Principles of atomic friction: from sticking atoms to superlubric sliding

BY HENDRIK HÖLSCHER1,2,3, ANDRÉ SCHIRMEISEN1,2 AND UDO D. SCHWARZ3,*

1Center for NanoTechnology (CeNTech), Heisenbergstrasse 11, 48149 Münster, Germany
2Physikalisches Institut, Westfälische Wilhelms-Universität Münster, Wilhelm-Klemm-Strasse 10, 48149 Münster, Germany
3Department of Mechanical Engineering, Yale University, PO Box 208284, New Haven, CT 06520-8284, USA

Tribology—the science of friction, wear and lubrication—is of great importance for all technical applications where moving bodies are in contact. Nonetheless, little progress has been made in finding an exact atomistic description of friction since Amontons proposed his empirical macroscopic laws over three centuries ago. The advent of new experimental tools such as the friction force microscope, however, enabled the investigation of frictional forces occurring at well-defined contacts down to the atomic scale. This research field has been established as nanotribology. In the present article, we review our current understanding of the principles of atomic-scale friction based on recent experiments using friction force microscopy.

Keywords: atomic-scale friction; friction force microscopy; nanotribology; Prandtl–Tomlinson model

1. Introduction

In basic physics textbooks, the frictional force $F_{\text{friction}}$ between two bodies in contact is typically described by the following phenomenological ‘law of friction’:

$$F_{\text{friction}} = \mu F_{\text{load}},$$

(1.1)

where $F_{\text{load}}$ represents the external loading force and $\mu$ the friction coefficient, which depends only on the actual combination of materials in contact. This equation is essentially the condensed form of observations first noted by Leonardo da Vinci (1452–1519). His findings had fallen into oblivion after his death, only to be independently rediscovered by Guillaume Amontons (1663–1705) who reported them to the French Académie Royale des Sciences on 19 December 1699. Hence, equation (1.1) is commonly referred to as Amontons’ law. It holds surprisingly well especially in the case of dry friction or Coulomb friction, which represents the case of

One contribution of 8 to a Theme Issue ‘Nanotribology, nanomechanics and applications to nanotechnology I’.

* Author for correspondence (udo.schwarz@yale.edu).

This journal is © 2007 The Royal Society
sliding without lubrication between the two interacting bodies. Unfortunately, it cannot be derived from first principles up to now. Even worse, our entire understanding of the fundamental mechanisms of friction on the atomic scale is still relatively limited since most macroscopic and microscopic frictional effects are usually dominated by the influence of wear, plastic deformation, lubrication, surface roughness and surface asperities. Macroscopic friction experiments are therefore difficult to analyse in terms of a universal theory. As this complexity of phenomena prevented the observation of pure wearless friction for many years, attempts to rigorously analyse the origins of friction on the molecular and atomic scale were largely unsuccessful.

In the last few decades, however, the field of nanotribology was established by introducing new experimental tools, which made the nanometre and atomic scales accessible to tribologists (Bhushan et al. 1995; Carpick & Salmeron 1997; Urbakh et al. 2004; Bhushan 2005). One of the basic ideas of nanotribology is that for a better understanding of friction in macroscopic systems, the frictional behaviour of a single-asperity contact should be investigated first. Figure 1 illustrates this concept. It is the hope that once the atomic-scale manifestations of friction at such a nanometre-sized single asperity have been clarified, macroscopic friction could be explained with the help of statistics, i.e. by the summation of the interactions of a large number of small individual contacts, which form the macroscopic roughness of the contact interface.

The first important step towards such experiments at the atomic level was taken when the surface force apparatus was equipped with special sample stages enabling the measurement of lateral forces between two molecularly smooth surfaces sliding against each other (Israelachvili & Tabor 1973; Briscoe & Evans 1982). The largest step forward concerning the reduction of the asperity size, however, was accomplished with the development of the friction force microscope (FFM; Mate et al. 1987). The FFM is derived from the atomic force microscope (AFM; Binnig et al. 1986) and enables the local measurement of lateral forces by moving a sharp tip, representing (approximately) a point contact, over a sample surface.

In this paper, we review recent contributions of friction force microscopy to the emerging field of nanotribology and their impact on our current theoretical understanding of the related underlying fundamental effects. First, we explain the working principle of friction force microscopy (§2). We then introduce in §3

---

Figure 1. Schematics illustrating the basic concept of nanotribology. (a) The apparent contact area between two bodies in contact observed on the macroscopic scale consists on close inspection of (b) many individual small asperities locally forming contacts with dimensions in the nanometre regime. (c) A nanometre-sized single-asperity contact, shown as an idealized sphere-on-flat geometry in atomic force microscopy (known as ‘Hertzian contact geometry’), can be realized in a friction force microscope and represents a model system for the isolated examination of one of the small contacts depicted in (b).
the Prandtl–Tomlinson model that has been used to analyse atomic-scale friction phenomena. Temperature- and velocity-dependent effects are presented in §4. Finally, we discuss the area dependence of friction in §5.

2. Atomic and friction force microscopy

The experimental set-up of an AFM is based on a simple idea. It detects forces acting between a sample surface and a sharp tip that is mounted on a soft leaf spring (the so-called cantilever). Such cantilevers are produced by standard microfabrication techniques, mostly with rectangular or V-like shapes, from silicon or silicon nitride. Spring constants for the bending of cantilevers typically used for AFM measurements are approximately $0.01$–$1$ N m$^{-1}$. The lateral spring constants for the torsion of the cantilever are in the range between $10$ and $100$ N m$^{-1}$.

