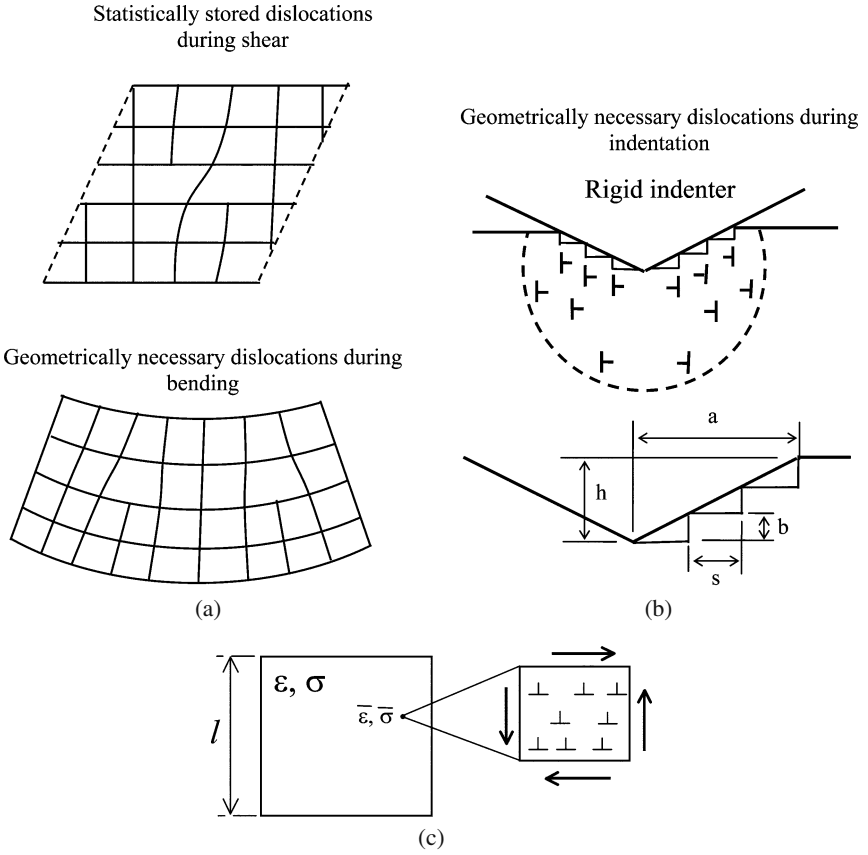



**Fig. 3.1. a** Statistically stored dislocation during bending and **b** geometrically necessary dislocations during indentation in strain-gradient plasticity [42]. **c** The multiscale framework of the strain gradient plasticity. Dislocation interaction at the microscale is considered via the Taylor relation. The higher-order strain gradient plasticity theory is established on the mesoscale representative cell of size $l$ (based on [156])

been proposed to study and explain the scale-dependence of friction. Many scholars have considered the so-called *scale effect* on friction or *scaling laws* of friction [4, 42–44, 46, 56, 87, 145, 158, 229, 326, 336, 353]. While the origin of the scaling laws is in the geometrical relations, such as surface-to-volume ratios [72, 272], the term "scale effect" implies more general laws dependent upon physical mechanisms, rather than pure geometrical relations. Johnson [168] paid attention to the fact that frictional stress is strongly dependent upon the scale of contact and suggested that gliding dislocations at the surface contribute to the frictional stress. Hurtado and Kim [158] (HK) proposed a model of single-asperity contact with a scale-dependent shear stress. Their model is based on the concept of dislocation-assisted sliding with dislocation loops nucleation at the perimeter of a circular contact zone. The model,

however, is limited to the case of commensurate interface between the bodies, which therefore should have the same orientation and spacing of the crystal lattices. This is not a likely situation in most cases. Adams et al. [4] applied the HK model for multiple-asperity elastic contact with a Gaussian statistical distribution of asperity heights and identified parameters responsible for the scale effect.

