The physics of atomic-scale friction: Basic considerations and open questions

Sergey Yu. Krylov*1,2 and Joost W. M. Frenken**,1

1 Huygens-Kamerlingh Onnes Laboratory, Leiden University, P. O. Box 9504, 2300 RA Leiden, The Netherlands
2 Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, 119071 Moscow, Russia

Received 28 August 2013, revised 25 February 2014, accepted 28 February 2014
Published online 3 April 2014

Keywords energy dissipation, friction force microscopy, nanotribology, superlubricity, thermolubricity

*e-mail krylov@redline.ru, Phone: +79104363421
**e-mail frenken@ARCNL.nl, Phone: +31-20-7547100

***This article is also included as a chapter in Volume 4 "Solid–Solid Interfaces and Thin Films" of the book series "Surface and Interface Science", edited by K. Wandelt (Wiley-VCH, Weinheim, 2014), ISBN 978-3-527-41157-3.

The field of friction is usually associated with its obvious practical importance. This tends to emphasize the engineering aspects of friction and thereby hides the fact that there is a wealth of interesting physics involved, part of which cannot be regarded as fully understood at present. New techniques, such as friction force microscopy, have started to provide access to the phenomenon of friction on the atomic scale. This has given a strong impulse to the field of tribology, pushing it significantly beyond the engineering level and into the regime of the fundamental aspects of frictional energy dissipation. This article reflects the authors’ personal view on matters of interest in the field of atomic-scale friction. Rather than to review important contributions in this field, we have chosen to summarize what has been learned and identify phenomena that may seem familiar to tribologists but actually should be regarded as non-trivial from a physical point of view.

1 Introduction Many physicists think, as did the present authors before having engaged in the field of nanotribology, that friction is mainly an engineering problem. After all, for the construction of a good ball bearing, the help of a physicist seems not to be essential. However, there are several simple observations that should serve as a warning against this naive point of view. So far, nobody has succeeded to predict the friction coefficient for a given system. Similarly, nobody has been able to come up with a more or less general way to reduce friction coefficients, not even by only 1%. This is in spite of the important fact that such a modest reduction in friction and the accompanying reduction in wear would annually save the industrialized countries billions of euros. Even though a considerable financial investment has been made in more than four decades of intensive development of the science of tribology, our real achievements in this field remain not more than modest. Usually, one is inclined to relate this to the complexity of “real” sliding configurations, with complex surfaces, with roughness on a wide range of length scales, with a variety of chemical effects and with the effects of lubricating media. Alternatively, one can recognize our lack of predictive power as a sign that at a fundamental level there is still much to be discovered about the origins of friction. In this article, we argue that the latter is indeed the case.

Our main aim here is to highlight the physics of friction, in particular in the relatively young and rapidly developing field of nanotribology, with special attention given to several of the non-trivial aspects of the phenomenon. A number of books and review papers have already summarized the state of the art in tribology with emphasis on the development of experimental techniques and theoretical models, see, e.g., [1–12]. However, in many cases the physical aspects have remained somewhat underilluminated, because of the dominant role of the technical or computational methodology. We concentrate mostly on this – physical – aspect, thus leaving out much of the work that has been performed in this area. Any selection of this type is subjective by definition. As a consequence, this article cannot pretend to review the state of the art, but mainly presents a personal view of the authors
on friction as a physical phenomenon. We believe that friction should be investigated not only because of its practical relevance, but also in view of its scientific appeal.

We concentrate on the seemingly most simple and certainly most fundamental aspect of the phenomenon – dry friction on the atomic scale, hoping that this will capture the essence of the behavior of the individual asperities that, together, establish the contact between sliding solid bodies, see Fig. 1a. Thus, we do not consider the practically important cases of friction on micro- and macro-scales, as well as lubrication and friction with wear.

The investigations of atomic scale friction started in 1987 [13] with the observation of atomic stick–slip motion, using a modified version of the new-born atomic force microscope (AFM) technique. This pioneering work, although in itself not aimed to investigate the origin of friction, stimulated the development of friction force microscopy (FFM) and thus manifested the birth of the new science of nanotribology.

Our first surprise will be that friction is often accompanied by stick–slip motion. What is the reason for this discrete, not-continuous mode of sliding? Why is stick–slip universal, in the sense that it takes place both on the macroscale (think of the motion of a bow over a violin string or the creaking

---

**Sergey Krylov** is a leading scientist of the A.N. Frumkin Institute of Physical Chemistry and Electrochemistry, at the Russian Academy of Sciences, and a Professor in Physics and Mathematics at Moscow State University of Land Use Planning. For many years he has been spending time as a visiting scientist at Leiden University. The Netherlands, in the last decade in close collaboration with the Interface Physics Group. Central to the research of Sergey Krylov is the fundamental understanding of nonequilibrium phenomena at surfaces and interfaces. Topics of interest include theory of particle-surface interactions, molecular transport in fine channels, surface diffusion, crystal growth, nanotribology, and frictional energy dissipation.

**Joost Frenken** is the Director of the Advanced Research Center for Nano-Lithography in Amsterdam and a Professor in Experimental Physics at Leiden University. Central to the research of Joost Frenken is the fundamental understanding of the dynamic aspects of surfaces and interfaces and their role in relevant, natural or industrial, processes under practical conditions. Topics of interest include surface diffusion, crystal growth (e.g. graphene), surface phase transitions, catalysis, and friction. For tailormade measurements in each of these areas, Frenken’s research group has developed a variety of special-purpose scanning probe microscopes as well as dedicated instrumentation for surface X-ray diffraction.
sound of a door) and on the atomic scale? A formal explanation of atomic stick–slip seems to be the natural outcome of a simple, mechanistic model proposed by Prandtl [14] and Tomlinson [15] as early as the 1920s, the basic ingredients of which are represented in Fig. 1b. However, the physics behind this phenomenon is far from simple. In particular, the appearance of regular, atomic-scale stick–slip, as observed in FFM experiments, requires the inherent dissipation rate to have a value in a specific range (close to "critical damping"). This is again surprising, since the dissipation rate depends not on the contact details, as one would certainly expect, but—perhaps counterintuitively—also on the stiffness and mass of the measuring system.

The alternative for stick–slip motion is the regime of continuous low-dissipative sliding, which is a general prediction of the Prandtl–Tomlinson model for those cases in which the contact potential corrugation is low enough and/or the stiffness of the measuring system is high enough. This regime is so unusual, that the so-called "superlubricity" transition from stick–slip to continuous sliding has been demonstrated in experiments only in the previous decade.

A direct prediction by the Prandtl–Tomlinson model, which is confirmed in FFM experiments, is that the average friction force (in the stick–slip regime) increases linearly with the ratio of the contact potential corrugation to the measuring system stiffness. It comes as no surprise that friction depends on the contact conditions, but the fact that it equally depends on properties of the driving system (its stiffness) is counterintuitive. This "dependence of friction on the driver" reflects a complex scenario of energy dissipation. It includes not only how the energy stored in the system is being transferred to heat but also how the driver first invests energy into the system. A stiff driver wins over a softer one, i.e., it experiences less friction.

Some deeper questions are whether and how the atomic scale friction depends on scanning velocity and system temperature. For both dependencies, the actual mechanisms turn out to be very different from initial, naïve expectations. Instead of a linear increase of friction with velocity (like for viscous friction in liquids), the simplest mechanistic analysis of dry friction, in the context of the Prandtl–Tomlinson model, leads to results that are independent of velocity (and temperature). In experiments, the velocity dependence in the stick–slip regime is found to be close to logarithmic. The decisive element turns out to be the thermal motion of the slider, which causes the slip events in the stick–slip sequence to start somewhat prior to the points of mechanical instability. The effect of such precritical slips on the mean friction force depends on velocity—this is how the logarithmic velocity dependence comes in.

Thermally activated motion is also at the origin of the temperature dependence of atomic-scale friction, but here the effect can be really dramatic. Under certain conditions, the thermally induced suppression of friction can even be orders of magnitude, an effect called "thermolubricity." Also, different behavior is possible, even featuring a highly counterintuitive increase of friction with temperature.