A feedback system, which controls the vertical $z$-position of the tip on the sample surface, keeps the deflection of the cantilever (and thus the force between tip and sample) constant. Moving the tip relative to the sample in the $x$–$y$-plane of the surface by means of piezoelectric drives, the actual $z$-position of the tip is recorded as a function of the lateral $x$–$y$-position with (ideally) sub-angstrom precision. The obtained data represent a map of equal forces, usually interpreted as the surface topography. The conversion of the cantilever deflection into the normal force is performed by applying Hooke’s law. The data can be analysed and visualized through computer processing. A more detailed description of the method and the generally used instrumentation can be found in the books of Meyer et al. (2004) and Bhushan (2005).

The by far most widespread detection scheme for atomic force microscopy experiments is the laser beam deflection technique sketched in figure 2 (Marti et al. 1990; Meyer & Amer 1990). As explained in the figure legend, not only the deflection but also the torsion of the cantilever can be measured. AFMs that can record bending and torsion of the cantilever simultaneously are referred to as either FFMs or, alternatively, lateral force microscopes.

To be able to collect quantitative data, the correct calibration of the bending and the torsion of the cantilever is an important issue. Many parameters have to be known, such as the cantilever dimensions, the elastic modulus of the cantilever material, the tip length, the exact position on the cantilever backside where the laser beam is reflected and the sensitivity of the four-quadrant photo diode. Different procedures for the quantitative analysis of lateral force microscopy experiments have been introduced (Lüthi et al. 1995; Putman et al. 1995; Ogletree et al. 1996; Schwarz et al. 1996; Bilas et al. 2004).

As already mentioned in §1, macroscopic friction experiments are often difficult to analyse due to the complexity of the experimental system. Measurements using friction force microscopy, however, can be carried out in nearly every environment (ambient conditions, vacuum or liquids) with well-defined tips representing approximative point contacts. Since loading forces can be kept very low, effects like wear and plastic deformation can be prevented in most cases. These features make friction force microscopy a powerful tool for nanotribology.
3. Atomic-scale friction and the Prandtl–Tomlinson model

(a) One-dimensional considerations and the occurrence of stick–slip motion

Despite the fact that much of our current knowledge on the origin of friction has been acquired during the last two decades, many of the newly observed atomic-scale effects can actually be understood in the framework of a surprisingly old model. This so-called Prandtl–Tomlinson model (Prandtl 1928; Tomlinson 1929) is sometimes also referred to as the independent oscillator model (McClelland 1989; Helman et al. 1994). A schematic of the simple spring model is shown in figure 3. A point-like tip is elastically coupled to a main body M with a spring possessing a spring constant $c_x$ in the $x$-direction and interacts with the sample surface via a periodic potential $V_{\text{int}}(x_t)$, where $x_t$ reflects the actual position of the tip. During sliding, the body M is moved in the $x$-direction with a velocity $v_M$.

The equation of motion for the tip in the interaction potential can be written as

$$m_x \ddot{x}_t = c_x(x_M - x_t) - \frac{\partial V_{\text{int}}(x_t)}{\partial x_t} - \gamma_x \dot{x}_t, \quad (3.1)$$
where $m_x$ is the effective mass of the system and $x_M = v_M t$ is the equilibrium position of the spring at the time $t$. The last term on the r.h.s. is a simple velocity-dependent damping term with a damping constant $\gamma_x$. It describes the energy dissipation of the tip independent from the exact physical nature of the actual dissipation channel (i.e. phonons or electronic excitations). The solution of this differential equation is the path of the tip $x_t(t)$. The lateral force $F_x$ to move the tip in the $x$-direction can be calculated from $F_x = c_x(x_M - x_t)$, whereas the friction force $F_{\text{friction}}$ is defined as the averaged lateral force $\langle F_x \rangle$.

Within this model, the point-like tip represents the average of the actual tip–sample contact or single-asperity contact, which might include hundreds of atoms. Alternatively, the actual tip–sample contact can also be treated as a system consisting of many individual atomically sharp tips interacting through springs, which leads to more complex models. A prominent example is the model proposed by Frenkel & Kontorova (1938), which is based on chains of elastically coupled atoms exposed to a periodic potential. However, we will restrict ourselves in this first part of the review to the simple model introduced previously, which has been proven to successfully describe the tip motion for many materials. Effects caused by the area dependence of friction are discussed in §5.

If the sliding velocities are low, analytical solutions of equation (3.1) can be derived with the assumption that the tip is always in its stable equilibrium position, which corresponds to the actual minimum of its total energy

$$E_{\text{tot}} = \frac{1}{2} (x_t - x_M)^2 + V_{\text{int}}(x_t).$$

For better clarity, the corresponding energy is plotted in figure 4 for two positions $x_M$ and a sinusoidal interaction potential. In figure 4a the tip is trapped in a local energy minimum. An energy barrier of height $\Delta E$ prevents the tip from reaching the next energy minimum to the right. Since the tip base $M$ moves along the $x$-direction, the energy barrier diminishes with time until the local minimum
vanishes at a certain position \( x_M = x_{M, \text{jump}} \). As a result, the tip ‘jumps’ to the next local minimum where it is trapped again. This situation is sketched in figure 4b. The ‘jump height’, i.e. the maximum force due to spring tension just before the ‘slip’ event occurs (cf. figure 3c), is a direct function of the tip–sample interaction potential \( V_{\text{int}} \), the spring constant \( c_x \) and the lattice constant \( a \). Owing to the periodicity of the interaction potential, this ‘stick’ and ‘slip’ mechanism repeats as long as the support \( M \) is scanned over the sample surface. For finite temperatures, it has to be considered that the tip might sometimes jump thermally activated over a non-vanishing potential barrier \( \Delta E \) to reach the next local minimum. This case will be discussed later in §4.