Bhushan and Nosonovsky [42–44, 46] took a different approach and considered scale-dependent distribution of surface heights combined with scale-dependent frictional stress due to dislocation nucleation from Frank–Read sources (rather than at the perimeter of the contact zone), as well as the strain-gradient plasticity. They later included in their model the effect of asperity and particle deformation, with scale-dependent densities of trapped particles at the interface. Zhang et al. [353] studied scale effects on friction using molecular dynamics (MD) simulation. He and Robbins [145] used MD simulation to study the origin of scale dependence on friction. Deshpande et al. [91] conducted numerical simulation of dislocation motion during frictional plastic deformation and showed that dislocation nucleation from the sources (rather than at the perimeter of the contact zone) results in scale-dependent frictional stress. Kogut and Etsion [189] proposed a model of elastic-plastic frictional contact with scale-independent plasticity, which resulted in the coefficient of friction strongly dependent upon the apparent area of contact, $A_a$, and normal load, $W$. Nosonovsky and Bhushan [239] also suggested that the mechanism of load-dependence of friction is similar to that of size dependence. Nosonovsky [235] also studied size, load, and velocity dependence of friction in combination. All these studies investigate some aspects of the scale effect on friction, however, they do not provide us with a general theory of scale dependence of friction.

A different approach is taken by the scholars who try to formulate empirical *friction laws at the nanoscale* rather than the scaling laws of friction [71, 302, 336]. Such friction laws are intended as substitutes for the classical Amontons–Coulomb's empirical laws (better called "rules," because situations in which these rules are not followed do not imply violation of any fundamental laws of nature) of friction, which state that the friction force between two bodies is (1) proportional to the normal load, (2) independent of the nominal contact area between the bodies, and (3) almost independent of the sliding velocity [32]. This approach, however, does not deal with the friction as a universal phenomenon and virtually considers nanoscale and macroscale friction as unrelated.

## 3.2 Mechanisms of Dry Friction

In this section we discuss major mechanisms of dry friction: adhesion, deformation of asperities, plastic yield, the ratchet, cobblestone, and third body mechanisms.

### 3.2.1 Adhesive Friction

Adhesion constitutes the most common and best studied mechanism of dry friction, which occurs at a wide range of length scales and conditions.

#### 3.2.1.1 Adhesion between Solid Surfaces

When two surfaces are brought into contact, adhesion or bonding across the interface can occur, and a finite normal force, called the adhesion force, is required to pull apart the two solids [30, 32, 63]. Since the typical range of the adhesion force is in nanometers, the role of adhesion is important at the nanoscale. As we discussed in the preceding sections, for chemically nonactive surfaces, there are two types of interatomic adhesive forces: the strong (chemical) forces, such as covalent, ionic, and metallic bonds, whose rupture corresponds to large absorption of energy (around 400 kJ/mol); and weak forces, such as hydrogen bonds and van der Waals forces (few kJ/mol) [222]. Weak conservative forces act at larger ranges of distance, whereas strong bonds act at short distances.

For macrofriction of nonadhesive surfaces, Bowden and Tabor [63] suggested that the friction force $F$ is directly proportional to the real area of contact $A_\mathrm{r}$ and shear strength at the interface $\tau_\mathrm{f}$

$$F = \tau_\mathrm{f} A_\mathrm{r}. \tag{3.2}$$

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 contact takes place only at the summits of the asperities (Fig. 3.2(a)). 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 Amontons–Coulomb's rule), assuming constant $\tau_\mathrm{f}$ [137]. The physical nature of the surface shear strength $\tau_\mathrm{f}$, however, remains a subject of discussion. For the pure interfacial friction, $\tau_\mathrm{f}$ may be viewed as the shear component of the adhesive force, which is required to move surfaces relative to each other.

Effect of adhesion on elastic contact has been investigated by many researchers [64, 82, 90, 169, 170, 221, 222]. When a smooth sphere comes into contact with elastic half-space, the contact area exceeds that predicted by the Hertzian elastic theory. The difference may be due to adhesion. Two competing models—by Johnson, Kendall and Roberts (JKR) [170] and Derjaguin, Muller, and Toporov (DMT) [90]—have been developed to account for adhesive force during elastic contact. The JKR model assumes that adhesive forces are confined to inside the contact area, whereas the DMT model also considers adhesive forces outside the contact area. Tabor [311] pointed out that these models are valid for different ranges of magnitude of elastic deformation compared to the range of surface forces, with JKR valid for large elastic deformations and DMT in the opposite case [3]. Adhesion of rough elastic surfaces has also been studied in the past years [64, 82, 264, 277, 345].