A natural question in this context is whether thermal motion can be of serious importance for macroscopically large sliding objects. Interestingly, the answer is "yes," which finds its origin in the simple fact that the system can not be modeled properly as a single, mechanical spring. The apex of an FFM tip is flexible, as is measured readily in experiments. The same should be true for the apices of nano-asperities at the surfaces of macroscopic, contacting bodies. Not only does this bring an additional spring into the problem, but, more importantly, it causes dramatic effects in the system dynamics as it can make the effective mass of the contact extremely low. Under certain conditions, the contact can be partially or even completely delocalized by the rapid thermally activated jumps of the tip apex between the surface potential wells. As a consequence, besides the familiar stick–slip and continuous low-dissipative sliding, there can be a variety of other, physically different regimes of motion and energy dissipation.

We will consider several other interesting FFM results, such as "multiple" slips (over more than one lattice spacing) and unusually structured slips. These reflect the possibility that the tip can move from one lattice position to another not only via a direct slip, but also mediated by several jumps back and forth. Again the background is to be found in the rapid dynamics of the nanocontact.

We conclude this initial survey by stressing that many questions remain far from being answered. For example, our present understanding of dry friction is at a semi-phenomenological level; the microscopic mechanisms of dissipation remain hidden. Apart from a formal reference to possible phononic or electronic mechanisms, we know little about the details of how and where the energy and momentum losses become irretrievable and how and where the transition of mechanical energy to heat takes place. In fact, one usually even does not distinguish between these two, i.e., irretrievable loss and transition to heat. A second example is that friction is obviously a nonequilibrium statistical phenomenon. Usually, theoretical calculations or computer simulations in this field are based on Langevin-type equations, even though a justification for the inherent approximations is lacking, in particular the neglect of memory effects and noise correlations. As far as we know, this issue has not been addressed properly from the perspective of the thermodynamics of nonequilibrium processes. Clearly, our understanding of friction as a physical phenomenon is still poor and will remain so until these general issues are clarified.

2 Fundamentals of tribology and nanotribology

2.1 Da Vinci’s laws and their failure at the nanoscale

The friction laws one is taught in high school were formulated as early as the 18th century by Amontons and Coulomb, and they had even been found much earlier by Leonardo da Vinci [1]. The kinetic friction force between two sliding macroscopic bodies is given by

\[ F = \mu N, \]  

where \( \mu \) is the friction coefficient and \( N \) is the normal force acting at the interface.
with \( N \) the normal load and \( \mu \) the friction coefficient determined by the contact conditions. The static friction force, which is the maximum lateral force before sliding starts, is usually also given by Eq. (1), albeit with a higher value for the coefficient of friction. The simple relation (1) implies three rules: the friction force is proportional to the normal load \( N \), friction is independent of the (visible/apparent) contact area \( S' \), and friction is independent of the sliding velocity \( v \). The linear proportionality to \( N \) and independence of \( S' \) are easy to understand, when we realize that practical surfaces are always rough and the real area of contact \( S \), which is usually composed of a large number of micro-asperities, is much smaller than \( S' \) [1, 16]. Both elasticity of the contacting bodies and, at higher contact pressures, their plastic response, make that the total contact area \( S \) is approximately proportional to the total normal load \( N \). A constant shear stress over the area of true contact then makes the friction force \( F \) linear in the area of true contact \( S \) and therefore also linear in the normal force \( N \).

These simple results hold for dry friction without wear. In the practically important cases of lubrication and of friction with wear, both of which we do not touch in this article, the situation is more complicated. We should realize that, even for dry friction without wear, the Amontons–Coulomb–DaVinci law (1) is no more than approximate. For example, experiments indicate that there is a finite time involved in the formation of micro-contacts between macroscopic bodies, which is derived from the observation of a logarithmic decrease of the friction force with increasing velocity, while the static friction force increases with the time that the system has been at rest, see, e.g., [17]. As mentioned by Feynman in 1963, “It is quite difficult to do quantitative experiments in friction, and the laws of friction are still not analyzed very well, in spite of the enormous engineering value . . . At the present time, in fact, it is impossible even to estimate the coefficient of friction between two substances” [18]. We have not come very far beyond this poor level of understanding, in spite of the intense investigations performed since then within the rapidly developing field of tribology [2–4]. Friction appears to have become even more of a puzzle since the advent of nanotribology.

The Amontons–Coulomb–DaVinci law fails at the nanoscale. It turns out that features of atomic scale friction are very different from those characteristic for macroscopic friction. This is so in many respects, including the dependence on contact area, normal load, velocity, and temperature, as well as the relation between the kinetic and static friction. One striking difference between behavior on the nano- and macro-scales that will be discussed below in detail, is that atomic-scale friction does not decrease but rather increases with increasing velocity.

It is important to realize that there is a world of difference between the atomic-scale friction that forms the focus of this article and the realm of practical friction between macroscopic bodies. Extrapolating between these scales is usually quite unsuccessful. In fact, as we will see in connection with the phenomenon of superlubricity, when we naively extrapolate the friction behavior that is found on the atomic scale, we should expect friction to be very low on the macroscopic scale. The background for this discrepancy lies in the addition of further mechanisms for dissipation of energy and momentum, that become dominant on the larger spatial scales involved and the corresponding temporal scales.

2.2 Basic experimental approaches in nanotribology

Three instruments play a central role in nanotribology experiments and we will introduce them briefly here. Most relevant to the results discussed in this article is the FFM, which was invented very soon after the AFM and is used to address frictional forces on the very atomic scale [13]. Like the AFM, an FFM has an atomically sharp tip that is brought in contact with the surface of a material and scanned over it, while the forces on the tip are recorded. Usually, this is done with the use of a cantilever, at the end of which the tip is located. Typically, the normal force, exerted by the substrate on the tip, is kept constant, as is also customary in so-called contact AFM. This is done by regulating the height of either the substrate or the support that holds the cantilever with the tip. Most FFMs are actually advanced versions of an AFM. The normal force is measured from the reflection of a laser beam from the end of the cantilever. The bending of the cantilever leads to a modest up- or downward shift of the reflected beam, which is readily measured by use of a split photodiode. The lateral force in the direction perpendicular to the length axis of the cantilever leads to a modest torsion of the cantilever. This torsion causes a sideways shift of the reflected laser beam. The normal and torsional forces can be measured simultaneously by use of a four-quadrant photodiode. Deriving the lateral force in the direction of the cantilever is in principle also possible by measuring the shape of the deformed cantilever more precisely; usually, this is not done and the friction measurement is restricted to a single, lateral direction, Fig. 2.

Several alternative FFM designs have been introduced, of which we show the one developed in the laboratory of the authors [19], see Fig. 3. It features a dedicated force sensor in the form of a MEMS device that is sensitive not only to the normal force but also to both components of the lateral force [20]. Readout of the three force components is done by use of four, simultaneously operated laser interferometers that reflect light of the four faces of a central pyramid on the MEMS structure. This sensor adds two favorable properties to the FFM, namely (i) sensitivity to both components of the lateral force and (ii) the combination of excellent sensitivity to these lateral forces with high stiffness in the normal direction; the latter property makes it possible to bring the tip and substrate in contact without the familiar and often destructive “snap to contact” due to the attraction between the two.

The second instrument that we mention here is the quartz crystal balance (QCB), which provides direct sensitivity to the sliding and friction between an atomic or molecular layer and a flat substrate [6]. The QCB measures the vibrational response and damping characteristics of a quartz resonator, placed in a vacuum system. In particular the shear motion...
is followed, parallel to the quartz surface. When thin layers are deposited either on this surface or on a thin film with which the quartz has been pre-coated, the eigenfrequency of the resonator is lowered, which can be quantified with extreme, sub-monolayer precision. However, when a layer is completely free to slide with respect to the substrate, it does not participate in the oscillatory motion and, hence, does not lower the frequency. The rate at which the slipping of the layer consumes energy, which is directly related to the (dynamic) friction force, is reflected in changes in the quality factor of the resonance, which can in principle be read off from the frequency profile, but is usually determined from the reduction of the amplitude at which the driven oscillator resonates.