For further analysis, we have to remember the two conditions for a local minimum \( \frac{\partial E_{\text{tot}}}{\partial x_t} = 0 \) and \( \frac{\partial^2 E_{\text{tot}}}{\partial x_t^2} > 0 \). Using these conditions, we see that the path of the tip \( x_t(x_M) \) can be determined from the solution of the equation

\[
c_x (x_M - x_t) = \frac{\partial V_{\text{int}}(x_t)}{\partial x_t}.
\]

For a stiff spring, there is only one solution for all \( x_M \), resulting in a continuous tip movement and vanishing friction, since the averaged lateral force \( \langle F_x \rangle \) is zero. However, the tip movement changes dramatically if the condition

\[
c_x < -\left[ \frac{\partial^2 V_{\text{int}}}{\partial x_t^2} \right]_{\text{min}}
\]

is fulfilled. Now, the tip moves discontinuously in a ‘stick–slip’-type motion over the sample surface, jumping from one potential minimum to another. This specific movement of the tip results in a saw-tooth-like function for the lateral force \( F_x \) (figure 3c). Since the averaged lateral force \( \langle F_x \rangle \) is non-zero in this case, a finite friction force \( F_{\text{friction}} \) is necessary to move the body \( M \) in the \( x \)-direction.

Figure 4. Potential diagram of the Prandtl–Tomlinson model at zero temperature. The indicated net potential represents the sum of a fixed sinusoidal interaction potential and a parabolic spring potential with its minimum at the position \( x_M \), which is moving with the velocity \( v \) from left to right (cf. equation (3.2)). In (a), where \( x_M \approx 0.66 \text{ nm} \), the tip is trapped in a local minimum and is separated from the next minimum to the right by an energy barrier of height \( \Delta E \). For (b), the tip base (and with it the minimum of the parabolic potential) has moved to the right (\( x_M \approx 0.8 \text{ nm} \)). As a consequence, the energy barrier has vanished, which causes the tip to ‘jump’ to the next local minimum. At non-zero temperature, however, there is a finite probability that even in (a) the tip might jump over the energy barrier due to thermal activation (dashed arrow).
In fact, it is exactly this behaviour that frequently occurs in atomic-scale friction force microscopy experiments. Observed first by Mate et al. (1987) for graphite(0001), experimentalists found atomic-scale stick–slip motion on many different materials such as mica (Fujisawa et al. 1993), NaF (Fujisawa et al. 1995b), TeS₂ (Kerssemakers & Hosson 1995), KBr (Lüthi et al. 1996), MoS₂ (Fujisawa et al. 1995a) and MoTe₂ (Hölscher et al. 1999).

A complete transition from stick–slip to pure sliding was reported by Socoliuc et al. (2004) for the NaCl(001) surface. Figure 5 shows a comparison between the experimental lateral force curves and the corresponding numerical simulation. Depending on the actual loading force of the tip, the measured lateral forces changed from a saw-tooth-like shape (figure 5a,b) to a continuous curve where forward and backward forces overlap (figure 5c). As a consequence, there is no hysteresis, and the frictional force reduces to ultra-low values as expected from the above discussed model. The graphs in figure 5d–f show numerical simulations where the simple sinusoidal potential

\[
V_{\text{int}} = -\frac{E_0}{2} \cos\left(\frac{2\pi}{a} x_t\right)
\]  

was used to describe the effective tip–sample interaction. The variation of the loading force can be inherently considered by the amplitude \(E_0\). Inserting this potential into condition equation (3.4), it is useful to define the parameter

\[
\eta = \frac{2\pi^2 E_0}{c_x a^2}.
\]  

Figure 5. Lateral forces acting on the tip during sliding in forward (solid lines) and backward (dotted lines) direction with different external loading forces. (a–c) Experimental results obtained on NaCl(001), where the external load was decreased from (a) 4.7 to (b) 3.3 to (c) 0.47 nN. (d–f) Numerical results where the parameter \(\eta\) was decreased from (d) 5 to (e) 3 to (f) 1 nN. (Reproduced with permission from Socoliuc et al. (2004).)
If $\eta > 1$, condition equation (3.4) is fulfilled and stick–slip occurs. This situation corresponds to figure 5d,e. Continuous sliding and no dissipation is observed for all $\eta \leq 1$. Figure 5f shows such a case, which is in good agreement with the experimental findings presented in figure 5c.

So far, we considered only the jump of the tip into the immediate neighbouring potential minimum. However, larger slips into the next but one minimum are also possible (Medyanik et al. 2006). Furthermore, there might be more than one minimum and maximum within one unit cell. In such a case, the stick–slip behaviour depends on the actual loading force since this parameter influences the height of the interaction potential as shown for NaF(001) (Fujisawa et al. 1996; Ishikawa et al. 2000) and $\beta$-MoT$_2$(001) (Hölscher et al. 1999).

