Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pergamon
Acta Materialia 51 (2003) 4331–4345
Acta MATERIALIA
www.actamat-journals.com

Scale effects in friction using strain gradient plasticity and dislocation-assisted sliding (microslip)

Bharat Bhushan *, Michael Nosonovsky

Nanotribology Laboratory for Information Storage and MEMS/NEMS, Department of Mechanical Engineering, 206 West 18th Avenue, The Ohio State University, Columbus, OH 43210-1107, USA

Received 12 March 2003; received in revised form 6 May 2003; accepted 6 May 2003

Abstract

Scale effects in dry friction at macro- to nanoscale are considered. According to the adhesional model of friction, the friction force depends on the real area of contact and the average shear strength of asperity contacts during sliding. The scale dependence of the so-called geometrically necessary dislocations causes enhanced hardness with decreasing scale. In the case of plastic contacts, enhanced hardness results in a decrease in the real area of contact. The average shear strength at the interface is associated with dislocation-assisted sliding (microslip) and increases with decreasing scale, from geometrical considerations. In cases of single-asperity and multi-asperity elastic or plastic contact, the scale dependence of the real area of contact and shear strength results in scale-dependent friction. Comparisons of the model with experimental data are also presented.

Keywords: Strain gradient plasticity; Dislocation mobility; Plastic deformation; Slip; Friction

1. Introduction

Microscale and nanoscale measurements of tribological properties, which became possible due to the development of the Surface Force Apparatus (SFA), Atomic Force Microscope (AFM), and Friction Force Microscope (FFM), demonstrate scale dependence of adhesion, friction, and wear as well as mechanical properties including hardness [1–10]. It has also been reported that large strain gradients in small indentations lead to geometrically necessary dislocations for strain compatibility conditions, that cause enhanced yield strength and hardening [1,6]. A qualitative explanation of the scale-dependent hardness was proposed as follows: as the volume of deformed material decreases, there is a lower probability of encountering material defects [3]. A quantitative theory of scale effects in tribology remains to be developed. Advances of Microelectromechanical Systems (MEMS) technology in the past decade makes understanding of scale effects in tribology especially important, since surface to volume ratio grows with miniaturization and surface phenomena dominate. However, conventional theories of con-

* Corresponding author. Tel.: +1-614-292-0651; fax: +1-614-292-0325.
E-mail address: bhushan.2@osu.edu (B. Bhushan).

1359-6454/03/$30.00
doi:10.1016/S1359-6454(03)00261-1

Nomenclature

a, $\bar{a}$ contact radius, mean contact radius, respectively
A_r, A_{re}, A_{rp} real area of contact, real area of elastic contact, real area of plastic contact, respectively
b Burgers vector
c constant, specified by crystal structure
c_p coefficient of friction scaling parameter
C_E, C_P constants, related to L_{lwl}, L_E, L_P, m, and n, respectively
d separation distance between reference planes of two surfaces in contact
D interface zone thickness
E^* effective elastic modulus
F_e, F_p friction force for elastic contact, friction force for plastic contact, respectively
G elastic shear modulus
h indentation depth
H hardness
H_0 hardness in absence of strain gradient
l_d length that characterizes depth dependence of hardness
l_s average climbing distance
L, L_{lwl}, L_s, L_d scan size, long wavelength limit, and length parameters related to l_s and l_d, respectively
m, n indices of exponents
N number of contacts
s spacing between slip steps on the indentation surface
W normal load
β^* correlation length
Γ gamma function
ε strain
η_{int}, η_{cr} density of dislocation lines per interface area, critical density of dislocation lines per interface area, respectively
θ indentation angle
κ curvature
μ_e, μ_{eo}, μ_{nano}, μ_p, μ_{p0} coefficients of elastic friction, macroscale elastic friction, nanoscale friction, plastic friction, macroscale plastic friction, respectively
ρ, ρ_G, ρ_S total density of dislocation lines per volume, density of GND per volume, density of SSD per volume, respectively
σ standard deviation of profile heights
τ, τ_a, τ_{a0}, τ_p, τ_Y, τ_{Y0} shear strength, average shear strength during sliding, average shear strength during sliding in the limit of large a, Peierls stress, shear yield strength, shear yield strength in absence of strain gradient, respectively

tact and friction lack characteristic length parameters, which would be responsible for scale effects.

Friction is a complex phenomenon, which involves asperity interactions involving adhesion and deformation (plowing), Fig. 1. Adhesion and plastic deformation involve energy dissipation, responsible for friction. A contact between two bodies takes place on high asperities, and the real area of contact (A_r) is a small fraction of the appar-

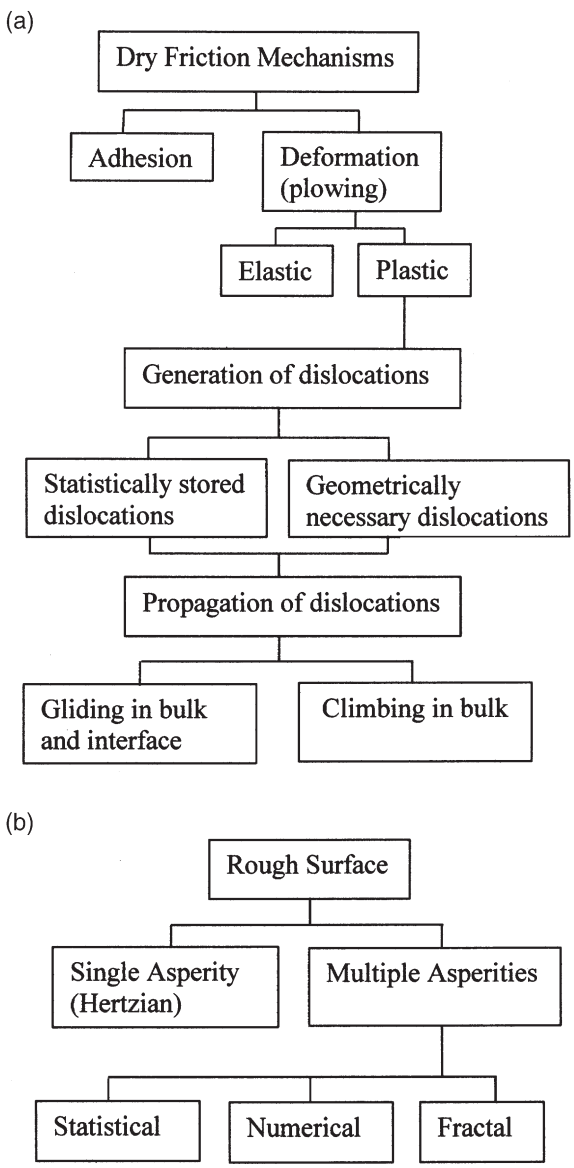

Fig. 1. (a) A block diagram showing generation and propagation of dislocations during sliding. (b) A block diagram of rough contact models.

ent area of contact [8]. It was shown by Greenwood and Williamson [11] and by subsequent modifications of their model, that for contacting surfaces with common statistical distributions of asperity heights, A_r is almost linearly proportional to the normal load. Based on classical theory of adhesion, the friction force (F) is defined as follows [8]:

$$F = \tau_a A_r \quad (1)$$

where τ_a is the average shear strength during sliding. A_r–load dependence and Eq. (1) result in linear dependence of the friction force on the normal load, or coefficient of friction being independent of the normal load. For a review of the numerical analysis of rough surface contacts, see Bhushan [12,13] and Bhushan and Peng [14]. The statistical and numerical theories of contact involve roughness parameters—e.g. the standard deviation of asperities heights and the correlation length [8]. The roughness parameters are scale dependent. In contrast to this, the theory of self-similar (fractal) surfaces solid contact developed by Majumdar and Bhushan [15] does not include length parameters and is scale-invariant in principle. The average shear strength of the contacts in Eq. (1) is also scale dependent. In addition to adhesion contribution to friction, elastic and plastic deformations on nano- to macroscale contribute to friction [8]. The deformations are scale dependent.