The third instrument that is widespread in the field of nanotribology is the surface forces apparatus (SFA) [21]. It brings two cylindrically bent mica surfaces in controlled contact with each other. The cylinder axes are oriented perpendicular to each other, so that they define a single contact point. Under the influence of the combination of the adhesive force between the two mica surfaces and the externally applied normal load, this point grows out into a macroscopic area of contact, with a diameter typically in the micrometer regime. The primary application of this instrument is to the forces, structural organization, and the squeeze-out and flow physics of thin films of molecules in the contact area between the two mica surfaces. The method takes advantage of the extreme, atomic flatness of well-prepared mica, which presents an ideal, flat, and completely parallel confinement geometry for the molecules.

In this article, we concentrate on the atomic-scale friction between unlubricated, solid surfaces, which is experimentally most directly addressed by the FFM technique.

### 2.3 Fundamental mechanisms of frictional energy dissipation

This constitutes the basic focus of nanotribology. Friction is concerned with the coupling between the relative motion of two bodies and their internal degrees of freedom. A variety of coupling mechanisms exist, depending on the nature of the solids (e.g., insulator, conductor, and semiconductor), sliding velocity, and temperature. A number of investigations have been devoted to the theoretical analysis of the possible phononic, see, e.g., [22–27], electronic, see, e.g., [28–31], electromagnetic [10], quantum [32] or non-contact Van der Waals friction [33–35], and other [36, 37] mechanisms. The simplest, phononic mechanism of frictional energy dissipation for slowly moving objects is traditionally associated with the generation of sound waves, as considered, e.g., in the motion of adsorbate atoms on surfaces [38] and in non-contact atomic force microscopy [33]. Unfortunately, each of these mechanisms has been considered separately, with the emphasis on specific situations, and a “unifying” description is lacking. Furthermore, the theoretical exercises devoted to these mechanisms contain tacit, sometimes disputable assumptions and, most importantly, they mostly lack experimental verification.

An important question is whether and how the basic mechanisms can be accessed experimentally. Interestingly, the traditional Prandtl–Tomlinson description of atomic-scale friction in terms of mechanical stick–slip instabilities appears so successful, that it obscures the actual mechanisms of energy dissipation. Within this model, in the typical stick–slip regime (discussed below in more detail), the...
As far as the transition to heat is concerned, friction is usually not made explicitly in the tribology literature. This final stage of the process is characteristic for friction and, apparently, in the underlying dissipation scenario, with tiny, at first glance inessential, changes in the system [39–41]. Some experimental results even provide indications for specific mechanisms. For example, a strong difference in friction between hydrogen- and deuterium-terminated single-crystal surfaces [39], when nothing else than the characteristic frequency of the surface atoms is changed, unambiguously indicates a phononic mechanism of dissipation. In another interesting experiment [41], a factor of 2 difference in friction between single- and bi-layer graphene was attributed to a dramatic difference in phonon–electron coupling. There is an indication for electron–hole pair creation in the damping of molecular motion relative to a solid substrate [42]. It was observed recently that the friction forces acting on a sharp tip sliding over a silicon surface are influenced by the local electronic doping conditions [43, 44], which has been attributed to different degrees of excitation of electronic processes in the substrate.

A direct experimental suggestion for an electronic-excitation based mechanism has been reported in QCB measurements of the friction between an adsorbate layer and its substrate [45, 46] at temperatures above and below the superconductivity transition of the substrate. However, these observations have not been confirmed independently [47, 48], and recent QCB experiments on a similar system produced results that rather suggested the absence of a sizeable electronic contribution to friction [49]. Recent non-contact AFM experiments on energy dissipation between a sharp silicon tip and a Nb film in a range of temperatures around the superconducting transition of the substrate [50] have convincingly shown that electronic friction was the main dissipative channel in the metallic state, and it smoothly decayed below the critical temperature, until phononic friction became dominant. These observations certainly do not settle the discussion, since the characteristic forces in the non-contact case were many orders of magnitude smaller than those in typical situations with sliding bodies in direct contact with each other. We have to conclude that at the present stage most of the experimental evidence on the dissipation mechanisms is indirect and some is even closer to speculation, see, e.g., [51]. More effort is required to come to a consistent theoretical description and to direct experimental access.

By definition, friction is due the irretrievable loss of momentum, accompanied by the transformation of mechanical energy into heat. Strictly speaking, the irretrievable loss and the transition to heat are two different parts of the dissipation scenario, which do not necessarily take place on the same temporal and spatial scales. This important distinction is usually not made explicitly in the tribology literature. As far as the transition to heat is concerned, all internal degrees of freedom of the substrate (vibrational, electronic, etc.) are involved. However, this final stage of the process is not necessarily relevant, since the friction force is determined only by the characteristic rate at which the momentum loss becomes irretrievable. Consequently, the central questions are (i) which degrees of freedom are directly coupled to the relative motion of the bodies and (ii) how are the energy and momentum irretrievably lost, that were invested into these internal modes?

Direct coupling of mechanical motion with electronic excitations seems insignificant at the low sliding velocities typical in the FFM. Most likely, the dominant scenario should be related with vibrational excitations, which further dissipate energy and momentum in a purely phononic manner or, for conductors, with the possible assistance of electron–phonon coupling. Even in this seemingly simple case, the full scenario is complex. Besides the dissipation that is inherent in the details of the contact and the tip and substrate materials, the friction force depends also strongly on the measuring system. Below, we will discuss a variety of dissipation scenarios and their dependence on the stiffness of the measuring system, the temperature, and other parameters.

### 3 Stick–slip motion versus continuous sliding

Atomic resolution in two-dimensional lateral force scans, as obtained routinely in FFM experiments, see Fig. 4, demonstrates regular stick–slip motion, with the tip following the substrate lattice periodicity. This forms the basis for several interesting physical questions that we discuss in this section.

**Figure 4** (a) Two-dimensional map of the lateral forces observed in an FFM experiment on graphite [52]. The yellow arrows indicate the 0.246 nm lattice spacing. (b) Similar force map computed with a two-dimensional, single-spring Prandtl–Tomlinson model with thermal excitations. Data in panel (a) reproduced by courtesy of Prof. A. Schirmeisen (Justus-Liebig-Universität, Giessen, Germany). Calculations in panel (b) performed by D.W. van Baarle (Leiden University, The Netherlands).
3.1 Universality of stick–slip motion As illustrated already by the very first FFM experiment [13], stick–slip motion occurs not only on macroscopic length scales, but also over atomic distances. Generally, the origin of the discrete, non-continuous sliding lies in the combination of two necessary elements, both naturally present in many situations. First, any driving system is flexible and there is an effective spring that actually forces one body to move with respect to the other. In the FFM, the driving spring is simply the cantilever, or, more accurately, it represents the combined stiffness of the cantilever and the tip. Second, the lateral interaction between the bodies should be a function of time or space. In the case of atomic-scale friction, the interaction is modulated in space with the periodicity of the substrate lattice. In the case of FFM experiments, the sliding body, i.e., the tip, will stick at a certain lattice position until the force exerted by the driving spring exceeds the maximal shear force, at which point it rapidly slips. During the slip event, the spring force is reduced, which makes the tip come to rest at the next lattice position. The spring force will have to build up again until it becomes large enough to initiate the next slip event.

In the case of the sliding of macroscopic objects over macroscopic distances, roughness and spatial variations in elasticity may serve to spatially modulate the interaction. However, even in the absence of such modulations, stick–slip motion can arise when the lateral interaction strength is an increasing function of time, or, to put it differently, if the static friction force is higher than the dynamic friction force. In that case, sliding will be initiated when the spring force overcomes the static friction force and the slider will come to rest when the spring force is reduced to the level of the dynamic friction force; after this, the spring force will have built up again to overcome the increasing, static friction force, in order to initiate the next slip event. Actually, the situation on the macroscopic scale is much more complex than this simple description, but this falls outside the scope of this article. We mention that the literature contains examples of appealing, nontrivial sliding physics on the macroscale. For instance, in the context of dry friction avalanches, there are experimental suggestions and theoretical models relating macroscopic scale stick–slip motion with self-organized criticality, when energy is released through rapid, avalanche-like relaxation events [53].