Since stick–slip movement has been identified as the underlying mechanism for atomic-scale friction, it would be of pre-eminent technological importance if strategies could be found that allow one to prevent or suppress this effect. It has been known for several years that the actuation of the tip–sample contact by small oscillations reduces friction (Schimmel et al. 1995; Dinelli et al. 1997; Kerssemakers & Hosson 1998; Bureau et al. 2000; Heuberger et al. 2000; Riedo et al. 2003). A recent study by Socoliuc et al. (2006) demonstrates the possible efficiency of such an approach on the atomic scale elegantly. Figure 6 illustrates their experimental set-up, where a small oscillatory force was added to the normal force by an external electrostatic potential applied between cantilever and sample. If that oscillatory force was modulated with a frequency equivalent to one of the vertical resonance frequencies of the tip–sample contact, the stick–slip motion of the cantilever and thus friction itself could be reproducibly switched on or off.

Figure 6. Friction can be reduced by a small oscillation of the normal force. Socoliuc et al. (2006) demonstrated that stick–slip behaviour can disappear if a small modulation with an appropriate frequency is added to the external loading force of the cantilever. (Reproduced with permission from Carpick (2006).)
(b) Extending the Prandtl–Tomlinson model to two dimensions

The original Prandtl–Tomlinson model considered only the one-dimensional movement of a tip along the x-direction. As real surfaces are two dimensional, any somehow realistic model has to include at least the x–y-periodicity of the crystalline lattice. A two-dimensional extension of the Prandtl–Tomlinson model has been introduced by Gyalog et al. (1995). In the following, this and similar models have been successfully applied to simulate the two-dimensional stick–slip movement of the tip on various crystalline surfaces, such as NbSe$_2$ (Kerssemakers & Hosson 1995), graphite (Sasaki et al. 1996; Toussaint et al. 1998; Hölscher et al. 1998), KBr (Litthi et al. 1996), NaF (Hölscher et al. 1996) and MoS$_2$ (Hölscher et al. 1997). In the following analysis, we will demonstrate with the example of graphite(0001) that the possibility of two-dimensional motion has significant effects on FFM images.

Assuming a two-dimensional interaction potential $V_{\text{int}}(x_t, y_t)$ leads to the following two coupled second-order differential equations:

$$m_x \ddot{x}_t = c_x(x_M t - x_t) - \frac{\partial V_{\text{int}}(x_t, y_t)}{\partial x_t} - \gamma_x \dot{x}_t, \quad (3.7a)$$

$$m_y \ddot{y}_t = c_y(y_M - y_t) - \frac{\partial V_{\text{int}}(x_t, y_t)}{\partial y_t} - \gamma_y \dot{y}_t, \quad (3.7b)$$

which describe the movement of the tip in the tip–sample potential of an idealized AFM (Hölscher et al. 1997). The simulation of experimental data is performed by moving the support M with constant velocity $v_M$ in ‘forward’ direction (i.e. from left to right) as well as in ‘backward’ direction (i.e. from right to left) continuously along a certain line in the x-direction while $y_M$ is held constant. Calculating the position of the tip $(x_t, y_t)$ from equations (3.7a) and (3.7b), the lateral forces $F_x = c_x(x_M - x_t)$ and $F_y = c_y(y_M - y_t)$ can be determined as a function of the support position $(x_M, y_M)$.

The interaction potential $V_{\text{int}}(x_t, y_t)$ depends on the properties of tip and sample and is generally unknown. However, in many practical cases it is sufficient to consider only the periodicity and symmetry of the sample surface. To represent the hexagonal structure of graphite(0001), we chose

$$V_{\text{HOPG}}(x_t, y_t) = -V_0 \left[ 2 \cos \left( \frac{2\pi}{a} x_t \right) \cos \left( \frac{2\pi}{a\sqrt{3}} y_t \right) + \cos \left( \frac{4\pi}{a\sqrt{3}} y_t \right) \right], \quad (3.8)$$

where $a=2.46$ Å is the lattice constant. Using this potential, numerical solutions of equations (3.7a) and (3.7b) can be computed with suitable parameters (Hölscher et al. 1998).

Figure 7 contrasts an experimentally measured force map (figure 7a) with such a simulation (figure 7b). The calculated force map reproduces even small features at the left side of the images—effects caused by the beginning tip movement. Both force maps show a structure with a distance of 2.46 Å between the individual maxima. This structure can be explained by the typical stick–slip movement of the tip on the graphite surface. The detailed tip movement is analysed in figure 7c, where three arbitrarily chosen paths of the tip in the interaction potential are plotted for different positions of $y_M$ ‘time-resolved’ by
points separated by equal time intervals. It can be clearly seen that the tip does not scan the sample surface continuously. Only the minima of the assumed tip–sample interaction potential—the so-called ‘hollow sites’—are in contact with the tip for a significant time; high-energy locations in the potential landscape are skipped. Note in particular that during the ‘slips’, the tip always jumps over a saddle point between two potential maxima marking the locations of carbon atoms. This results in a two-dimensional ‘zigzag’ motion, where the tip ‘sneaks around’ the atoms in a fashion much like a slalom skier around the pylons.