The linear elasticity and conventional plasticity theories are scale-invariant and do not include any material length scales. A strain gradient plasticity theory has been developed, for microscale deformations, in recent years by Fleck et al. [1], Nix and Gao [6], Hutchinson [16], and others, which is based on statistically stored and geometrically necessary dislocations (to be described later). Their theory predicts a dependence of mechanical properties on the strain gradient, which is scale dependent: the smaller the size of the deformed region, the greater the gradient of plastic strain, and, the greater the yield strength and hardness. Gao et al. [17] and Huang et al. [18] proposed a mechanism-based strain gradient (MSG) plasticity theory, which is based on a multi-scale framework, linking the microscale (10–100 nm) notion of statistically stored and geometrically necessary dislocations to the mesoscale (1–10 μm) notion of plastic strain and strain gradient. Bazant [19] analyzed scale effect based on the MSG plasticity theory in the limit of small scale, and found that corresponding nominal stresses in geometrically similar structures of different sizes depend on the size according to a power exponent law.

Mechanism of slip involves motion of large

number of dislocations, which is responsible for plastic deformation during sliding. Dislocations are generated and stored in the body and propagate under load. There are two modes of possible line (or edge) dislocation motion: gliding, when dislocation moves in the direction of its Burgers vector *b* by a unit step of its magnitude, and climbing, when dislocation moves in a direction, perpendicular to its Burgers vector, Fig. 2(a). Motion of dislocations can take place in the bulk of the body or at the interface. Due to periodicity of the lattice, a gliding dislocation experiences a periodic force, known as the Peierls force (Weertman and Weertman [20]). The Peierls force is responsible for keeping the dislocation at a central position between symmetric lattice lines and it opposes dislocation's gliding, Fig. 2(b). Therefore, an external force should be applied to overcome Peierls force resistance against dislocation's motion. Weertman [21] showed that a dislocation or a group of dislocations can glide uniformly along an interface between two bodies of different elastic properties. In continuum elasticity formulation, this motion is equivalent to a propagating interface slip pulse; however, the physical nature of this deformation is plastic, because dislocations' motion is irreversible. The local plastic deformation can occur at the interface due to concentration of dislocations even in the predominantly elastic contacts. Gliding of a dislocation along the interface results in a relative displacement of the bodies for a distance equal to the Burgers vector of the dislocation, whereas a propagating set of dislocations effectively results in dislocation-assisted sliding, or microslip, Fig. 3. Several types of microslip are known in the tribology literature [8], the dislocation-assisted sliding is one type of microslip, which propagates along the interface. Conventional mechanism of sliding is considered to be concurrent slip with simultaneous breaking of all adhesive bonds. Based on Johnson [22] and Hurtado and Kim [23], for contact sizes on the order of a few nanometers to a few micrometers, dislocation-assisted sliding is more energetically profitable than a concurrent slip. Their argument is based on the fact that experimental measurements with the SFA demonstrated that, for mica, frictional stress is of the same order as Peierls stress, which is required for dislocation's

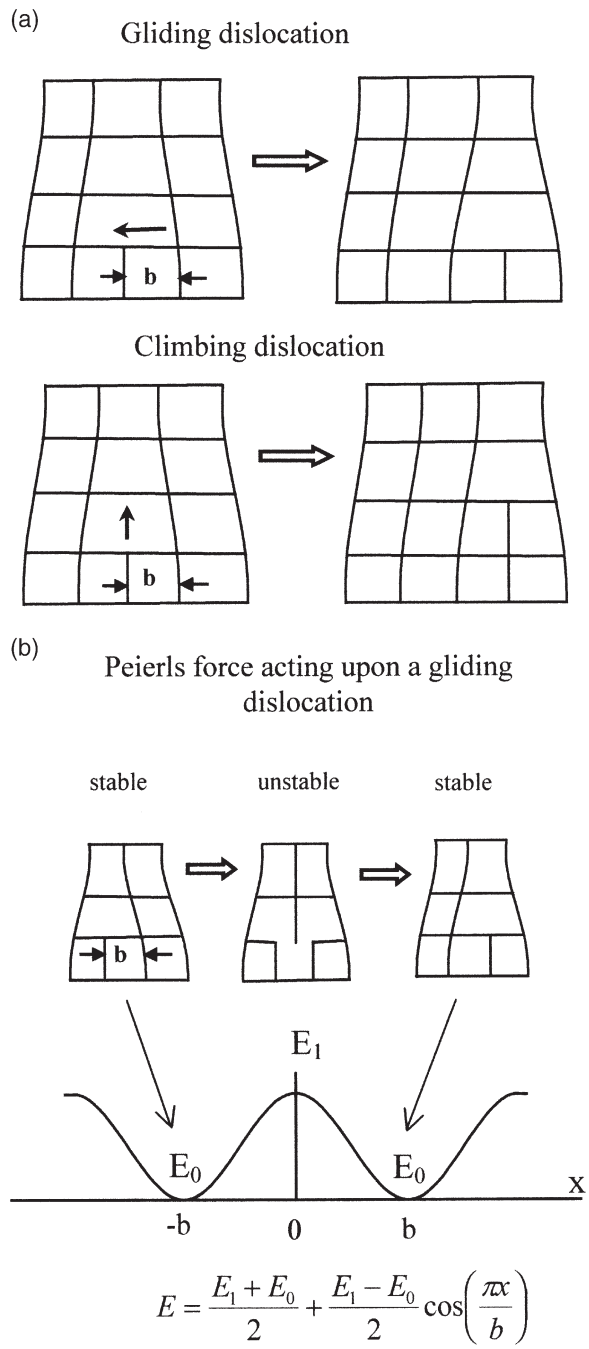

Fig. 2. (a) Schematics of gliding and climbing dislocations motion by a unit step of Burgers vector *b*. (b) Origin of the periodic force acting upon a gliding dislocation (Peierls force). Gliding dislocation passes locations of high and low potential energy.

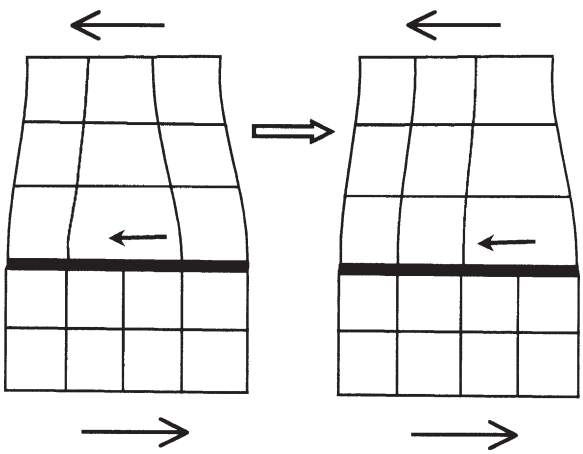

Fig. 3. Schematic showing microslip due to gliding dislocations at the interface.

gliding. They considered a complicated dislocation-assisted friction mechanism, which involves a transition from a concurrent slip of a circular asperity to a single dislocation-assisted slip by nucleation of a new dislocation ring, and then to multiple dislocation cooperated slip. Their model involves numerous parameters and they did not propose a simple scaling rule.