Another example of stick–slip motion is its realization in ultrathin liquid films confined between two moving atomically flat surfaces, due to the crystallization and shear yielding of the confined fluid, see [54] and references therein. Recently, this type of behavior was directly observed in FFM experiments at elevated humidity, which revealed regular stick–slip motion with the period corresponding to the lattice of ice and thus provided evidence for ice formation at room temperature [55].

What we see is that stick–slip motion is a rather natural phenomenon, while continuous sliding would require a sufficiently hard driver or sufficiently low lateral forces between the sliding bodies. We also realize that the dissipation scenarios include the investment of the driver energy into the stretching of the spring during the stick–part of the stick–slip cycle, and the release of this energy in the slip event, accompanied by irretrievable losses to internal degrees of freedom.

Returning to the main theme of this article, dry nanoscale friction, we will see that atomic stick–slip motion, though simple at first glance, is actually a rich phenomenon, the details of which are not yet fully understood.

3.2 Atomic stick–slip motion A typical manifestation of atomic stick–slip motion in an FFM experiment is shown in Fig. 5a. One observes a saw-tooth-like dependence of the lateral force as a function of the support position (or time), somewhat perturbed by thermal noise but with the period of the substrate lattice clearly recognizable. Per lattice spacing, the lateral force goes through significant variations that may even involve a change in sign, from reverse to forward forces. The friction force is defined as the average value of the lateral force and it corresponds to the average amount of energy dissipated per unit traveled distance.

Atomic stick–slip motion is easily described by the above-mentioned model, first proposed by Prandtl [14] and by Tomlinson [15], see Fig. 1b. The sliding object (the tip) is considered to move in a periodic potential field formed by its
interaction with the substrate lattice. It is dragged along the surface by a rigid, external support, via a macroscopic spring (the cantilever), which is at the same time used to measure the lateral force experienced. The total potential energy of the system consists of the sum of the periodic tip–surface interaction, \( U_i(x) \), usually approximated by a sinusoidal form, and the parabolic contribution from the deformed spring:

\[
\begin{align*}
U_{\text{tot}}(x, X) &= U_i(x) + \frac{1}{2} k_{\text{eff}}(x - X)^2, \\
U_i(x) &= \frac{1}{2} U_0 \left[ 1 - \cos \left( \frac{2\pi x}{a} \right) \right].
\end{align*}
\]

Here, \( x \) is the position of the tip and \( X = V t \) is the position of the support, with \( V \) the scanning velocity, \( k_{\text{eff}} \) represents the stiffness of the effective spring, \( U_0 \) is the corrugation of the interaction potential (the height of the potential barriers with respect to lateral displacement), and \( a \) is the substrate lattice constant (period). \( U_{\text{tot}} \) as a function of \( x \) is a corrugated parabola, whose shape depends on \( X \), see Fig. 6. When the corrugation is sufficiently large, it contains a number of wells, which periodically appear, evolve, and disappear as a function of \( X \), i.e., in the cause of scanning. In the simplest variant of the model, neglecting thermal effects, the tip is assumed to reside at the bottom of a certain well. This is “the stick” part of the motion cycle. When the support moves, the barrier between this well and the next one is reduced. At the point where this barrier vanishes, the system is mechanically unstable and the tip necessarily “slips” to the next well.

The recorded lateral force is given by \( F = k_{\text{eff}}(X - x) \). Minimizing \( U_{\text{tot}} \) with respect to the tip position \( x \), i.e., \( \delta U_{\text{tot}}/\delta x = 0 \), and plotting the force \( F \) at the optimal tip position against the support position \( X \), one obtains the variation of the lateral force as shown in Fig. 7. The simple, sinusoidal dependence, that would characterize the spring-free case for which \( x = X \), is re-shaped by the presence of the external spring. The lower branches of the force loops cannot be reached. Instead, the tip exhibit slips to the next well, as shown by arrows in Fig. 7. As a result, the observed variation of the lateral force with support position has a saw-tooth shape (Fig. 5b), very similar to the typical experimental observations (Fig. 5a).

As long as the characteristic frequency of the system, \( \nu_c \sim \sqrt{k_{\text{eff}}/M_{\text{eff}}} \), with \( M_{\text{eff}} \) the effective mass of the slider, is high with respect to \( V/a \), as is typical for FFM experiments, the slips events should follow the vertical arrows in the \( F(X) \) graph. This makes the mean friction force \( \langle F \rangle \), i.e., the time or space average of the lateral force \( F \), independent of the details of the slip event. Instead, it is determined completely by the maximum lateral force \( F_{\text{max}} \) and the slope of \( F(X) \) in the sticking state. The former is determined by the potential corrugation \( (F_{\text{max}} = \pi U_0)/a \), for the sinusoidal potential), while the latter is basically given by the spring stiffness \( k_{\text{eff}} \) (strictly speaking, there is an easily derived correction to the slope that depends also on \( U_0 \) and \( a \), see, e.g., \[57\]). Consequently, in the simplest model, the friction force is completely determined by three system parameters: the corrugation of the interaction potential \( U_0 \), the effective spring stiffness \( k_{\text{eff}} \), and the substrate lattice constant \( a \).

One should not think that the Prandtl–Tomlinson model is only applicable to ideal, single-atom contacts, which exhibit an obvious periodicity in the interaction potential with the substrate. Also for a multi-atom contact, the total tip–surface interaction, i.e., sum of all pair interactions between the tip and substrate atoms, will always be periodic, with the period still determined by the substrate lattice, see Fig. 8. This is so, regardless of the number of atoms in the contact, the tip’s internal structure (commensurability, crystalline versus amorphous) and other details of the contact. This explains the success of the model to explain the saw-tooth-like shape of
Figure 8. Even when the tip is not atomically sharp, the interaction between the tip and the (crystalline) surface is always periodic, with the period of the substrate lattice $a_1$.

the lateral force loops typically observed in FFM experiments under a variety of contact conditions.

In terms of the dissipated energy, which is half the area included in the force loop between forward and back scans, see Fig. 5, the stick–slip scenario is easy to understand. The driver (in this case the support) invests energy in the effective spring by stretching it during the stick-part of each cycle. The invested energy is released and dissipated in every slip event. In this simple model, the dissipation is independent of scanning velocity and temperature. We also observe that this model also does not contain a single parameter that is related to the physical dissipation mechanism. This reflects the simplification that the nonequilibrium aspects of the stick–slip, which are hidden behind the tacitly made, yet crucial assumption that the excess energy is rapidly and completely dissipated in each slip event. In order to understand the arguments behind this assumption and its implications, we should extend the Prandtl–Tomlinson model with an explicit treatment of the inherent dissipation rate.

3.3 The problem of critical damping A first step to incorporate damping explicitly in our friction model, is via the following equation of motion of the tip,

$$M_{\text{eff}} \ddot{x} = -\frac{dU_i}{dx} - k_{\text{eff}}(x - Vt) - M_{\text{eff}} \eta \dot{x}.$$  \hfill (4)

The damping factor $\eta$ has the meaning of the momentum relaxation rate. It reflects the energy dissipation in the motion of the tip along the surface, without specifying the nature of this dissipation (phononic, electronic, etc.). The term $M_{\text{eff}} \eta \dot{x}$ is sometimes referred to as “viscous friction,” in view of its formal analogy with viscous friction in liquids, but this analogy is actually poor. Rather, the reason for the linear relation between the damping force and the tip velocity $\dot{x}$ can be found in linear response arguments. Note that the mass $M_{\text{eff}}$ is introduced in this term only for the formal reasons to provide $\eta$ with the convenient dimension of a frequency.

Note, that the Prandtl–Tomlinson model can be viewed as the stationary solution of Eq. (4), with which the stick epochs are described, in combination with the assumption of instantaneous and complete dissipation in the slip events.