#### 3.2.1.2 Adhesion Hysteresis

It was recently suggested [211, 284, 310, 351] that nanofriction is not related to adhesion per se, but to adhesion *hysteresis*. The energy needed to separate two surfaces

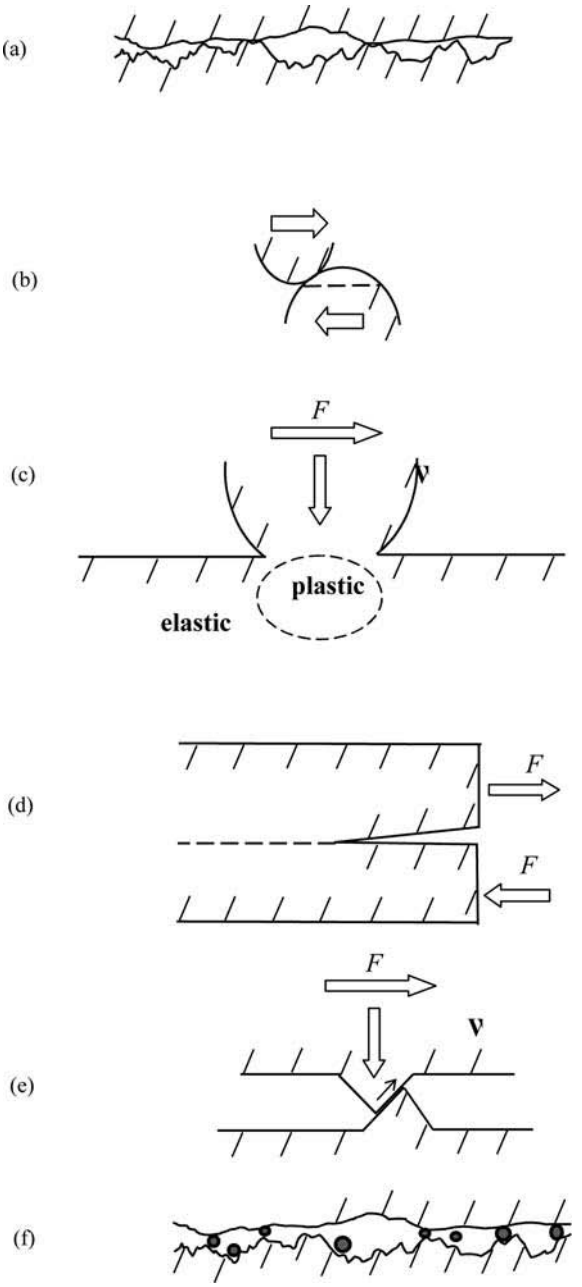



**Fig. 3.2.** Fundamental mechanisms of friction **a** adhesion between rough surfaces, **b** plowing, **c** the plastic yield, **d** the similarity of a mode II crack propagation and friction, **e** the ratchet mechanism, **f** the third-body mechanism

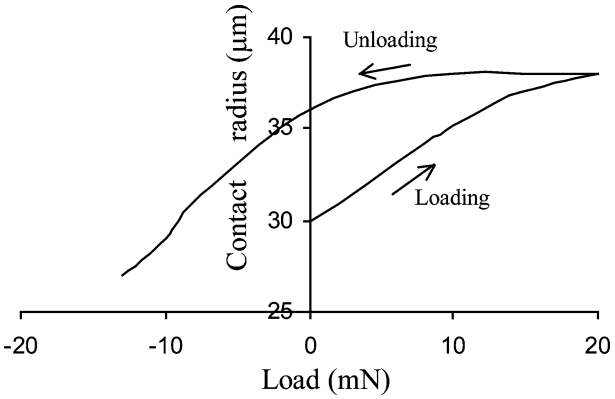



**Fig. 3.3.** Adhesion hysteresis. Adhesion force is different when surfaces are approaching contact and when separating for polystyrene (based on [211])

is always greater than the energy gained by bringing them together (Fig. 3.3). As a result, the energy is dissipated during the separation process. Adhesion hysteresis, or surface energy hysteresis, can arise even between perfectly smooth and chemically homogeneous surfaces supported by perfectly elastic materials. Adhesion hysteresis exists due to surface roughness and inhomogeneity [211].

The van der Waals force itself is conservative and does not provide a mechanism of energy dissipation. However, adhesion hysteresis due to surface heterogeneity and chemical reactions leads to dissipation [211, 284, 310, 351]. 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 $\Delta W$ is dissipated per unit area.

Since the underlying physical reason of 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 that combine the elastic deformation and adhesion [169], however, these theories do not address the issue of adhesion hysteresis.