4. Thermal activation and the velocity dependence of friction

As we have seen above, the Prandtl–Tomlinson model and its extensions are quite successful in describing the qualitative overall behaviour of atoms at nanometre-scale dry, single-asperity contacts in relative motion. We argued that the tip moves in a stick–slip like fashion over the sample surface and that this specific movement is responsible for the observed friction and energy dissipation. This theory is based on the idea that the tip slips from one energy minimum to the next if the energy barrier $\Delta E$ between the current minimum and the next one vanishes owing to the changes induced by the cantilever movement on the total energy (equation (3.2)). Since these jumps are typically much faster than the scan velocities applied in friction force microscopy, the frictional forces can be assumed to be nearly constant for scan velocities well below the slip velocities (Zwo¨rner et al. 1998).

However, this simple picture does not hold if we consider finite temperatures. As already indicated in figure 4, the tip might actually jump over a non-vanishing energy barrier $\Delta E$ due to thermal activation, leading to reduced frictional forces. Even though this effect has already been discussed in the original publication by Prandtl (1928), it is still a topic of current research.
The velocity dependence on atomically flat surfaces has been measured by FFMs on mica (Liu et al. 1994; Koinkar & Bhushan 1996; Riedo et al. 2003), Si(100) (Koinkar & Bhushan 1996), Cu(111) (Bennewitz et al. 2000), NaCl(100) (Gnecco et al. 2000), silicon oxide (Schirmeisen et al. 2006) and graphite (Krylov et al. 2005; Schirmeisen et al. 2005; Evstigneev et al. 2006). Most authors report a logarithmic increase of the frictional force with velocity, which might be explained by a phenomenological ‘thermally activated Eyring model’ (Bouhacina et al. 1997).

Nonetheless, recent theoretical analyses of these phenomena are often based on a thermally activated Prandtl–Tomlinson model. A good starting point is the following ‘master equation’ (Gnecco et al. 2000):

\[
\frac{dp(t)}{dt} = -f_0 \exp \left(-\frac{\Delta E(t)}{k_B T} \right) p(t),
\]

(4.1)

where \(\Delta E(t)\) reflects the energy barrier at the time \(t\); \(f_0\) is the resonance frequency of the tip in its actual minimum; and the probability that the tip does not jump is denoted as \(p(t)\).

For the analysis of experimental data, we have to change the variable \(t\) to the actual lateral force \(F_x\) (Gnecco et al. 2001). After this transformation, we get

\[
\frac{dp(F_x)}{dF_x} = -\frac{f_0}{c_x v_M} \exp \left(-\frac{\Delta E(F_x)}{k_B T} \right) p(F_x).
\]

(4.2)

The maximum of the probability for jumping can be obtained from the condition \(d^2 p/dF_x^2\) and leads to

\[
\frac{\partial \Delta E}{\partial F_x} = -\frac{k_B T f_0}{c_x v_M} \exp \left(-\frac{\Delta E(t)}{k_B T} \right).
\]

(4.3)

This equation can be solved for an analytical description of \(\Delta E\). To do so, Gnecco et al. (2000, 2001) assumed a linear dependence of the energy barrier on \(F_x\). However, most jumps occur very close to the critical position, as pointed out by Sang et al. (2001). In this case, the energy barrier is relatively small and more adequately described by a function of the type \(\Delta E = (1/\beta)(F_{\text{max}} - F_x)^{3/2}\), where \(F_{\text{max}}\) is the maximum possible lateral force occurring at zero temperature and \(\beta\) a constant. Introducing this relationship into equation (4.3), we get an implicit function for the frictional force (Riedo et al. 2003)

\[
\ln \left(\frac{v_0}{v_M}\right) - \frac{1}{2} \ln \left(1 - \frac{F_x}{F_{\text{max}}} \right) = \frac{1}{\beta k_B T} (F_{\text{max}} - F_x)^{3/2}.
\]

(4.4)

Here, the parameter \(v_0\) has been defined as

\[
v_0 = \frac{2f_0 \beta k_B T}{3 c_x \sqrt{F_{\text{max}}}};
\]

(4.5)

it is usually referred to as ‘critical velocity’ (Gyalog et al. 2006).

The nature of \(v_0\) as a critical velocity is best explored by approximating equation (4.4) for a ‘low velocity’ and a ‘high velocity’ regime relative to \(v_0\). As shown by Gyalog et al. (2006), this approach results in

\[
F_{\text{low}}(v_M) = F_{\text{max}} - \left(\frac{\beta k_B T \ln \left(\frac{v_0}{v_M}\right)}{2\beta k_B T} \right)^{2/3}
\]

(4.6)
for the low velocity limit and in
\[
F_{\text{high}}(v_M) = F_{\text{max}} \left(1 - \left(\frac{v_0}{v_M}\right)^2\right)
\]
for the high velocity limit. Figure 8a compares the resulting force curves for the exact and approximated solutions using the parameters \(T=300\) K, \(c_x=1.0\) N m\(^{-1}\), \(f_0=25\) kHz, \(\beta=10^6\) N\(^{3/2}\) J\(^{-1}\) and \(F_{\text{max}}=1.5\) nN, leading to a critical velocity of \(v_0=1.78 \times 10^{-6}\) m s\(^{-1}\). The solid line represents the exact solution of the implicit equation (4.4). Note that the frictional force increases logarithmically until it reaches a plateau for velocities higher than the critical velocity \(v_0\). The logarithmic increase can be described by the approximation equation (4.6), while the dependence near the plateau is well reproduced by equation (4.7). Therefore, \(v_0\) represents the velocity where the logarithmic dependence transits towards the plateau value. The vertical dashed line in figure 8a labels the critical velocity \(v_0\) and marks this border. Figure 8b displays experimental force versus velocity curves for different loads on mica. Both the logarithmic increase and the plateau for higher velocities were found. The data could be also fitted with equation (4.4), as shown by the solid and dotted lines.