From the perspective of nonequilibrium statistical mechanics, one can think about the tip–substrate complex as a system with many degrees of freedom. We are interested in

a reduced description of the evolution of only a single “reaction coordinate” (the tip position with respect to the surface, $x$) averaged over all other degrees of freedom of the complex system. Generally, such a reduction leads to an equation of motion that includes a dissipative force, which results from the nonretrievable loss of energy and momentum to the other degrees of freedom, plus a random force (noise) that is experienced due to fluctuations in the other degrees of freedom. If the deviation from equilibrium can be considered modest and if memory effects can be neglected, one arrives at a Langevin equation [58, 59]

$$M_{\text{eff}} \ddot{x} = -\frac{dU_i}{dx} - k_{\text{eff}}(x - Vt) - M_{\text{eff}} \eta \dot{x} + R,$$  \hfill (5)

in which the random force $R$ should be related with the damping factor $\eta$ via the fluctuation–dissipation theorem,

$$\langle R(t)R(t') \rangle = 2M_{\text{eff}} \eta k_B T \delta(t - t').$$  \hfill (6)

We can view Eq. (4) as the limiting case of Eq. (5) when thermal effects can be neglected. For the sake of accuracy, we mention that, strictly speaking, $U_i(x)$ in Eqs. (2), (4), and (5) is not just a mechanistic potential, but rather an effective potential that also incorporates entropic effects and, generally speaking, can depend on temperature.

Equations (4) and (5) have been widely used for computational studies of atomic-scale stick–slip (see, e.g., the review papers [4, 5] and references therein), even though the underlying assumptions have not been thoroughly justified. In the next sections, we will return to equations of type (5) in connection with velocity and temperature effects in friction.

Nanotribological computations typically assume the damping to be critical or close to critical, usually without any explanation. Apparently, one is forced to do so, since in the strongly underdamped case, the simulated force-versus-position loops would lose the atomic-scale regularity, in contrast to the typical experiments (see Figs. 4a and 5a), while in the strongly overdamped case they would lose the characteristic saw-tooth shape, again in contrast to experiments. In order to reproduce sliding dynamics, similar to what is observed in typical FFM experiments, one is forced to assume that the damping factor $\eta$ falls in a narrow range around its critical value; see, e.g., [60]. More direct information on the value of $\eta$ can be obtained by analyzing the occurrence of long (multiple) slips, which will be discussed in more detail below. In this way, the tip motion has been found to be slightly underdamped at low normal load and slightly overdamped at high load [61]. The hidden assumption of “instantaneous” dissipation in the mechanistic Prandtl–Tomlinson model, which so nicely reproduces the regular character and the characteristic shape of $F(X)$ loops, Fig. 5b, can be seen to reflect critical or supercritical damping. Clearly, the FFM experiments suggest that damping is always close to critical.

With very few exceptions [10, 62], one usually does not pay attention to the fact that the necessity for this specific
choice of the damping factor looks strange from a physical point of view. In fact, critical damping means that \( \eta = 2 \omega_0 \), with \( \omega_0 \) the characteristic frequency of tip motion in a well of the total potential \( U_{\text{tot}}(x) \), see Eq. (2). It depends equally on the curvature of the contact potential and the effective spring stiffness \( k_{\text{eff}} \), and it is inversely proportional to \( \sqrt{M_{\text{eff}}} \).

One might have expected \( \eta \), which is a measure of the inherent, microscopic dissipation process, to be completely determined by the contact conditions and material (dissipation) parameters of the solids in contact. The actual situation turns out to be different. What is the reason for the damping factor to be mainly determined by the stiffness of the measuring system and the mass? So far, this issue has not received serious attention.

### 3.4 The crucial role of stiffness

Tomanek et al. [63] were possibly the first to realize that “the friction force depends not only on the interaction potential, but even more critically on the construction parameters” [of the measuring system]. This has become common knowledge in the nanotribology field, but there are further counterintuitive aspects.

As predicted by the Prandtl–Tomlinson model (see above) and observed in experiments (see below), friction monotonically increases with increasing corrugation of the interaction potential. This result is easy to understand, since higher potential barriers will more strongly resist lateral motion of the tip. What the model also predicts is that, in the stick–slip regime, friction reduces monotonically with increasing stiffness of the driving spring, \( k_{\text{eff}} \), see Fig. 9a. Formally, this results from the increase in the slope of the force-versus-position curve \( F(X) \) that automatically (for a given potential corrugation and, hence, given \( F_{\text{max}} \)) leads to a decrease in the value of \( \langle F \rangle \), as illustrated in Fig. 9b. In terms of the energy dissipation, the dependence of the friction force on the system stiffness simply reflects that more important than the efficiency of the transfer of mechanical energy into heat, which is described by \( \eta \) in Eqs. (4) and (5), is how much energy is invested in the spring. At the same force, a stiffer spring (higher \( k_{\text{eff}} \)) invests less energy and, hence, experiences a lower mean friction force.

As we are about to see, the situation is more complex, even in the seemingly simple setting of FFM experiments. The result is a variety of dissipation scenarios and corresponding friction regimes.

### 3.5 From stick–slip motion to continuous sliding

One of the important predictions of the Prandtl–Tomlinson model is the transition from stick–slip motion to continuous, nearly frictionless sliding. The necessary condition for stick–slip is that, for certain ranges of the support position, the total potential (2) should exhibit two or more local minima with respect to the tip position; it is this situation that leads to slips of the tip from one energy minimum to the next. If we stay with our earlier assumption that \( U_{\text{tot}}(x) \) is sinusoidal, the number of wells in the total potential is determined by the dimensionless parameter

\[
\gamma = 2 \pi^2 \frac{U_0}{k_{\text{eff}} a^2}.
\]

For \( \gamma > 1 \), there are two or more wells, and their number increases with \( \gamma \). Generally, there will be \( n \) or \( n + 1 \) wells (for \( n \geq 2 \)), depending on the support position, if \( \gamma \) is in the interval between approximately \( (n - (1/2)) \pi \) and \( (n + (1/2)) \pi \). In this multi-well case, the system is in the stick–slip regime and the mean friction force \( \langle F \rangle \) increases with \( \gamma \), reflecting the two main trends of its increase with increasing \( U_0 \) and with decreasing \( k_{\text{eff}} \), as discussed above. It depends only weakly on the inherent damping coefficient \( \eta \).

For \( \gamma < 1 \), there is always only a single local minimum in the potential, regardless of the support position. In this case, there are no mechanical instabilities and the tip moves continuously together with the single minimum in the total potential. The instantaneous lateral force can still exhibit considerable periodic variations, but these are nearly symmetric in sign, reflecting periodic acceleration and deceleration of the tip motion by the surface potential. The friction force \( \langle F \rangle \) is very low in this case, though nonzero, since a component of the order of \( M_{\text{eff}} \eta V \) remains (see Eqs. (4) and (5)), with \( V \) the scanning velocity, which is due to direct damping of the tip motion with respect to the surface.

Figure 9b shows how the friction force \( \langle F \rangle \) is reduced as the effective spring is made stiffer, i.e., as \( \gamma \) is reduced, see
Eq. (7). It is useful to think of $\gamma$ as an order parameter. At the critical value $\gamma_c = 1$, the total potential changes its shape qualitatively (see the two diagrams in Fig. 10), and this is accompanied by a principal change in character of the motion from stick–slip to continuous sliding and of the corresponding dissipation scenario. There are further critical values of $\gamma$ (4.6, 7.79, etc.) when $n$ (see above) is changed from 1, for which there are one or two wells in $U_{\text{tot}}(x)$, depending on the support position $X$, to 2 or 3, etc. At these critical points subtle details of dissipation scenario can change.

For obvious reasons, the possibility to reduce friction by orders of magnitude, by going over from the stick–slip regime to continuous sliding has attracted significant attention. This transition should be possible even at relatively high corrugations $U_0$ if the measuring system stiffness ($k_{\text{eff}}$) is high enough. Also the value of the lattice constant $a$ matters. Among these three parameters that determine Eq. (7), the potential corrugation is the only one that can be manipulated relatively easily in experiments, as we will see in the next section.

Note, that we are limiting the discussion to the case of relatively low scanning velocities in the nm s$^{-1}$ to $\mu$m s$^{-1}$ range, characteristic for AFM and FFM experiments. At high velocities, when $V/a$ is high compared with the characteristic frequency of the tip motion in the surface potential well, the system will always exhibit continuous sliding. This situation is not accessible by FFM experiments, but it is investigated (see, e.g., [64]) with respect to other tribological applications.