Nosonovsky [233] considered a very simple model which, however, can account for adhesion hysteresis. Physically the van der Waals adhesion force and the elastic force are both caused by the atomic interaction. However, at the scale of nanometers, the contacting bodies can still be treated as a continuum, but the effects of adhesion forces are important [169]. The usual approach for the elasto-adhesive problems is to consider the bodies in contact as a continuum media and the interaction between them governed by an adhesive potential. In this section we will study a simple two-dimensional model of solid–solid contact with adhesion. It is expected that the two-dimensional model, while simple, can catch qualitatively the behavior during three-dimensional contact as well.

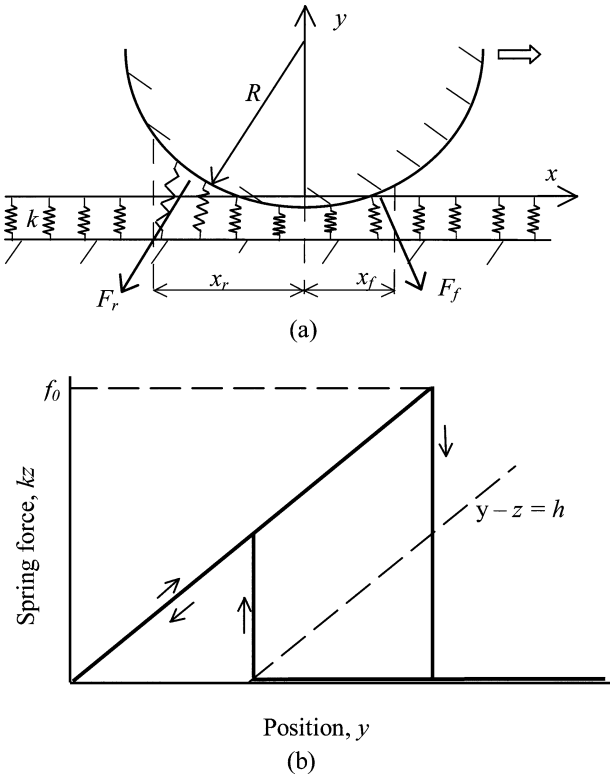


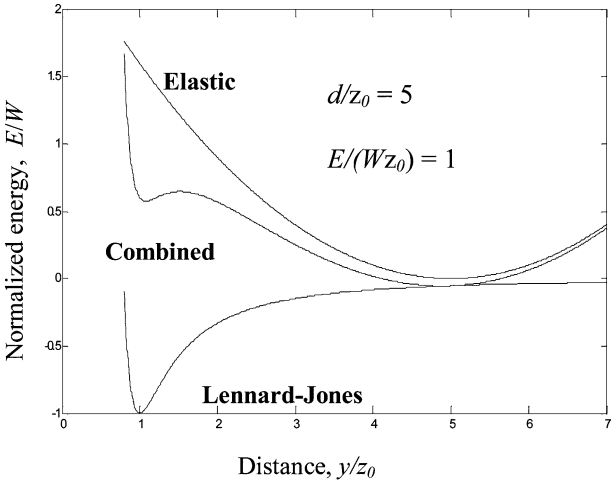

**Fig. 3.4. a** Schematics of a rigid spherical asperity sliding upon a deformable substrate (represented by springs), $z = d - y$, where $d = R - (R^2 - x^2)^{1/2}$, with adhesion force between them. Due to the hysteresis, the position of the springs on approach is different from that at detaching. **b** Dependence of the force, acting upon a spring, on the vertical position of the asperity $y$ during the motion (loading–unloading cycle). **c** The Lennard-Jones, elastic and combined potentials, with the combined potential having two minimums. **d** Normalized energy difference of the two equilibrium states, $\Delta W z_0 / W$, as a function of the normalized elastic modulus, $\alpha = E z_0 / W$ [233]

Consider a solid continuum deformable surface in contact with a rigid cylinder with the van der Waals adhesion force acting between them (Fig. 3.4(a)) and the separation distance $z$. The cylinder presents an asperity in contact with a substrate. The total energy, $T$, of a point at the surface is given by [233]