5. Area dependence of friction: structural lubricity versus Amontons’ law

One of the most debated subjects in nanotribology in recent years concerned the question of how the above principles, valid for virtual point contacts, can be generalized to finite contact sizes or, in other words, how the frictional
force $F_{\text{friction}}$ experienced at a finite, atomically flat interface of nanoscopic dimensions scales with the actual contact area $A_{\text{contact}}$. Macroscopically, Amontons’ law (equation (1.1)) applies and friction is independent of the dimensions of the interface, as seen in §1. If, however, we perform the transition from the apparent macroscopic contact area to the true dimensions of actual nanometre-sized single-asperity contacts that support the weight of macroscopic sliders (cf. figure 1), it is reasonable to assume that the situation will change. This section reviews the current theories on this subject based on atomic-scale principles.

Let us start by recognizing that there are two different spring constants that determine the overall behaviour of the contact. First, we saw above that the effective spring constant of the contact in the $x$-direction $c_x$ (i.e. the overall spring constant between the fixed sample base and the slider base that is in motion) is quite soft, which leads to the observed stick–slip motion of the atoms at the interface. This relative softness of the overall spring constant $c_x$, however, is an effect arising from summing up thousands of individual, comparatively hard springs, each represented by a single interatomic bond. As a consequence, neighbouring atoms at the interface can be regarded as separated by a first approximation a fixed distance $a$ (the lattice constant). This finding allowed us to formally reduce the motion of a large ensemble of atoms at an atomically smooth interface to the motion of a virtual point contact located at the position $x_t$ that is moving in an effective interaction potential $V_{\text{int}}(x_t)$ for fixed contact areas and loads, as pursued above. Or, with other words, we successfully mapped the complex multibody problem onto an easy-to-solve effective single-body problem. To include effects of variations in contact area into this picture, we have to understand how the effective potential $V_{\text{int}}$ depends on the contact area.

A frequently used illustration for this case arises from the purely geometrical argument that friction is caused by the interlocking of surface asperities. Bé�idor modelled such surface asperities as spherical asperities arranged to form commensurate crystalline walls as early as 1737 (Dowson 1979). His approach was later revived by Homola et al. (1989, 1990) who interpreted the spherical asperities as atoms and called it the ‘cobblestone model’. However, asperities on real surfaces do not match as well as envisioned in these models. On average, for every asperity or atom going up a ramp, there is another going down. One concludes that the mean friction between rigid surfaces vanishes unless they happen to have the same periodicity and alignment. Detailed calculations show that elastic deformations are generally too small to alter this conclusion (Hirano & Shinjo 1990, 1993; Sørensen et al. 1996; Müser & Robbins 2000).

Figure 9 illustrates the situation from a slightly different viewpoint. It shows the surface atoms of a substrate and the surface potential they cause. If only one atom was placed in this surface potential (figure 9a), it drops into a deep minimum and a huge energy barrier has to be overcome to displace it by one lattice constant $a$. However, if we place two atoms in a fixed structural relation with each other on the surface, featuring a lattice constant $b \neq a$, the energy barrier that has to be overcome by each individual atom to move the entire two-atom cluster by $a$ has shrunk considerably (figure 9b). This principle continues for increasing contact sizes (figure 9c,d) and ultimately results in a vanishing energy barrier and therefore ultra-low friction as long as substrate and slider have incommensurate lattices or feature disordered (amorphous) surfaces. While
it is difficult to give an analytic expression describing the case of incommensurate lattices, one can show that the generalization of these principles in terms of simple analytical models leads for the latter case, i.e. the static friction $F_{\text{static}}$ between dry, amorphous and flat surfaces, to

$$F_{\text{static}} \propto \frac{F_{\text{load}}}{\sqrt{A_{\text{contact}}}}, \quad (5.1)$$

due to the averaging effect of the law of large numbers. This effect was originally denoted as superlubricity (Hirano & Shinjo 1990; Shinjo & Hirano 1993); however, as it is a purely structural effect, it has been suggested to denote it as structural lubricity by Müser (2004). In this way, it can be distinguished from other effects reducing friction such as the stick–slip prevention techniques reviewed in §3.

Even though many aspects of structural lubricity are still under debate, such as the exact circumstances under which a superlubric state can actually be established, a growing number of experimental studies seem to corroborate its existence (Martin et al. 1993; Hirano et al. 1997; Crossley et al. 1999; Dienwiebel et al. 2004). Results from Dienwiebel et al. (2004) are presented in figure 10, which reflect the friction measured between an FFM tip and a graphite sample as a function of the angle between tip movement and graphite lattice. In order to realize this measurement, the authors used a home-built FFM with specifically designed sensors that allow the simultaneous and independent detection of the lateral forces acting on the tip in both $x$- and $y$-directions (Zijlstra et al. 2000). A significant increase of the friction for angles of approximately $0^\circ$ and $60^\circ$ is observed, but ultra-low friction is found for all other angles.