Polonsky and Keer [24] considered the pre-existing dislocation sources and carried out a numerical microcontact simulation based on contact plastic deformation representation in terms of discrete dislocations. They found that when the asperity size decreases and becomes comparable with the characteristic length of the material's microstructure (distance between dislocation sources), resistance to plastic deformation increases, which supports the conclusions drawn from strain gradient plasticity. Deshpande et al. [25] conducted discrete plasticity modeling of cracks in single crystals and considered dislocation nucleation from Frank–Read sources distributed randomly in the material. Pre-existing sources of dislocations, considered by all these authors, are believed to be a more realistic reason for increasing number of dislocations during loading, rather than completely newly nucleated dislocations [26].

In general, dislocations are emitted under loads from pre-existing sources and propagate along slip lines (Fig. 4). As shown in the figure, in regions of higher loads, the number of emitted dislocations is higher. Their approach was limited to numerical analysis of special cases.

Strain-gradient plasticity, applied for scale effects in the indentation deformation process, is well established. In the present work, scale effects in friction are modeled using the strain-gradient plasticity on contact size as well as dislocation-assisted sliding on shear strength. Strain-gradient plasticity combined with dislocation-assisted sliding has not been exploited to date to study scale effects in tribology. Comparisons of the model with experimental data are also presented.

2. Plasticity and dislocation-assisted sliding

In this section, scale dependence of deformation and dislocation-assisted sliding caused by gliding dislocations at the interface will be considered.

2.1. Plastic deformation and scaling

Plastic deformation occurs during asperity contacts because a small real area of contacts results in high contact stresses, which are often beyond the limits of applicability of the elasticity. As stated earlier, during loading, generation and propagation of dislocations are responsible for

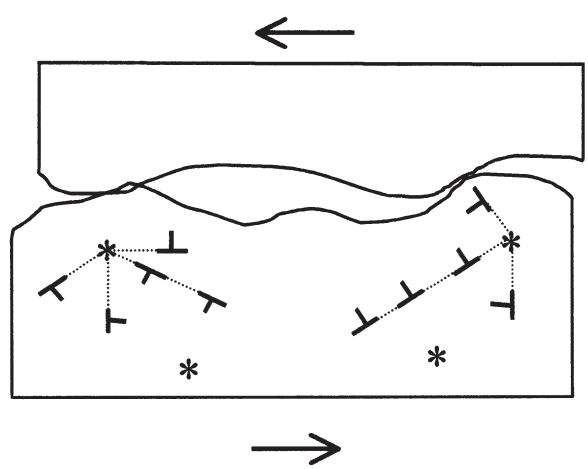

Fig. 4. Generation of dislocations from sources (∗) during plowing due to plastic deformation.

plastic deformation. Because dislocation motion is irreversible, plastic deformation provides a mechanism for energy dissipation during friction. The strain-gradient plasticity theories [1,6,16–18] consider two types of dislocations: randomly created Statistically Stored Dislocations (SSD) and Geometrically Necessary Dislocations (GND). The GND are required for strain compatibility reasons. For illustration, randomly created SSD during shear and GND during bending are presented in Fig. 5(a). The density of the GND (total length of dislocation lines per volume) during bending is proportional to the curvature κ and to the strain gradient

$$\rho_G = \frac{\kappa}{b} = \frac{1}{b}\frac{\partial \varepsilon}{\partial z} \propto \nabla \varepsilon, \tag{2}$$

where ε is strain and $\nabla\varepsilon$ is the strain gradient.

The GND during indentation, Fig. 5(b), are located in a certain sub-surface volume. The large strain gradients in small indentations require GND to account for the large slope at the indented surface. SSD, not shown here, also would be created and would contribute to deformation resistance, and are functions of strain rather than strain gradient. According to Nix and Gao [6], we assume that indentation is accommodated by circular loops of GND with Burgers vector normal to the plane of the surface. If we think of the individual dislocation loops being spaced equally along the surface of the indentation, then the surface slope

$$\tan\theta = \frac{h}{a} = \frac{b}{s} \tag{3}$$

where θ is the angle between the surface of the conical indenter and the plane of the surface, a is the contact radius, h is the indentation depth, b is the Burgers vector, and s is the spacing between individual slip steps on the indentation surface, Fig. 5(b). They reported that for geometrical (strain compartibility) considerations, the density of the GND is

$$\rho_G = \frac{3}{2bh}\tan^2\theta = \frac{3}{2b}\left(\frac{\tan\theta}{a}\right) = \frac{3}{2b}\nabla\varepsilon. \tag{4}$$

Thus, ρ_G is proportional to strain gradient (scale dependent) whereas the density of SSD, ρ_S, is

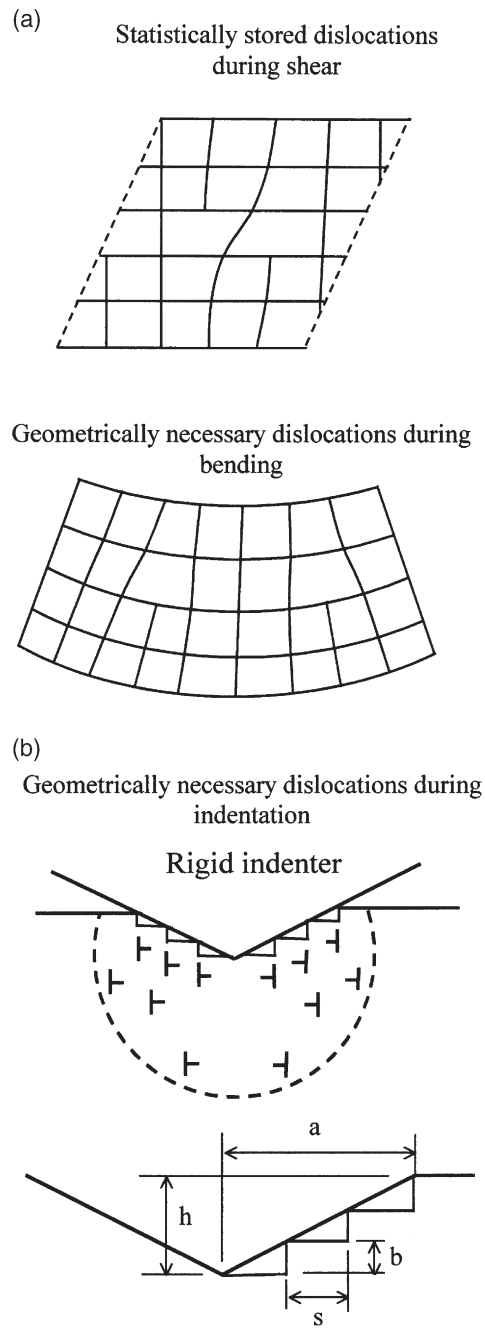

Fig. 5. (a) Illustration of statistically stored dislocations during shear and geometrically necessary dislocations during bending. (b) Geometrically necessary dislocations during indentation.

dependent upon the average strain in the indentation, which is related to the slope of the indenter (tan θ). Based on experimental observations, ρ_S is approximately proportional to strain [1].