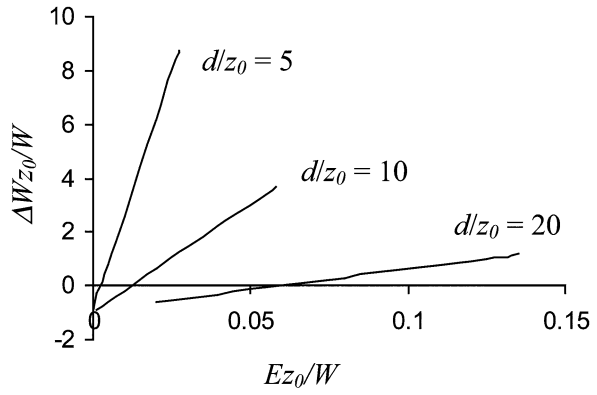
$$T = p(z) + W_{\mathrm{E}}, \tag{3.3}$$

where $p(z)$ is the Lennard-Jones adhesion potential for plane surfaces [169]

$$p_{\mathrm{a}}(z) = -\frac{W}{3}\left[\left(\frac{z_0}{z}\right)^8 - 4\left(\frac{z_0}{z}\right)^2\right], \tag{3.4}$$




(c)




(d)

**Fig. 3.4.** (*Continued*)

and $W_{\mathrm{E}}$ is the elastic energy, which can be approximated by

$$W_{\mathrm{E}} = \frac{Ey^2}{2z_0^2}, \tag{3.5}$$

where $y$ is the vertical displacement of the point (individual spring), $z$ is the distance between the point and the rigid asperity, $z_0$ is the equilibrium distance, and $E$ is the elastic modulus (Fig. 3.4(b)). Equation (3.5) represents a simplified linear elastic law, which may be visualized as a linear spring. Combining (3.3)–(3.5) and noting from Fig. 3.4(a) that $z = d - y$, where $d = R - (R^2 - x^2)^{1/2}$ is a constant distance at a given coordinate $x$, yields [233]

$$E(z) = -\frac{W}{3}\left[\left(\frac{z_0}{z}\right)^8 - 4\left(\frac{z_0}{z}\right)^2\right] + E\frac{(d-z)^2}{2z_0^2}. \tag{3.6}$$

As observed from Fig. 3.4(c), the combined potential can have two minimum points which correspond to equilibriums, and thus makes the hysteresis possible. The first equilibrium corresponds to the adhesion forces dominating over the elastic force and is achieved on approach when an element of the deformable surface "jumps to contact" with the rigid asperity. The second equilibrium corresponds to the elastic force dominating over the adhesion and is achieved at separation when an element of the surface detaches from the asperity. The energy barriers between the two states, $\Delta W$, are equal to the hysteresis. Note that even though both the adhesion and elastic forces are reversible, and the energy potential (3.4) is conservative, the hysteresis occurs in the system, which leads to a nonreversible process. The energy is consumed for excitation of elastic vibrations and waves [233].

The normalized energy difference of the two equilibrium states, $\Delta W z_0/W$, as a function of the normalized elastic modulus, $\alpha = Ez_0/W$, is presented in Fig. 3.4(d) for the values of the normalized distance $d/z_0 = 5, 10$, and 20. For $d/z_0 = 5$, there are two equilibriums when $0.0201 < \alpha < 0.1353$, which correspond to $1.01 < z/z_0 < 1.21$ and $3.75 < z/z_0 < 4.92$. Obviously, the first equilibrium (near $z/z_0 = 1$) corresponds to the substrate, just slightly deformed by the adhesion force, whereas the second equilibrium (near $z/z_0 = d/z_0 = 5$) corresponds to a significant deformation of the substrate, adhered to the asperity. For $d/z_0 = 10$, there are two equilibriums when $0.0012 < \alpha < 0.059$, which correspond to $1.001 < z/z_0 < 1.187$ and $7.99 < z/z_0 < 9.98$. For $d/z_0 = 20$, there are two equilibriums when $0.0001 < \alpha < 0.0273$, which correspond to $1.0013 < z/z_0 < 1.1923$ and $17.51 < z/z_0 < 19.997$. It is observed from Fig. 3.4(d) that the energy difference is almost linearly proportional to the normalized elastic modulus. This is because the energy of the second equilibrium (when the substrate is attached to the asperity) is greater than that of the first equilibrium, and the $W_e$ term, which is proportional to $E$, dominates [233].

#### 3.2.1.3 Shear Strength Due to 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, $\Delta W A_r$, is equal to the work of the friction force $F$ at the distance $d$; therefore, the friction force is given by [233]

$$F = A_r \Delta W/d. \tag{3.7}$$

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 manner similar to rolling. Based on the adhesion hysteresis approach, the frictional force during sliding is also given by (3.7),

and all considerations presented in the preceding section are also valid for the sliding friction.