### 3.6 Nearly vanishing friction in experiments

Two straightforward methods exist by which the corrugation of the interaction potential can be made very low. Conceptually, the easiest one is to reduce the normal load of the contact. The potential energy of the interaction between the slider and the substrate is expected to be a function of the normal load. When the load is made low or even negative, in order to partly counteract the adhesive force between slider and substrate, this potential energy can be made arbitrarily low. Hand-in-hand with this reduction in (average) potential also its variation with lateral position should reduce. In principle, this should allow one to make $\gamma$ smaller than unity.

This method of lowering the normal force has been exercised in experiments by Socoliuc et al. [57] for a sharp tip at the end of a silicon cantilever sliding over a NaCl surface. Figure 11 shows force traces recorded in this measurement at different values of the normal force. While the forward and reverse measurements display a distinct saw-tooth shape with clear hysteresis at the higher normal-load values, characteristic of stick–slip motion with non-zero friction, the lateral force shows continuous variations and hysteresis is no longer measurable when the external load is made sufficiently negative, suggesting that $\gamma$ has been reduced below unity in this experiment. We will return to this observation later in this article.

Having to make the normal load small or even negative may be highly undesirable from a practical point of view. Luckily, an alternative approach, in which the normal load does not have to be made small, has been proposed as early as 1990 by Hirano et al. [11, 65]. Their idea was to consider the friction between two atomically flat surfaces and make active use of the mismatch between the two contacting crystal lattices, that of the substrate and that of the slider. When the lateral stiffness of the two lattices is high enough that each can be considered a rigid object, the corrugation of the total interaction potential between the two surfaces scales sub-linearly with the contact area, due to the high degree of cancellation of the lateral forces on the individual atoms in the contact area. This can drive $\gamma$ below unity and make the contact highly slippery, an effect that Hirano et al. nick-named “superlubricity” and that is related to the incommensurability transition introduced before by Aubry [66].

First experimental indications of superlubricity have been presented in Refs. [67] and [68], but the first convincing experimental demonstration of the effect has been given by Dienwiebel et al. [69]. In a measurement of the friction between a tungsten tip and a graphite surface, the tip was shown to pick up a graphite flake, so that the measured lateral force was really that between two graphite surfaces. The corrugation of the interaction potential between the graphite flake and the graphite substrate could be continuously varied by simply rotating the substrate with respect to the flake. Figure 12 shows that the friction force was high for two narrow ranges of relative orientation for which the graphite lattices were almost or completely aligned, while it fell below the detection limit outside these ranges, where the lattices were fully mis-oriented with respect to each other. Individual force traces show the expected transition between large-amplitude variations with hysteresis to small-amplitude...
There have been further observations of nearly vanishing friction, for example in elegant experiments by Dietzel et al. [71], in which small antimony particles on a graphite substrate were pushed around sideways by the tip of an FFM under ultrahigh-vacuum conditions. The observations were naturally divided over two categories. Whereas one class behaved “normally,” exhibiting significant friction, proportional to the area of the particle (i.e., the area of contact), the other was “superlubric,” moving with nearly zero friction, independent of the size of the particles. The incommensurability between the lattices of the antimony and the graphite is responsible for the superlubricity of the second class of particles. This superlubricity is thought to be frustrated for the first class as the consequence of a low density of atoms or molecules, trapped in the contact area between particle and substrate [72–74]. The latter notion was tested by deliberate contamination of the system, which indeed led to a reduction of the fraction of superlubric particles. Recently, indications have also been obtained of superlubric behavior of contacts between graphite surfaces as large as $10 \times 10 \mu m^2$ [75].

Maier et al. [40] have observed the transition from atomic stick–slip motion to the regime of smooth sliding with ultralow friction as the result of a tailor-made modulation in the interaction potential. This modulation was caused by a periodic rumpling at the interface between heteroepitaxial films of KBr and a NaCl(100) substrate. As a result, the motion varied from stick–slip to smooth within each of the $6 \times 6$ unit cells of the rumpling superstructure.

A remark is in place on the name “superlubricity,” originally coined by Hirano and Shinjo [65] and now widely used in the field of tribology [76], also referring to other phenomena than the original, e.g., static, structural, and variations without hysteresis. The variation in friction force was found for a wide range of normal forces, from negative to strongly positive, and it was observed for all sliding directions with respect to the lattice of the graphite substrate (and that of the graphite flake on the tip) [56].

From the size and shape of the friction maximum in Fig. 12, it is possible to estimate the size and shape of the graphite–graphite contact in these FFM measurements [70]. The simple “rule” here is that the minimum angle required to twist the two surfaces so much out of registry that the forces on the atoms in the contact region cancel each other out, is the angle for which the twist adds up to a mismatch of one full lattice unit cell over the diameter of the contact. This geometrical rule directly relates the full-width-at-half-maximum of the measured friction peaks in Fig. 12 with the inverse diameter of the contact, from which a contact area of 96 carbon atoms was derived [56, 69, 70]. Asymmetric contact shapes were shown to result in asymmetric friction peaks; Fig. 12 was shown to be consistent only with a more or less circular shape of the graphite–graphite contact.

There have been further observations of nearly vanishing friction, for example in elegant experiments by Dietzel et al. [71], in which small antimony particles on a graphite substrate were pushed around sideways by the tip of an FFM under ultrahigh-vacuum conditions. The observations were naturally divided over two categories. Whereas one class

Figure 11 Vanishing friction between a silicon tip and the NaCl(001) surface, measured in ultrahigh vacuum at room temperature [57]. The bottom three panels show atomically resolved “stick–slip” patterns at three values of the normal load $N$, with which the tip was externally pressed against the surface, namely 4.7, 3.3, and $-0.47\text{ nN}$, respectively. The negative load for the rightmost panel indicates an external pulling force that was used to compensate a large fraction of the natural tip–surface adhesive force. The force loops in the left and central panels display the familiar “sawtooth” shape for stick–slip motion and exhibit the hysteresis associated with energy dissipation: friction. The force variations in the right panel are different in character; they show continuous force variations, i.e., motion without discontinuities, and the hysteresis is reduced to nearly zero. The top panel shows the friction force, i.e., the average lateral force, plotted as a function of the effective normal load (combination of the external load and the measured adhesion). Friction is seen to disappear at low loads. All plots reproduced by courtesy of Prof. E. Meyer (Universität Basel, Switzerland).
Support position $X$ [nm]
Lateral force $F$ [pN]

Figure 12  Superlubricity between a tungsten tip and a graphite surface [69]. The lower three panels show force loops at three orientations of the tip relative to the graphite surface, namely 60, 72, and 38°, measured at normal loads of 18, 18, and 30.1 nN, respectively. All three panels show a signature from the graphite lattice periodicity, but the hysteresis in the central and rightmost panels is nearly absent, i.e., friction is reduced to nearly zero at these orientations. The top panel shows the friction force, i.e., the average lateral force, plotted as a function of the relative orientation between the tip and the graphite surface. Two narrow peaks of high friction are observed at 0 and 61°. Between these peaks a wide angular range with ultra-low friction, close the detection limit of the instrument was found. The variation has been interpreted in terms of the friction between the graphite surface and a nanoscale graphite flake, firmly attached to the apex of the tungsten tip. The curve through the data shows a calculation based on the Prandtl–Tomlinson model, evaluated for a symmetric 96-atom flake.

3.7 Contact size effects Why is superlubricity not routinely observed on the macro-scale? After all, the chances are very slim for a large contact to be ideally commensurate. Here, we touch upon a few effects that may be of importance, when going from ideal nano-contacts to “real” macro-contacts.

In an ideally commensurate contact, the potential corrugation and the corresponding quasi-static forces should increase linearly with the number of contacting atoms $N$ (for $N$ well above 1). If the contact is not commensurate, the potential corrugation increases much more slowly, as $\sqrt{N}$. This can be understood as the statistical result of the summation over the “random” collection of forces on the individual atoms in the contact, each with a different value and a different direction. As a consequence, the friction coefficient, experienced on a macroscopic scale should then be expected to reduce as $1/\sqrt{N}$.