According to the Taylor model of plasticity [26], dislocations are emitted from Frank–Read sources. Due to interaction with each other, the dislocations may become stuck in what is called the Taylor network, but when externally applied stress reaches the order of Peierls stress for the dislocations, they start to move and the plastic yield is initiated. The magnitude of the Peierls stress τ_p is proportional to the dislocation's Burgers vector b divided by a distance between dislocation lines s [20,26],

$$\tau_p = Gb/(2\pi s) \tag{5}$$

where G is the elastic shear modulus. An approximate relation of the shear yield strength τ_Y to the dislocation's density at a moment when yield is initiated is given by [26]

$$\tau_{0Y} = cGb/s = cGb\sqrt{\rho} \tag{6}$$

where c is a constant on the order of unity, specified by the crystal structure, and ρ is the total length of dislocation lines per volume, which is a complicated function of strain ε and strain gradient ($\nabla\varepsilon$)

$$\rho = \rho_S(\varepsilon) + \rho_G(\nabla\varepsilon). \tag{7}$$

The shear yield strength τ_Y can be written now as a function of SSD and GND densities [6],

$$\tau_Y = cGb\sqrt{\rho_S + \rho_G} = \tau_{Y0}\sqrt{1 + (\rho_G/\rho_S)} \tag{8}$$

where

$$\tau_{Y0} = cGb\sqrt{\rho_S} \tag{9}$$

is the shear yield strength value in the limit of small ρ_G/ρ_S ratio (large scale) that would arise from the SSD, in the absence of GND. Note that the ratio of the two densities is defined by the problem geometry and is scale dependent. Based on the relationships for ρ_G (Eq. (4)) and ρ_S, the ratio ρ_G/ρ_S is inversely proportional to a and Eq. (8) reduces to

$$\tau_Y = \tau_{Y0}\sqrt{1 + (l_d/a)} \tag{10}$$

where l_d is a plastic deformation length that characterizes depth dependence on shear yield strength. According to [16], this length is physically related to an average distance a dislocation travels, which was experimentally determined to be between 0.2 and 5 μm for copper and nickel. Note that l_d is a function of the material and the asperity geometry and is dependent on SSD. The ratio (l_d/a)will be referred to as a characteristic depth dependence parameter.

Using von Mises yield criterion, hardness $H = 3\sqrt{3}\tau_Y$. From Eq. (10), the hardness is also scale dependent [6]:

$$H = H_0\sqrt{1 + (l_d/a)} \tag{11}$$

where H_0 is the hardness in absence of strain gradient. Eq. (10) provides dependence of the resistance force to deformation upon the scale in a general case of plastic deformation.

2.2. Dislocation-assisted sliding and scaling

Now we consider a sliding contact between two bodies. Slip along the contact interface is an important special case of plastic deformation. The local dislocation-assisted microslip can exist even if the contact is predominantly elastic due to concentration of dislocations at the interface. Due to these dislocations, the stress at which yield occurs at the interface is lower than shear yield strength in the bulk. This means that average shear strength at the interface is lower than in the bulk.

An assumption that all dislocations produced by externally applied forces are distributed randomly throughout the volume would result in vanishing small probability for a dislocation to be exactly at the interface. However, many traveling (gliding and climbing) dislocations will be stuck at the interface as soon as they reach it. As a result of this, a certain number of dislocations will be located at the interface. In order to account for a finite dislocation density at the interface, we assume that the interface zone has a finite thickness D. Dislocations within the interface zone may reach the contact surface due to climbing and contribute into the microslip. In the case of a small contact radius a, compared to the average distance dislocations can climb l_s, interface zone thickness

D, which is scale dependent, is approximately equal to a. However, in the case of a large contact radius, D is approximately equal to l_s. An illustration of this is provided in Fig. 6. The depth of the subsurface volume, from which dislocations have a high chance to reach the interface, is limited by l_s and by a, respectively, for the two cases considered here. Based on these geometrical considerations, an approximate relation can be written as

$$D = \frac{al_s}{l_s + a} \tag{12}$$

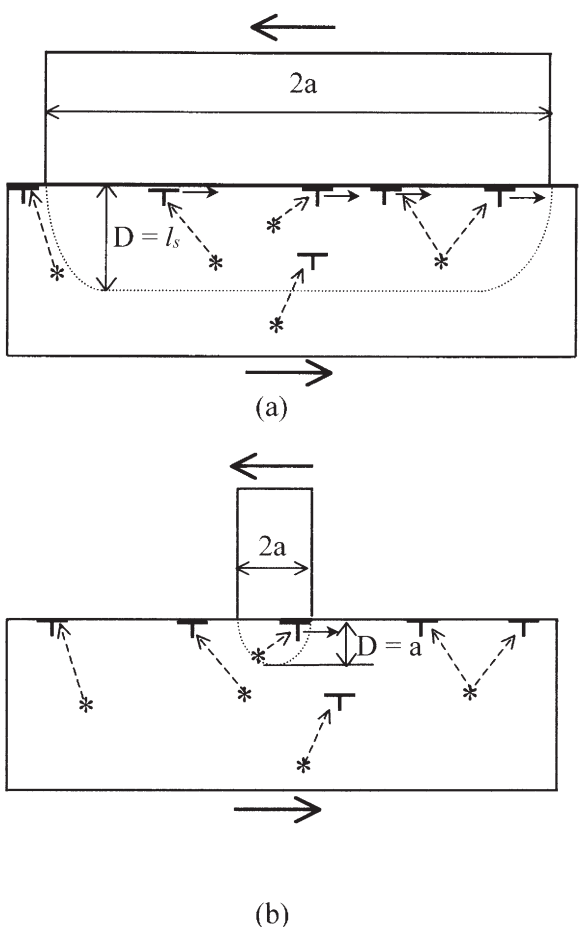

Fig. 6. Gliding dislocations at the interface generated from sources (∗). Only dislocations generated within the interface zone can reach the interface. (a) For a large contact radius a, thickness of this zone D is approximately equal to the average distance dislocations climb l_s. (b) For small contact radius a, the thickness of the interface zone is approximately equal to a.

The interface density of dislocations (total length of dislocation lines per interface area) is related to the volume density as

$$\eta_{int} = \rho D = \rho\left(\frac{al_s}{l_s + a}\right) \tag{13}$$

During sliding, dislocations must be generated at the interface with a certain critical density $\eta_{int} = \eta_{cr}$. The corresponding shear strength during sliding can be written following Eq. (10) as

$$\tau_a = \tau_{a0}\sqrt{1 + (l_s/a)} \tag{14}$$

where

$$\tau_{a0} = cGb\sqrt{\frac{\eta_{cr}}{l_s}} \tag{15}$$

is the shear strength during sliding in the limit of $a \gg l_s$. The ratio (l_s/a) is the characteristic microslip parameter. Similarity between Eq. (10), obtained from strain-gradient plasticity, and Eq. (14) should be noted.