However, it is well known from the experiments that sliding friction is usually greater than the rolling friction [31, 32]. 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 harder material. The shear strength during plowing is often assumed to be proportional to the average absolute value of the surface slope [31, 32]. It is therefore assumed that in addition to the adhesion hysteresis term, there is another component, $H_{\mathrm{p}}$, responsible for friction due to surface roughness and plowing [233]

$$F = A_{\mathrm{r}}(\Delta W/d + H_{\mathrm{p}}). \tag{3.8}$$

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 in (3.8) directly, however, $A_{\mathrm{r}}$ depends upon the normal load. The right-hand side of (3.8) involves two terms: a term that is proportional to adhesion hysteresis and a term that is proportional to roughness. Nosonovsky [233] pointed out the similarly of (3.8)—that governs energy dissipation during solid–solid friction—to the equations that govern energy dissipation during solid–liquid friction, which will be discussed in the next part of this book.

Summarizing, the adhesive friction provides the mechanism of energy dissipation due to breaking strong adhesive bonds between the contacting surfaces and due to adhesion hysteresis. The value of the force is given by (3.2). In order for adhesive friction to exist, either irreversible adhesion bonds should form or the contacting bodies should be deformable and thus nonideally rigid. The adhesive friction mechanism involves weak short-range adhesive force and strong long-range bulk forces.

### 3.2.2 Deformation of Asperities

Another important mechanism of friction is deformation of interlocking asperities ([30, 32], as shown in Fig. 3.2(b)). 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, a 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 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 plastic deformation and plowing. The transition from adhesive to deformational friction mechanism depends on load and yield strength of materials and usually results in a significant increase of the friction force [32].

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.

### 3.2.3 Plastic Yield

Chang et al. [74] proposed a single-asperity contact model of friction based on plastic yield, which was later modified by Kogut and Etsion [189]. 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. 3.2(c)). 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 Etsion [189] calculated the tangential load at the onset of sliding as a function of the normal load using the finite element 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.

### 3.2.4 Fracture

For brittle material, asperities can break forming wear debris. Therefore, fracture can also contribute to friction. There is also an analogy between mode II crack propagation and the sliding of an asperity [129, 178, 280] (Fig. 3.2(d)). When an asperity slides, the bonds are breaking at the rear, while new bonds are being created at the front. Thus, the rear edge of asperity can be viewed as the 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 [42]. Calculations have been performed to relate the stress intensity factors with friction parameters [129, 178, 280]. Crack and dislocation propagation along the interface implies that the interface is weak compared to the bulk of the body.

### 3.2.5 Ratchet and Cobblestone Mechanisms

Interlocking of asperities may result in one asperity climbing upon the other, leading to the so-called ratchet mechanism [30, 32]. In this case, in order to maintain sliding, a horizontal force should be applied which is proportional to the slope of the asperity (Fig. 3.2(e)). 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 the “cobblestone mechanism” [161]. This mechanism implies that the strong bonds are acting in the bulk of the body, whereas interface bonds are weak.

### 3.2.6 "Third Body" Mechanism

During the contact of two solid bodies, wear and contamination particles can be trapped at the interface between the bodies (Fig. 3.2(f)). Along with liquid, which condensates at the interface, these particles 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 [30, 32] and thus the third body mechanism leads to an increase in the coefficient of friction. At the atomic scale, adsorbed mobile molecules can constitute the "third body" and lead to significant friction increase [146]. The third body has much weaker bonds to the surface, than those in the bulk of the body.

### 3.2.7 Discussion

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 following chapters, we will discuss linearity of friction as a result of the presence of a small parameter, nonlinearity of friction, related to heterogeneity and hierarchical structure and multiscale nature of the frictional mechanisms.

## 3.3 Friction as a Linear Phenomenon

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

$$F = \mu W, \tag{3.9}$$

where $\mu$ 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, (3.9) 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. These two concepts are considered in the following sections.