In order to explain their results, Dienwiebel et al. (2004) assumed that the tip has picked up a small graphite flake during scanning. This flake would then be commensurate with the underlying lattice for sliding angles of $0^\circ$ and $60^\circ$, but incommensurate for all angles in between. They estimated the diameter of the flake that they picked up to approximately 7–12 lattice spacings, which they conclude from the width of the peaks located at $0^\circ$ and $60^\circ$. 

**Figure 9.** Illustrating the effect of incommensurability on the average barrier between potential minima: while the number of atoms that have to overcome a barrier increases from one to four for (a–d), respectively, the height of the individual barriers shrinks significantly. For increasingly larger contacts, the effective overall barrier height will approach zero even though a large number of atoms contribute to the frictional resistance.
Equation (5.1) might even have much bigger consequences than the so far small number of publications referring explicitly to structural lubricity effects suggests. In the past, it has frequently been found that the friction between an FFM tip and a sample surface depends nonlinearly on load (Carpick et al. 1996, 1997; Meyer et al. 1996; Lantz et al. 1997; Schwarz et al. 1997a,b; Enachescu et al. 1998). This has been explained by a constant shear stress \( \tau = F_{\text{friction}}/A_{\text{contact}} \), while \( A_{\text{contact}} \) was represented by one of the popular sphere-on-flat contact mechanical models such as the Johnson–Kendall–Roberts (JKR) model, Maugis’ approximation to the Derjaguin–Muller–Toporov model (DMT-M) or the Maugis–Dugdale (MD) model. All mentioned contact models are derivatives of the basic Hertzian contact model

\[
A_{\text{contact}} = \pi \left( \frac{R}{K} \right)^{2/3} F_{\text{load}}^{2/3},
\]

where \( R \) reflects the tip radius and \( K \) the ‘effective elastic modulus’, which combines the elastic properties of tip and sample. As the Hertz model does not consider adhesion, the JKR, DMT-M and MD models essentially differ in the way that adhesion is included (see Schwarz (2003) for details and the exact definition of \( K \)). In fact, despite the usually small dimensions of the FFM tips, molecular dynamics simulations showed that the atomic discreteness within the bulk of solids does not have a significant effect (Luan & Robbins 2005). The (atomic) surface roughness of tip and sample, however, may lead to significant deviations from continuum theory. An underestimation of contact areas and an overestimation of friction may be the consequence, which nevertheless heavily depends on the exact atomic structure of the tip apex. Despite these reservations, additional simulations that included friction reproduced surprisingly well the experimentally observed dependencies for the more realistic tip apex models considered (Luan & Robbins 2006). In particular, the simulations seem to corroborate the above finding of a largely pressure-independent shear stress, even though our earlier considerations showed that this is not expected theoretically.

![Figure 10. Average friction force versus rotation angle of a graphite sample around an axis normal to the sample surface. Two narrow peaks of high friction were observed at 0° and 61°, respectively. Between these peaks a wide angular range with ultra-low friction close to the detection limit of the instrument was found. The first peak has a maximum friction force of 306±40 pN and the second peak has a maximum of 203±20 pN. The solid curve shows results from a model calculation for a graphite flake of 96 atoms. (Reproduced with permission from Dienwiebel et al. (2004).)
Wenning & Müser (2001) explained this apparent contradiction for adhesion-free Hertz-type contacts by introducing equation (5.2) into equation (5.1), leading to $F_{\text{friction}} \propto A_{\text{contact}}$ in accordance with the experiment. For contacts subject to adhesion, we can generalize this argument by defining an effective Hertzian load $F_{\text{Hertz}}$ as the force that has to be introduced into equation (5.2) to obtain the actual contact area. As argued elsewhere, $F_{\text{Hertz}}$ can be regarded as reflecting the effectively acting loading force including adhesional contributions (Schwarz 2003). Thus, $F_{\text{Hertz}}$ replaces $F_{\text{load}}$ not only in equation (5.2) but also in equation (5.1), again leading to $F_{\text{friction}} \propto A_{\text{contact}}$ despite a potentially complex influence of adhesion of contact area. Ultimately, our line of arguments implies the important conclusion that small contacts (some nm$^2$) are frequently subject to structural lubricity, even though this might be hidden in a nonlinear frictional behaviour. Also note that within this approach, both the apparent proportionality of contact area and friction and the apparent independence of the shear stress from the contact pressure are a product of the specific contact area–load dependence of spherical contacts and not a general finding.

Despite the above discussion, a structure-induced superlubric state still seems to be more exotic than widespread. Not only have FFM experiments been reported where the friction–load dependence appears to be linear, but also macroscopically measured friction coefficients are always substantial and do not vanish with increasing contact area. In fact, all the existing experimental demonstrations of structural lubricity have either been achieved using scanning probe microscopes, which feature very small contact areas as discussed above, or they are based on the sliding of basal planes of layered crystals relative to each other (Martin et al. 1993), an effect that is widely held responsible for the excellent lubrication properties of solid lubricants such as graphite or molybdenum disulphide. Realistically sized nanocontacts, which can have sizes anywhere between several tens and several hundred thousands of nm$^2$, have not yet been found to show structural lubricity.