3. Adhesional friction and scaling in single-asperity contact

The real area of contact A_r is related to the contact radius in the case of a circular contact as

$$A_r = \pi a^2 \tag{16}$$

For a single-asperity contact and elastic deformation, based on the Hertz analysis, the contact radius is related to the normal load [8],

$$a = \left(\frac{3WR}{4E^*}\right)^{1/3} \tag{17}$$

where R is asperity tip radius of curvature, $E*$ is the effective elastic modulus of the two bodies, and W is the normal load. Combining Eqs. (1), (14) and (16), we get an expression for friction force for a single-asperity elastic contact:

$$F_e = \pi\tau_0 a^2\sqrt{1 + (l_s/a)}. \tag{18}$$

The coefficient of elastic friction can be written as

$$\mu_e = \mu_{e0}\sqrt{1 + (l_s/a)} \tag{19}$$

where μ_0 is the coefficient of macroscale elastic friction. It should be noted that elastic modulus needed for a (Eq. (17)) is scale independent.

In the case of predominantly plastic deformation, the contact area can be calculated according to

$$A_{rp} = \frac{W}{H} \tag{20}$$

where H is hardness of the softer material [8]. In accordance with the strain-gradient plasticity approach, the hardness depends on the contact radius (Eq. (9)). The relationship between the normal load and the friction force can be presented as

$$F_p = A_{rp}\tau = \frac{W\tau_0\sqrt{1 + (l_s/a)}}{H_0\sqrt{1 + (l_d/a)}}. \tag{21}$$

The coefficient of friction can be written as

$$\mu_p = \mu_{p0}\sqrt{\frac{1 + (l_s/a)}{1 + (l_d/a)}} \tag{22}$$

where μ_{0p} is a coefficient of macroscale plastic friction. A coefficient of friction scaling parameter

$$c_p = \frac{\mu_{nano}}{\mu_{0p}} = \sqrt{\frac{l_s}{l_d}} \tag{23}$$

is responsible for coefficient of friction scaling in transition from nanoscale μ_{nano} (in the limit of $a \to 0$) to macroscale μ_{0p} (in the limit of $a \to \infty$) in the case of predominantly plastic contact.

It is known from experimental data that the coefficient of friction tends to decrease with decreasing scale. The experimental data for coefficients of nano/microscale and macroscale friction [27] are presented in Table 1. It must be assumed, therefore, based on Eq. (23), that $l_d > l_s$.

Table 1
Surface roughness (standard deviation of surface heights or σ) and coefficients of friction on nano- and macroscales of various samples in air [27]

Material	σ (nm)	Coefficient of nanoscale friction vs. Si_3N_4 tip[a]	Coefficient of macroscale friction vs. Si_3N_4 ball[b]
Graphite (HOPG)	0.09	0.006	0.1
Natural diamond	2.3	0.04	0.2
Si(1 0 0)	0.14	0.07	0.4

[a] Tip radius of about 50 nm in the load range of 10–150 nN (2.5–6.1 GPa) and a scanning speed of 0.5 nm/s and scan area of 1 nm × 1 nm for HOPG and a scanning speed of 5 μm/s and scan area of 1 μm × 1 μm for diamond and Si(1 0 0).

[b] Ball radius of 3 mm at a normal load of 1 N (0.6 GPa) and average sliding speed of 0.8 mm/s.

4. Multi-asperity contact

4.1. Elastic contact

Any nominally flat surface has a number of asperities and the contacts take place only on the top of some of the asperities. A rough surface with Gaussian height distribution can be characterized by the standard deviation of profile height σ and correlation length β^* [8]. The correlation length β^* is a spatial parameter and it can be viewed as a measure of randomness, since the degree of randomness of surface increases with an increase in the magnitude of β^*. The correlation length is also a measure of spatial decay of any asperity, since profile heights at any two points separated by a distance greater than β^* are statistically independent. Therefore, β^* is responsible for the horizontal scale of the profile, whereas σ is responsible for vertical scale of the profile.

Surface roughness is found at scales ranging from atomic dimensions to millimeter scale [8]. A surface is composed of a large number of length scales of roughness that are superimposed on each other. Therefore, on a surface, it is not that different asperities come in different sizes, but that one asperity comes in different sizes. This means that asperity is not a "definite object". Distribution of size and shape of asperities is dependent on the short- and long-wavelength limit of measurement. An asperity relevant in contact mechanics is defined as that which makes a contact in a particular application and carries some load. The length of the nominal contact zone or scan size L defines the long-wavelength limit in contact problems. Based on AFM measurements, both σ and β^*

initially increase with the scan size and then approach a constant value, when the scan size is approaching the long-wavelength limit L_{lwl}, Fig. 7. Based on the data, presented in Fig. 7, it can be noted that for glass–ceramic disk, long-wavelength limit is about $L_{lwl} \sim 25$ μm[8]. Based on the data of Fig. 7, let us assume that σ and β^* depend on the scan size according to an exponential law,

$$\sigma \propto L^n \qquad \beta^* \propto L^m \tag{24}$$

where n and m are indices of corresponding exponents.

The mean of surface height distribution corresponds to the so-called reference plane of the surface. Separation d is a distance between reference planes of two surfaces in contact, normalized by σ. For a given d and statistical distribution of surface heights, the total real area of contact (A_r), number of contacts (N), and elastic normal load W_e can be found, using statistical analysis of contacts. The real area of contact, number of contacts and elastic normal load are related to the separation distance d [28]:

$$A_r \propto F_A(d) \qquad N \propto \frac{1}{(\beta^*)^2} F_N(d) \qquad W_e \propto \frac{E^* \sigma}{\beta^*} F_W(d) \tag{25}$$

where $F_A(d)$, $F_N(d)$, and $F_W(d)$ are integral functions defined by Onions and Archard [28]. It should be noted that A_r and N as a function of d are prescribed by the contact geometry (σ, β^*) and do not depend on whether the contact is elastic or plastic. Based on Onions and Archard data, it is observed that the ratio F_W/F_A is almost constant for moderate $d < 1.4$ and increases slightly for $d > 1.4$. The ratio F_A/F_N decreases rapidly with d and becomes almost constant for $d > 2.0$. For moderate loads, the contact is expected to occur on the upper parts of the asperities ($d > 2.0$), and a linear proportionality of $F_A(d)$, $F_N(d)$, and $F_W(d)$ can be assumed. The mean contact radius can be calculated as

$$\bar{a} \propto \sqrt{\frac{A_r}{N}} \propto \beta^* \propto L^m. \tag{26}$$

Based on Eq. (25) and the observation that F_W/F_A is almost constant, elastic real area of contact is given by

$$A_{re} \propto \frac{\beta^*}{E^* \sigma} W \propto W L^{m-n}. \tag{27}$$

Using Eqs. (1), (14), and (27), the friction force

$$F_e = A_{re} \tau_a \propto W L^{m-n} \sqrt{1 + (L_s/L)^m} \tag{28}$$

where L_s is a constant dependent upon l_s and the coefficient of proportionality in Eq. (26).

$$L_s = \left(\frac{L^m}{\bar{a}} l_s \right)^{1/m} \tag{29}$$

is a constant, related to l_s. Dependence of the coefficient of friction on scale length L in the domain of $L < L_{lwl}$ for a given nominal pressure is

$$\mu_e = \frac{F_e}{W} = \mu_{e0} C_E L^{m-n} \sqrt{1 + (L_s/L)^m} \tag{30}$$

where μ_{e0} is value of the coefficient of friction at macroscale (or at $L \geq L_{lwl}$) and C_E is a constant

$$C_E = L_{lwl}^{n-m} / \sqrt{1 + (L_s/L_{lwl})^m} \tag{31}$$

chosen in order to satisfy $\mu = \mu_{e0}$ at $L = L_{lwl}$.