### 3.3.1 Friction, Controlled by Real Area of Contact

Every nominally flat surface is not ideally smooth and has roughness due to small asperities. A contact between the two bodies during friction occurs only at the summits of the asperities. As a result, the real area of contact, $A_r$, constitutes only a small fraction of the nominal area of contact and depends upon the normal load. For metals at typical loads, 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 pioneering work by Greenwood and Williamson [137]. Using numerical computations, these models conclude 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 [3]. For the elastic contact of a smooth surface and a rough surface with the correlation length $\beta^*$ and standard deviation of profile height $\sigma$, the real area of contact is given by

$$A_r \propto \frac{\beta^*}{E^*\sigma} W, \tag{3.10}$$

where $E^*$ is the composite elastic modulus of the two bodies [32]. Note that $\sigma$ is the vertical and $\beta^*$ is the horizontal roughness parameters with the dimension of length. The smoother the surface (higher the ratio $\beta^*/\sigma$), the larger $A_r$. 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 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. 3.5).

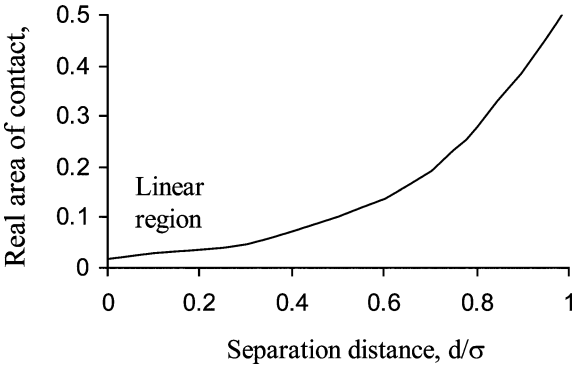



**Fig. 3.5.** The number of contacts and contact area as a function of separation between the contacting bodies (based on [257]

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 $H_s$ [32]

$$A_r = \frac{W}{H_s}. \tag{3.11}$$

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 depend upon the roughness parameters, elastic modulus, and hardness. Interestingly, Greenwood and Williamson [137] showed that whether the contact is elastic or plastic does not depend upon the load, but solely upon the so-called plasticity index $\psi = (\sqrt{\sigma/R_p})E^*/H$, where $\sigma$ is the standard deviation of peak heights and $R_p$ is mean asperity peak radius.

Based on Bowden and Tabor's model (Eq. (3.2)), the friction force due to adhesion is proportional to the real area of contact and adhesive shear strength $\tau_a$. Combining (3.2) and (3.9)–(3.11) yields a linear dependence of $F$ upon $W$.

Fractal models provide an alternative description of a rough surface. Long before the discovery of fractals by mathematicians, Archard [12] studied multiscale roughness with small asperities on top of bigger asperities, with even smaller asperities on top of those, and so on (Fig. 3.6). According to the Hertzian model, for the contact of an elastic sphere of radius $R$ with an elastic flat with the contact radius $a$, the contact area $A_r = \pi a^2$ is related to the normal load as

$$A_r = \pi\left(\frac{3RW}{4E^*}\right)^{2/3}. \tag{3.12}$$

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

$$p = \left(\frac{6WE^{*2}}{\pi^3 R^2}\right)^{1/3}\sqrt{1-(r/a)^2}. \tag{3.13}$$

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 a distance $r$ from the center, the load is proportional to the stress given by (3.13). The area of contact of this small asperity is still given by (3.12)

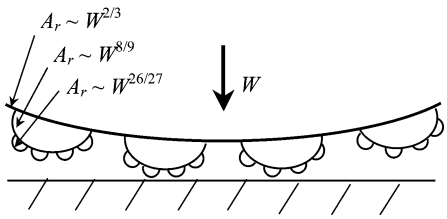



**Fig. 3.6.** A multiscale rough elastic surface in contact with a flat surface

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 [12]

$$\begin{aligned} A_\mathrm{r} &\propto \int_0^a \left[ W^{(1/3)} \sqrt{1 - r^2/a^2} \right]^{2/3} 2\pi r \, \mathrm{d}r \\ &\propto \int_0^\pi \left[ W^{(1/3)} \cos\phi \right]^{2/3} 2\pi (a \sin\phi) a \cos\phi \, \mathrm{d}\phi \\ &\propto W^{(2/9)} a^2 \propto W^{(2/9)} W^{(2/3)} \propto W^{(8/9)}. \end{aligned} \tag{3.14}$$

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 an even smaller radius, the total area of contact can be calculated in a similar way as

$$A_\mathrm{r} \propto \int_0^a \left[ W^{(1/3)} \sqrt{1 - r^2/a^2} \right]^{8/9} 2\pi r \, \mathrm{d}r \propto W^{(8/27)} a^2 \propto W^{(26/27)}. \tag{3.15}$$

For elastic contact, it is found that

$$A_\mathrm{r} \propto W^{(3^n - 1)/3^n}, \tag{3.16}$$

where $n$ is the number of orders of asperities, leading to an almost linear dependence of $A_\mathrm{r}$ upon $W$ with increasing $n$. Later more sophisticated fractal surface models were introduced, which led to similar results [213].