So why not? According to a suggestion by He et al. (1999), the problem is that realistic contacts almost always feature adsorbed, but mobile molecules that are trapped between the sliding surfaces. Examples for such molecules are small hydrocarbon or water molecules that adhere on any surface (this effect can be only prevented by ultra-high vacuum conditions). Often referred to as ‘dirt particles’, these molecules prevent a direct interaction of the surface potentials of the sliding interfaces by acting as spacers. Since their mobility allows them to simultaneously lock at surface potential minima for both sliders, an area-independent friction coefficient is obtained for any surface geometry (figure 11). Owing to the ubiquity of contamination, this effect even represents the assumed reason for the apparent universality of Amontons’ law (Müser et al. 2001).

Unfortunately, the predictions by He et al. (1999) and Müser et al. (2001) are difficult to validate. To date, only two studies based on realistically sized nanocontacts have been published (Sheehan & Lieber 1996; Ritter et al. 2005). The reason for this apparent deficiency is that established experimental procedures are severely limited due to a size gap between the small contact areas of scanning probe microscopes and the contact areas offered by the surface

---

$F_{\text{Hertz}}$ may be a nonlinear function of the external load $F_{\text{load}}$, the tip radius $R$, the work of adhesion $\gamma$ and the ‘critical force’ $F_{\text{critical}}$ needed to separate tip and sample, which is often also referred to as ‘pull-off force’ (Schwarz 2003).

---

Phil. Trans. R. Soc. A (2008)
Figure 11. Illustrating the effect of contamination on the resulting friction coefficient. (a) Two incommensurable, atomically flat surfaces are in relative motion. Owing to the effects discussed in the text, the barrier between stable potential minima decreases with increasing contact size (cf. figure 9). (b) Same contact as in (a), but this time, mobile contamination molecules (‘dirt’) are present in the interface, effectively preventing a direct interaction of the surface potentials of the two sliding bodies. As the mobile bodies can always lock at suitable potential minima, a friction coefficient independent of the contact area is obtained.

Figure 12. (a–d) Illustration of the manipulation procedure for antimony particles on HOPG (image sizes: 1×1 μm², the height of the particle labelled with ‘a’ is 26 nm). (a) Overview of the particle of interest (particle ‘a’) and the surrounding area. A white and black arrow indicate the path of the subsequent tip motion and the resulting dislocation of the particle, respectively. (b) Topography after the manipulation. Comparison with (a) shows that the particle ‘a’ experienced a lateral translation and an in-plane rotation. For the next manipulation step, another contact point between the particle and the tip was selected, visualized again by a white arrow. (c) Result of the second manipulation step, revealing again a translation and an in-plane rotation. (d) Final result after the third manipulation step. (e) Plot of the minimum values of power dissipation needed for translation of differently sized antimony nanoparticles on graphite (filled triangles) and MoS₂ (open circles), respectively. The threshold values for both substrates are in the same range and scale linearly with the contact area of the translated particles. Note that according to a recent theoretical study (Aruliah et al. 2005), the threshold values scale linearly with static friction. The straight lines represent linear fits of the measured data.
force apparatus, which are of the order of some ten thousands of \( \mu \text{m}^2 \) (Homola et al. 1989, 1990). The two studies quoted above present experiments where the frictional resistance of nanoparticles is measured while they are pushed by the tip of an FFM, which has rarely been applied for quantitative friction experiments so far. In both cases, friction grows linearly with contact area, as expected (see figure 12 for an example). Finally, additional support for our above-developed theories comes from the surface force apparatus studies mentioned earlier (Homola et al. 1989, 1990). In these studies, it was found that friction grows linearly with contact area (and nonlinear with load) as long as the interface is contamination free (note that the contact in a surface force apparatus is analogous to a Hertzian-style sphere-on-flat contact). However, as soon as contamination enters the interface, a linear friction–load dependence is observed, corroborating the above predictions.

6. Conclusions

In this article, we have reviewed various aspects of the most popular theories believed to deliver useful descriptions of the atomic-scale frictional behaviour of nanoscale contacts in relative motion. In particular, friction force microscopy has been demonstrated to be a powerful tool in our quest to gain direct insight into the fundamental processes of friction. Atomic-scale stick–slip friction is one of the most prominent manifestations of nanoscale friction, which can rather successfully be described in the framework of the Prandtl–Tomlinson model. By including effects due to a finite temperature into the theory, a logarithmic velocity dependence is predicted, which has been confirmed experimentally.

Despite its successes, important limitations of the Prandtl–Tomlinson model concern the influence of more complex, realistic (i.e. extended) interfaces, including the existence of the real-life omnipresent surface adsorbates. In fact, clean crystalline surfaces are expected to exhibit very low friction due to the effect of structural lubricity, unless in the rare case of perfect commensurability. These recent advances in our understanding of friction nurture our hopes that these efforts will lead to new tailor-made materials that will eventually allow us to fully control the frictional properties of moving surfaces in mechanical contact.

H.H. acknowledges financial support from the German Federal Ministry of Education and Research (BMBF; grant no. 03N8704). U.D.S. thanks the donors of the American Chemical Society Petroleum Research Fund (grant ACS/PRF 42259-AC5) for partial support of this research.

References


Principles of atomic friction


Fujisawa, S., Kishi, E., Sugawara, Y. & Morita, S. 1995b Two-dimensionally discrete friction on the NaF(100) surface with the lattice periodicity. Nanotechnology 6, 8–11. (doi:10.1088/0957-4484/6/1/002)


Phil. Trans. R. Soc. A (2008)


*Phil. Trans. R. Soc. A* (2008)