The values of m and n can be determined from the measurements as a function of scan size. For example, for the data in Fig. 7, $m = 0.5$ and $n = 0.2$.

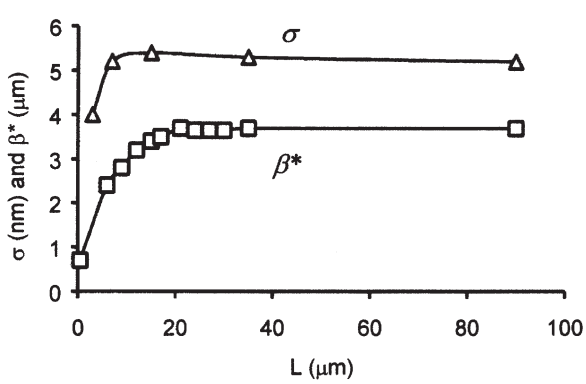

Fig. 7. Roughness parameters as a function of scan size for a glass–ceramic disk measured using AFM [8].

4.2. Plastic contact

In the case of plastic contact Eq. (26), which is based on the contact geometry, still holds. The real area of contact should be determined from Eq. (20). Using Eqs. (1), (11), (14) and (20), the friction force

$$F_p = A_{rp}\tau \tag{32}$$
$$= \frac{W\tau_{a0}\sqrt{1 + (l_s/\bar{a})}}{H_0\sqrt{1 + (l_d/\bar{a})}} \propto W\sqrt{\frac{1 + (L_s/L)^m}{1 + (L_d/L)^m}}$$

where L_d is a constant dependent on l_d and the coefficient of proportionality in Eq. (26).

$$L_d = \left(\frac{L^m}{\bar{a}} l_d\right)^{1/m} \tag{33}$$

Dependence of the coefficient of friction on scale length L in the domain of $L < L_{wl}$ for given nominal pressure is

$$\mu_p = \frac{F_p}{W} = \mu_{p0} C_P \sqrt{\frac{1 + (L_s/L)^m}{1 + (L_d/L)^m}} \tag{34}$$

where μ_{p0} is the value of the coefficient of friction at macroscale and C_P is a constant

$$C_P = \sqrt{\frac{1 + (L_d/L_{wl})^m}{1 + (L_s/L_{wl})^m}} \tag{35}$$

chosen in order to satisfy $\mu = \mu_{p0}$ at $L = L_{wl}$.

The scale effects predicted by the proposed models so far for single-asperity and multi-asperity contacts are summarized in Table 2 and presented in Fig. 8. In the case of multi-asperity elastic contact, change of scale affects the coefficient of friction in two different ways: through the change of τ_a and A_{re}. As seen from Fig. 8, by comparison of the curve with $L_s/L_{wl} = 0$ (no scale effect on τ_a), $L_s/L_{wl} = 1$, and $L_s/L_{wl} = 1000$ (significant scale effect on τ_a), the results for the normalized coefficient of friction are close; thus, the main contribution to the scaling effect is due to change of the contact area. In the case of plastic multi-asperity contact, the results presented are for $L_d/L_s = 0.25$, $L_d/L_s = 5$ and $L_d/L_{wl} = 1$ and $L_d/L_{wl} = 1000$.

5. Comparison with AFM/SFA data and discussion

5.1. Single-asperity, predominantly elastic contact

Nanoscale dependence of friction force upon the normal load was studied for Pt-coated AFM tip vs. mica in ultra-high vacuum (UHV) by Carpick et al. [29], for Si tip vs. diamond and amorphous carbon by Schwarz et al. [30] and for Si_3N_4 tip on Si, SiO_2, and diamond by Bhushan and Kulkarni [2], Fig. 9(a). Homola et al. [31] conducted SFA experiments with mica rolls with a single contact zone (before onset of wear), Fig. 9(b). Contacts relevant in these experiments can be considered as single-asperity, predominantly elastic in all of these cases. For a single-asperity elastic contact of radius a, expression for F is given by Eq. (18). For the limit of a small contact radius $a \ll l_s$, Eq. (14) combined with the Hertzian dependence of the contact area upon the normal load (Eq. (17)) yields

$$F_e \approx \pi a^2 \tau_0 \sqrt{l_s/a} \propto a^{3/2} \propto W^{1/2}. \tag{36}$$

If an adhesive pull-off force W_0 is large, Eq. (29) can be modified as

$$F_e = C_0\sqrt{W + W_0} \tag{37}$$

Table 2
Scaling factors for elastic and plastic deformations

Single-asperity elastic contact	Single-asperity plastic contact	Multiple-asperity elastic contact	Multiple-asperity plastic contact
$\mu_e = \mu_{e0}\sqrt{1 + (l_s/a)}$	$\mu_e = \mu_{e0}\sqrt{\frac{1 + (l_s/a)}{1 + (l_d/a)}}$	$\mu_e = \mu_{e0} C_E L^{m-n}\sqrt{1 + (L_E/L)^m}$	$\mu_p = \mu_{p0} C_P \sqrt{\frac{1 + (L_s/L)^m}{1 + (L_d/L)^m}}$

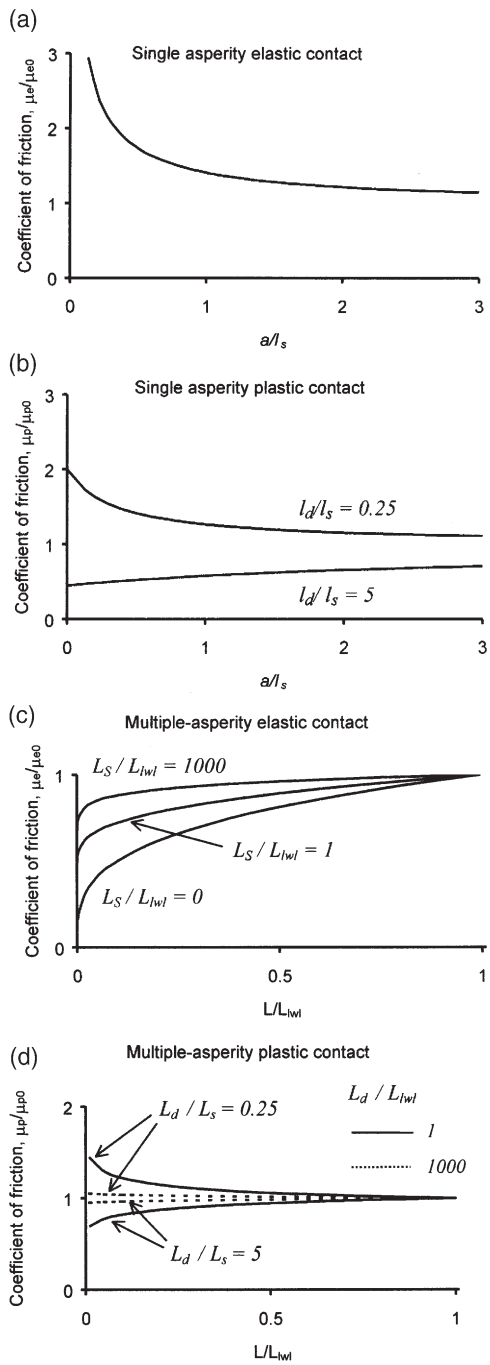

Fig. 8. Normalized results of the coefficient of friction, as a function of scale (a/l_s for single-asperity contact and L/L_{lwl} for multi-asperity contact). In the case of single-asperity plastic contact, data are presented for two values of l_d/l_s. In the case of multi-asperity contact, data are presented for $m-n = 0.3$. For multi-asperity elastic contact, data are presented for three values of L_S/L_{lwl}. For multi-asperity plastic contact, data are presented for two values of L_d/L_s.