Thus, both statistical and fractal roughness for elastic and plastic contact, combined with the adhesive friction law (3.2) results in an almost linear dependence of the friction force upon the normal load.

### 3.3.2 Friction Controlled by Average Surface Slope

Another type of dry friction model is based on the assumption that during sliding asperities climb upon each other (the ratchet mechanism) (Fig. 3.7). 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

$$F = W \tan\theta, \tag{3.17}$$

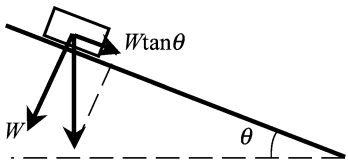



**Fig. 3.7.** Slope-controlled friction. For a body moving without acceleration upon an inclined surface with slope $\theta$, the shear force, $W \tan\theta$, is proportional to the normal load, $W$

where $\theta$ is the slope angle of the asperities. Comparing (3.9) and (3.17) it may be concluded that, for a rough surface, the coefficient of fiction is equal to the average absolute value of its slope

$$\mu = |\tan\theta|. \tag{3.18}$$

The sign of the absolute value appears in (3.18) because asperities can climb only if the slope is positive. Similar to the ratchet mechanism is the cobblestone mechanism, which is typical for atomic friction.

### 3.3.3 Other Explanations of the Linearity of Friction

Among other attempts to explain the linearity of the friction force with respect to the load, two modeling approaches are worth mentioning. Sokoloff [301] suggested that the origin of the friction force is in the hardcore atomic repulsion. The vertical component of the repulsion force's vector, which contributes to the normal load, is proportional to the horizontal component of the same vector, which contributes to friction because the vector has a certain average orientation. In a sense, this is still the same slope-controlled mechanism, but considered at the atomic level.

Ying and Hsu [348] suggested an interesting macroscale approach. They noticed that for a spherical asperity of radius $R$, slightly indented into a substrate, the contact radius, $a$, is proportional to the second power of the penetration $h$ (Fig. 3.8)

$$a \propto W^{1/3}. \tag{3.19}$$

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) $A_\mathrm{p}$ is given by a cubic function of $a$ and thus is proportional to the normal load

$$A_\mathrm{p} = \frac{2a^3}{3R} \propto W. \tag{3.20}$$

This is the case of "elastic plowing," the force resisting to sliding is proportional to $A_\mathrm{p}$ and, therefore, is linearly proportional to the normal load.

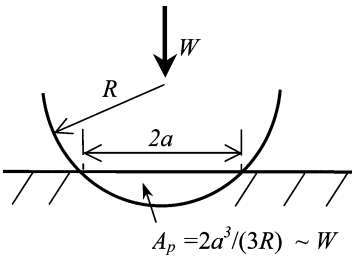



**Fig. 3.8.** "Elastic plowing:" the trans-sectional area of the asperity is linearly proportional to the Hertzian normal load

### 3.3.4 Linearity and the "Small Parameter"

We have found that several physical mechanisms result in a linear dependence of the friction force upon the normal load. Mathematically, a linear dependence between 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

$$F = f(W) \approx f(0) + f'(0)W + \frac{f''(0)}{2}W^2 = \mu W + \frac{f''(0)}{2}W^2 \tag{3.21}$$

the dependency can be linearized as $F = \mu W$ if

$$W \ll \frac{2\mu}{f''(0)}. \tag{3.22}$$

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.

## 3.4 Summary

In this chapter we considered several mechanisms of friction that result in a linear dependence of the friction force upon the normal load (Table 3.2). We also discussed the role of the small parameter in the linearity. In more general terms, linearity of the friction is a consequence of the interface forces being small compared to the binding forces acting in the bulk of the body. Since this ratio is small, the ratio of real to nominal areas of contact is also small, which guarantees validity of (3.10) based on the statistical models, as it was explained in the preceding sections. In a similar manner, the small extent of the contact at the interface, compared to the bulk of the material, provides the linear dependencies given by (3.10)–(3.13) and (3.20).

**Table 3.2.** Mechanisms of friction and linear dependence of the friction force upon the normal load

| | 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 | $F = \tau_a A_p = \frac{2a^3}{3R} \propto W$ |