In practice, large contacts are almost never commensurate. One might think elasticity to re-establish local commensurability at a network of regions within each large contact, thus re-introducing linear scaling with $N$. However, calculations show [78] that this is not the case for contacts between three-dimensional bodies. This leaves us with a big problem. Not only does the Prandtl–Tomlinson model predict dynamic superlubricity [77]. This name is confusing, if not misleading, in view of the suggested analogy with superconductivity or superfluidity [78]. Superlubricity is not a quantum mechanical effect and it does not reduce the friction force rigorously to zero, even when it can be made very low. It is only in the simplest version of the Prandtl–Tomlinson model that zero friction is predicted [63], but this should be regarded as a mere zero-order approximation. A more appropriate name that is sometimes used is “structural lubricity” [78], but in this article we will follow the convention and refer to it as “superlubricity”.

The interpretation of the experiments discussed in this section seems straightforward, in terms of simple transitions from stick–slip motion to smooth, low-dissipation sliding (superlubricity). This seems to be supported by the quantitative fits that have been produced, for example, to the data of the two initial experiments of this type [57, 69]. Interestingly, a more advanced model, introducing extra flexibility in the tip, in combination with effects of thermal activation [79, 80], shows that true superlubricity, i.e., with $\gamma < 1$, has not been reached in either of these experiments. Instead, the nearly vanishing friction in both cases was due to “thermolumbricity” of the rapidly moving nanocontact. This alternative interpretation will be discussed in more detail below.
static friction forces to be much lower than those observed in practice. Also the $\sqrt{N}$ scaling should make the friction parameter $\gamma$ decrease with increasing contact size $N$. Consequently, the model predicts the disappearance of stick–slip instabilities for large contacts and the transition to superlubricity, again in contradiction with practice.

Several “escapes” from the problem have been considered. Müser and co-workers [72–74] have proposed that the presence of mobile molecules (“third bodies”) at the interface can lead to static friction and the appearance of stick–slip instabilities. This is surprising, since naively, one might have expected such molecules to lubricate the contact, rather than to enhance friction. Support for this explanation has come from the experiments with antimony clusters sliding over graphite, mentioned above [71]. Another potential reason for non-vanishing friction is multi-scale surface roughness [81]. Contact between rough surfaces with fractal properties can be established by small-length-scale asperities, sufficiently small to produce sizable static and sliding friction. There have been further ideas [82, 83]. In spite of their physical merits, we wonder whether any of these models provides the proper explanation why high friction is almost always observed in the macroscopic world. Below, we will further refine the Prandtl–Tomlinson model and reveal a variety of new friction regimes. However, also this will not settle the issue of friction in macro-contacts.

Next to the size of the contact, also the system size in the direction perpendicular to the contact can be of importance. Recent experiments with atomically thin sheets of graphene, molybdenum disulfide, and other layered materials [84] revealed a monotonic increase in friction when the number of atomic layers was reduced. This effect was ascribed to out-of-plane elastic deformations that resulted in a larger effective tip–surface contact area for thinner sheets. A thickness effect has also been predicted due to lattice vibrations. Model calculations of friction at high velocities (continuous sliding) [26] indicated friction to increase with the number of atomic layers in the solid, as a consequence of the role of bulk atoms in the vibrational dissipation process.

### 3.8 Nontrivial slip dynamics: Multiple slip events and damping

Stick–slip motion with simple, single-lattice-spacing slips is typical but not unique. If there are three or more local minima in the total potential (2) and if the dynamics is underdamped, multiple (double, triple, ...) slips are possible. Starting from a mechanical instability point, the tip can pass – ballistically – over one or more lattice positions before becoming trapped, see Fig. 13a, since the time needed for dissipation of the excess energy is longer than the time to traverse one or more intermediate wells.

Although double slips have been observed already in the very first FFM experiment [13] and multiple slips were observed in several recent studies, single slips are typically dominant. This is similar to surface diffusion of atoms and small atom clusters: multiple jumps of adparticles are rarely observed, see, e.g., [85]. The situation there is simplified by the fact that the number of surface wells accessible for an adparticle is infinite, so that the occurrence of multiple jumps is completely determined by the damping rate [86], which turns out to be close to critical. In the context of friction, the occurrence of multiple slips requires two conditions to be satisfied, namely sufficiently low damping (for details see, e.g., [87]) and the availability of more than a single potential well. The number of wells is determined by the friction parameter $\gamma$ (7), and it can be large if the effective spring is weak or if potential corrugation is strong. Recent FFM measurements on graphite [88] showed transitions from single to double and to triple slips with increasing load, in agreement with this trend, assuming that an increasing normal load is accompanied by an increasing corrugation. Another recent observation of multiple slips [61] is illustrated in Fig. 13b. One notices different slips within a single scan line, nicely demonstrating the probabilistic nature of the process. This stochastic element directly reflects the random force $R$ in the equation of motion (5).

Important is that the observed multiple slip events enable us to estimate the value of the damping factor $\eta$, the essential, yet hardly accessible parameter in the theory. From a comparison between computer simulations and experimental...
observations of multiple slips [61], it was concluded that damping increases with normal load. This is in qualitative accordance with our expectation, because a higher load should correspond to stronger coupling of the relative motion of the tip and the substrate with their internal degrees of freedom.

Obviously, multiple slip events reduce the mean friction force. The calculations of [89] show that \( \langle F \rangle \) can depend non-monotonously on the contact potential corrugation \( U_0 \), with discontinuities at the “critical points,” i.e., the critical values of \( y \) \( (7) \), where the number of wells changes. Further interesting multiple-slip physics [90] will be discussed below, in connection with the temperature dependence of friction.

In typical FFM experiments, the dynamics of individual slip events is not resolved. Relative to the low scan velocities, slip events are perceived as instant jumps. However, recent high-resolution experiments [91] revealed rich slip dynamics, in which the tip can slip from one lattice position to another not only directly, but also via several jumps back and forth, or even via an apparent “intermediate position.” The authors of [91] invoked a complex, multi-tip contact to explain these observations. Below, we discuss a more simple explanation, in which we give account of the flexibility of the tip apex itself.

3.9 The role of dimensionality For completeness, we emphasize that, even though the description provided in this article has been restricted to one-dimensional motion, with the tip being pulled by the support through a one-dimensional energy landscape, the relative sliding of two contacting surfaces is inherently two-dimensional. An appropriate theoretical approach is to solve two coupled equations of motion of type \( (5) \), one for the \( x \) direction and one for \( y \). Even though quantitative complexity is introduced by zigzag-like motion of the tip, no essentially new phenomena are introduced in this way. Examples of experiments and computer simulations demonstrating such motion can be found in Refs. [92–95]. Recent measurements of slip length distributions [52] could not be described by a mere one-dimensional, thermally activated Prandtl–Tomlinson model and were indicative of the two-dimensional character of the interaction potential.

Another consequence of the two-dimensional nature of the motion is the possibility of so-called “friction anisotropy,” which is the dependence of the sliding motion and friction force on the sliding direction with respect to the main crystallographic directions (see, e.g., [12] and references therein). Interesting complications arise when friction is studied on the inherently non-periodic surfaces of quasi-crystals [96, 97]. Friction anisotropy is emphasized on surfaces with steps [98]. Interestingly, both step-up and step-down scans exhibit higher frictional forces at the edge, but they were different in their load dependence: while the additional frictional force due to the step increases linearly with load if the tip has to climb a step, it remains constant when the tip descends. This represents a universal effect that can be explained in terms of a modified Prandtl–Tomlinson model featuring a Schwoebel–Ehrlich barrier at steps.

We close this section by briefly mentioning the coupling with the rotational coordinate that has been considered experimentally and in computer simulations in relation to the geometry of graphite flakes sheared over a graphite surface [99]. Even when the sum of all forces on the carbon atoms in the flake is close to zero and the flake can be translated without stick–slip instabilities, the torques on all carbon atoms can still add up to a significant total torque, especially when the flake has been twisted out of registry with the substrate only over a modest angle. If the flake is free to rotate and respond to the torques to which it is subjected, it will tend to rotate into registry [99]. During the translational motion, the flake may be frequently thrown out of registry, but in its rotational random walk, it displays a significant preference for aligned geometries. When flakes are sheared between two, parallel but azimuthally misoriented graphite surfaces, there is rotational frustration between each flake and at least one of the two graphite surfaces, which keeps the entire contact in the superlubric state [100].