where C_0 is a constant. Friction force increases with square root of the normal load, opposed to the two third exponent in scale independent analysis.The results in Fig. 9 demonstrate a reasonable agreement of the experimental data with the model. The platinum-coated tip vs. mica [29] has a relatively high pull-off force and the data fit with $C_0 = 23.7$ $(\text{nN})^{1/2}$ and $W_0 = 170$ nN. For the silicon tip vs. amorphous carbon and natural diamond [30], the fit is given by $C_0 = 8.0$, 19.3 $(\text{nN})^{1/2}$ and small W_0. For the virgin Si (1 1 1), SiO_2, and natural diamond sliding versus Si_3N_4 tip [2], the fit is given by $C_0 = 0.40$, 0.76, 0.86 $(\text{nN})^{1/2}$ for Si(1 1 1), SiO_2, and diamond, respectively, and small W_0. For two mica rolls [31], the fit is given by $C_0 = 10$ $\text{N}^{1/2}$ and $W_0 = 0.5$ N.

AFM experiments provide data on nanoscale, whereas SFA experiments provide data on microscale. Next we study scale dependence on the shear strength based on these data. In the AFM measurements by Carpick et al. [29], the average shear strength during sliding for Pt–mica interface was reported as 0.86 GPa, whereas the pull-off contact radius was reported as 7 nm. In the SFA measurements by Homola et al. [31], the average shear strength during sliding for mica–mica interface was reported as 25 MPa, whereas the contact area during high loads was on the order of 10^{-8} m^2, which corresponds to a contact radius on the order 100 μm. To normalize shear strength, we need shear modulus. The shear modulus for mica is $G_{\text{mica}} = 25.7$ GPa and for Pt is $G_{\text{Pt}} = 63.7$ GPa. For mica–Pt interface, the effective shear modulus is

$$G = 2G_{\text{mica}}G_{\text{Pt}}/(G_{\text{mica}} + G_{\text{Pt}}) = 36.6 \text{ GPa}. \quad (38)$$

This yields the value of the shear stress normalized by the shear modulus $\tau_a/G = 2.35 \times 10^{-2}$ for Carpick et al. [29] AFM data and 9.73×10^{-4} for the

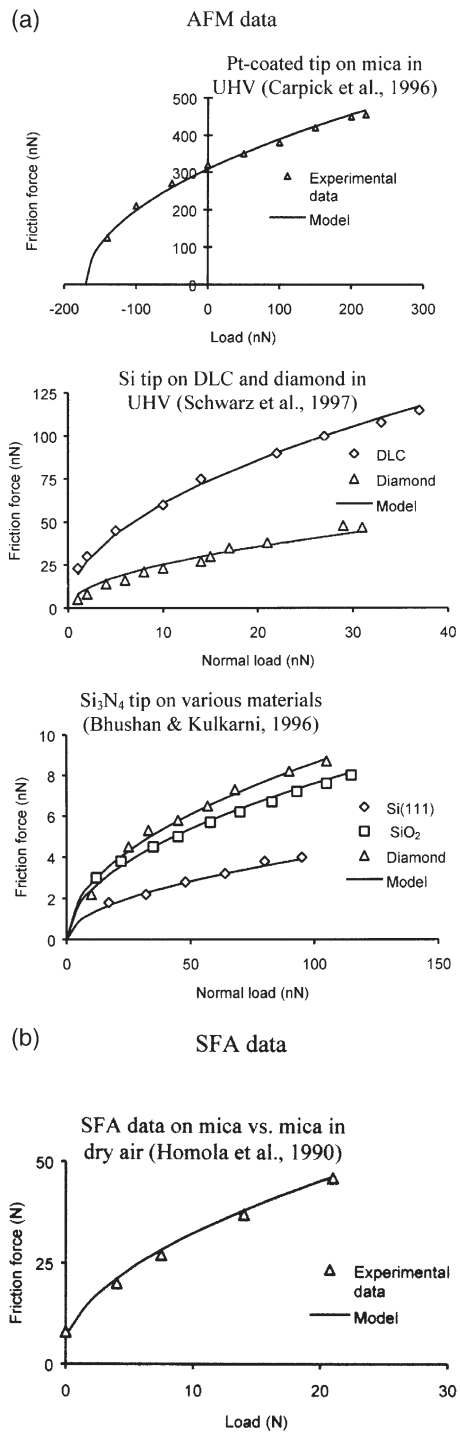

Fig. 9. Summary of (a) AFM data and (b) SFA data for friction force as a function of normal load.

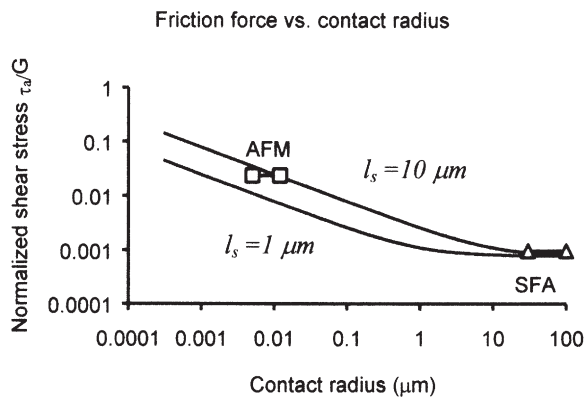

Fig. 10. Shear stress as a function of contact radius. Microscale and nanoscale data compared with the model for l_s = 1 μm and l_s = 10 μm.

SFA data. These values are presented in the Fig. 10 together with the values predicted by the model for assumed values of l_s = 1 and 10 μm. It can be seen that the model provides an explanation of friction increase with a scale decrease.

5.2. Transition from elastic to plastic and single-asperity to multiple-asperity contact

Bhushan and Kulkarni [2] studied Si_3N_4 tip vs. Si, SiO_2, and natural diamond for higher loads, Fig. 11. They reported that the coefficient of friction increases with increasing load after a certain load. It is noted that the critical value of loads for Si and SiO_2 corresponds to stresses equal to their hardness

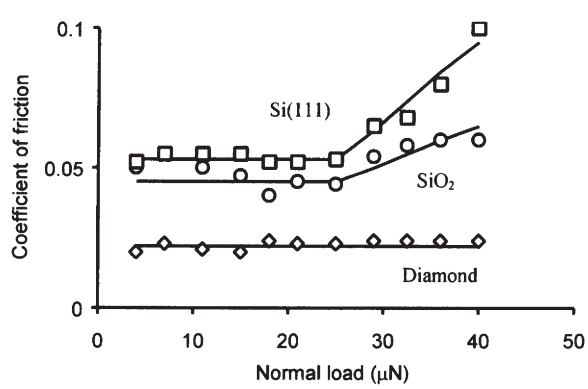

Fig. 11. Coefficient of friction as a function of normal load [2].

values, which suggests that transition to plasticity plays a role in this effect. The friction values at higher loads for Si and SiO_2 approach that of macroscale values. The data in Table 1 indicate that the coefficient of nanoscale friction is smaller than that at the macroscale [8].

In the data in Fig. 11 and Table 1, contacts at very low loads in AFM experiments are expected to be single-asperity and predominantly elastic. Whereas at high loads in AFM experiments and macroscale experiments, contacts are expected to be multiple-asperity and predominantly plastic. It is this transition which is responsible for an increase of the coefficient of friction. Further modeling is necessary to study the transition effect.

6. Conclusions

A model, which explains scale effects on friction, has been presented. During the contact of two bodies, adhesion and elastic and plastic deformation occur at contact. Adhesion and plastic deformation involve energy dissipation responsible for friction. According to the adhesional model of friction, the friction force depends on the real area of contact and the shear strength of asperity contacts during sliding. Plastic deformation involves propagation and generation of dislocations in the bulk and at the interface. Based on the strain-gradient plasticity approach, two types of dislocations exist: statistically stored dislocations and geometrically necessary dislocations. SSD are created by homogeneous strain, whereas large strain gradients in small deformation regions lead to GND. The scale dependence of GND causes enhanced hardness with decreasing scale. In the case of plastic contacts, enhanced hardness results in a decrease in the real area of contact. The average shear strength at the interface is lower than in the bulk, due to concentration of dislocations at the interface, which is scale dependent from geometrical considerations. Lower strength facilitates sliding, referred as dislocation-assisted sliding or microslip. The average shear strength increases with decreasing scale.

The model predicts that, in the case of single-asperity elastic contact, the coefficient of friction increases with decreasing scale. In the case of multi-asperity elastic contact, the coefficient of friction decreases with decreasing scale. In the case of single-asperity or multi-asperity plastic contact, the coefficient of friction may increase or decrease with scale, depending on the material parameters. The proposed model for single-asperity elastic contact is consistent with the existing experimental data for macro- to nanoscale friction measurements made by the AFM and SFA.

References

[1] Fleck NA, Muller GM, Ashby MF, Hutchinson JW. Strain gradient plasticity: theory and experiment. Acta Metall Mater 1994;42:475–87.
[2] Bhushan B, Kulkarni AV. Effect of normal load on microscale friction measurements. Thin Solid Films 1996;278:49–56 [vol. 293, p. 333].
[3] Bhushan B, Israelachvili JN, Landman U. Nanotribology: friction, wear and lubrication at the atomic scale. Nature 1995;374:607–16.
[4] Bhushan B, Kulkarni AV, Bonin W, Wyrobek JT. Nano/picoindentation measurement using a capacitive transducer system in atomic force microscopy. Philos Mag 1996;74:1117–28.
[5] Bhushan B, Sundararajan S. Micro/nanoscale friction and wear mechanisms of thin films using atomic force and friction force microscopy. Acta Mater 1998;46:3793–804.
[6] Nix WD, Gao H. Indentation size effects in crystalline materials: a law for strain gradient plasticity. J Mech Phys Solids 1998;46:411–25.
[7] Bhushan B. Nanoscale tribophysics and tribomechanics. Wear 1999;225-229:465–92.
[8] Bhushan B. Introduction to tribology. NY: Wiley, 2002.
[9] Bhushan B, Dandavate C. Thin-film friction and adhesion studies using atomic force microscopy. J Appl Phys 2000;87:1201–10.
[10] Sundararajan S, Bhushan B. Development of AFM-based techniques to measure mechanical properties of nanoscale structures. Sensors Actuators A 2002;101:338–51.
[11] Greenwood JA, Williamson JBP. Contact of nominally flat surfaces. Proc R Soc London 1966;A295:300–19.
[12] Bhushan B. Contact mechanics of rough surfaces in tribology: single asperity contact. Appl Mech Rev 1996;49:275–98.
[13] Bhushan B. Contact mechanics of rough surfaces in tribology: multiple asperities contact. Tribol Lett 1998;4:1–35.
[14] Bhushan B, Peng W. Contact modeling of multilayered rough surfaces. Appl Mech Rev 2002;55:435–80.
[15] Majumdar A, Bhushan B. Fractal model of elastic–plastic contact between rough surfaces. ASME J Tribol 1991;113:1–11.

[16] Hutchinson JW. Plasticity at the micron scale. Int J Solids Struct 2000;37:225–38.
[17] Gao H, Huang Y, Nix WD, Hutchinson JW. Mechanism-based strain-gradient plasticity—I. Theory. J Mech Phys Solids 1999;47:1239–63.
[18] Huang Y, Gao H, Nix WD, Hutchinson JW. Mechanism-based strain-gradient plasticity—II. Analysis. J Mech Phys Solids 2000;48:99–128.
[19] Bazant ZP. Scaling of dislocation-based strain-gradient plasticity. J Mech Phys Solids 2002;50:435–48.
[20] Weertman J, Weertman JR. Elementary dislocations theory. New York: The MacMillan Company, 1966.
[21] Weertman JJ. Dislocations moving uniformly on the interface between isotropic media of different elastic properties. J Mech Phys Solids 1963;11:197–204.
[22] Johnson KL. Adhesion and friction between a smooth elastic spherical asperity and a plane surface. Proc R Soc London 1997;A453:163–79.
[23] Hurtado J, Kim K-S. Scale effect in friction of single-asperity contacts. I. From concurrent slip to single-dislocation-assisted slip. II. Multiple-dislocations-cooperated slip. Proc R Soc London 1999;A455:3363–400.
[24] Polonsky IA, Keer LM. Scale effects of elastic–plastic behavior of microscopic asperity contact. ASME J Tribol 1996;118:335–40.
[25] Deshpande VS, Needleman A, Van der Giessen E. Discrete dislocation plasticity modeling of short cracks in single crystals. Acta Mater 2003;51:1–15.
[26] Friedel J. Dislocations. Oxford, New York: Pergamon Press, 1964.
[27] Ruan J, Bhushan B. Atomic-scale friction measurements using friction force microscopy: part I—general principles and new measurement technique. ASME J Tribol 1994;116:378–88.
[28] Onions RA, Archard JF. The contact of surfaces having a random structure. J Phys D Appl Phys 1973;6:289–304.
[29] Carpick RW, Agrait N, Ogletree DF, Salmeron M. Measurement of interfacial shear (friction) with an ultrahigh vacuum atomic force microscope. J Vac Sci Technol B 1996;14:1289–95.
[30] Schwarz UD, Zwörner O, Köster P, Wiesendanger R. Quantitative analysis of the frictional properties of solid materials at low loads. 1. Carbon compounds. Phys Rev B 1977;56:6987–96.
[31] Homola AW, Israelachvili JN, McGuiggan PM, Gee ML. Fundamental experimental studies in tribology: the transition from “interfacial” friction of undamaged molecularly smooth surfaces to normal friction with wear. Wear 1990;136:65–83.