4 Effects of temperature and velocity On the macro-scale, dry friction is often (though not generally) observed to decrease modestly, i.e., logarithmically, with increasing velocity, which has been associated with creep-induced growth of the load-bearing contact area, see, e.g., [17] and references therein. In atomic-scale friction, when the contact size is kept fixed, Eqs. (4) and (5) would rather predict the opposite behavior, with, to first approximation, a linear increase of friction with velocity.

It may not be directly clear what temperature dependence to expect. Based on the dissipation rate in the related phenomenon of surface diffusion of adsorbed particles (see, e.g., [101]), one might expect only a weak temperature effect. On the other hand, the phononic mechanism of dissipation may be expected to result in a significant drop of friction at low temperatures (see, e.g., [25]), in view of the Bose statistics of phonons. Actually, all these speculations are applicable only to the superlubricity regime of low-dissipation, continuous sliding, when the friction force is linear in the scanning velocity and it is directly proportional to momentum dissipation rate (damping factor \( \eta \)), as discussed above.

Stick–slip friction is, to a first approximation, independent of both velocity and temperature, as derived in the context of the basic Prandtl–Tomlinson model. This is, because thermal effects in the tip motion are ignored. Experiments indicate, however, that thermal effects are noticeable, even in the stick–slip regime at low temperatures. This observation is remarkable in view of the macroscopically large mass of the slider (the tip–cantilever). As we are about to see, thermal effects in the tip motion lead to surprising velocity and temperature dependences.

4.1 Thermal activation effects on stick–slip motion The role of thermal effects and the related dependence of friction on velocity was anticipated nearly a century...
E. Meyer (Universität Basel, Switzerland).

Fig. 6) and decreases while the lateral force is increased and scanning velocity has indeed been observed in experiments [105]. The relation between friction, temperature, and scanning velocity in the stick–slip regime can then be written as [106, 107]

\[ \langle F_{\text{max}} \rangle = F_{\text{max}}^* - \beta k_B T \ln \left( \frac{V_c}{V} \right)^{2/3} \]  

for \( V < V_c \). The parameter \( \beta \) depends on corrugation and shape of the surface potential, with \( \beta = (3\pi \sqrt{2a}) / (2\sqrt{2a}) \) for a sinusoidal \( U_c \). The characteristic velocity \( V_c \) is proportional to temperature and to the prefactor \( r_0 \) of the activated jump rate (8), \( V_c = (2\beta r_0 k_B T) / (3k_{\text{eff}} \sqrt{2a}) \). Below \( V_c \), friction increases nearly logarithmically with velocity, as discussed above. Above \( V_c \), the thermal energy can not assist slip anymore, and \( \langle F_{\text{max}} \rangle \) will approach the plateau value of \( F_{\text{max}}^* \). The transition from a logarithmic increase to a plateau with increasing velocity has indeed been observed in experiments [105].

For a more detailed description of thermally activated motion of the tip, one can numerically solve the Langevin Eq. (5), see, e.g., [60, 104], or the corresponding Fokker–Planck equation [58] for the probability density [60]. If the relevant potential barriers are large with respect to \( k_B T \), this approach is somewhat excessive, since it follows not only the relevant motion of the tip, between the surface potential wells, but also the irrelevant, rapid vibrations in a well, when the system is in equilibrium. Computationally lighter are approaches based on rate theory [80, 108–110], which account for the tip motion, but average over rapid vibrations in a well. Within this approach, the jump rate (8) enters the theory in a natural way, as a solution of Eq. (5), with the prefactor \( r_0 \) dependent on the damping rate \( \eta \) [58, 59]. According to the classical Kramers solutions, \( r_0 \) scales as \( \eta \) at low damping and as \( 1/\eta \) at high damping. In the intermediate case, close to critical damping, the prefactor reaches its maximum value. The familiar transition state theory approximation, in which \( r_0 \) is simply equated to the vibrational frequency at the bottom of a well, provides an estimate for the prefactor in this case [111].

Calculations based on the Langevin equation or on rate-theory extensions of the thermally activated Prandtl–Tomlinson model have yielded a number of interesting and instructive results. Here, we touch upon a few. Langevin simulations [112] have shown that in two dimensions the friction force is somewhat lower and it varies with velocity somewhat more strongly than in one dimension. A similar, but quantitatively more pronounced effect of dimensionality was obtained in rate theory-like calculations [113]. The influence of the dimensionality can be viewed as an entropic effect, related with the multiplicity of the possible trajectories.

Figure 14 Friction as a function of scanning velocity, measured in an FFM experiment under ultrahigh vacuum at room temperature with a silicon tip on NaCl(001) at normal loads of 0.44 nN (circles) and 0.65 nN (squares) [102]. The reference velocity of the horizontal axis is \( V_0 = 1 \text{ nm s}^{-1} \). Both the average lateral force (friction; open symbols) and the mean value of the maximum lateral force (solid symbols) are shown. Graph reproduced by courtesy of Prof. E. Meyer (Universität Basel, Switzerland).
between which the tip can choose on a two-dimensional surface. Another interesting effect [90] is found when multiple slips occur frequently, in which case there are two competing temperature dependences. The first is the thermally induced reduction of the lateral force needed to initiate slip, as described by Eq. (9). The other is the effect that when a thermally activated slip event starts earlier, the subsequent potential barriers are still higher, as is seen clearly in Figs. 6 and 15b. This reduces the fraction of multiple slips and thereby adds to the friction. Due to the competition between these two trends, the calculations of [90] predict a rich and possibly non-monotonous temperature dependence.

We will discuss later how the dependence of atomic-scale friction on velocity and temperature can be influenced strongly by the rapid dynamics of the tip apex, an effect that is completely ignored within the traditional single-spring Prandtl–Tomlinson model that we have addressed so far. Nevertheless, even within this traditional approach, the role of thermal effects is variable and it is not exhausted with the activated pre-critical slips.

4.2 From stick–slip to thermal drift

As discussed above, a necessary condition for atomic stick–slip motion is that \( \gamma > 1 \), which guarantees that there are two or more local minima in the total potential (2), so that the tip can stick and slip. There are two further conditions: the temperature should not be too high, while the scanning velocity should be neither too high, nor too low.

In order to organize the possible friction regimes over the full range of temperatures and velocities [80, 109, 114] (see Fig. 16a), we identify four basic time scales or frequencies: (i) the frequency with which the support moves over one atomic lattice period, \( V/a \); (ii) the characteristic oscillation frequency of the tip–cantilever combination in a well of the total potential, \( \nu_r = (1/2\pi)\sqrt{(k_a + 2\pi^2U_0/a^2)/M_a} \); (iii) the characteristic rate \( \nu_c \) of thermally activated jumps between potential wells, as given by Eq. (8) with \( U_c \sim U_0 \); and (iv) the damping parameter \( \eta \), which is the rate of momentum dissipation into the phonon bath or other excitations of the substrate or tip. Assuming that \( \eta \) is on order of \( \nu_c \), i.e., motion is not strongly under- or overdamped, we introduce two dimensionless system parameters, in addition to the friction parameter \( \gamma \) (7):

\[
\alpha = \frac{V}{av_c},
\]

\[
\beta = \frac{V}{ar_c} = \frac{V}{ar_0} \exp \left( \frac{U_0}{k_BT} \right).
\]

Here, \( \alpha \) shows how fast or slow the scanning process is with respect to the inherent motion of the tip–cantilever combination in a potential well. \( \beta \) shows how fast or slow the scanning process is with respect to the characteristic rate of activated jumps of the tip between neighboring potential wells, as shown by arrows (C) in Fig. 15a. With the transition-state-theory approximation for the pre-exponential factor, \( r_0 \approx r_c \), these two parameters are related, with always \( \beta > \alpha \), and typically \( \beta \gg \alpha \).
The system will be in a regime of continuous sliding not only when \( \gamma < 1 \), as discussed above (Section 3.5), but also for \( \gamma > 1 \) when the velocity is sufficiently high, \( \alpha \gtrsim 1 \), so that the tip can no longer skip between potential energy wells. In both cases, the mean friction force is linear in velocity and for low velocities it is very small.

For \( \gamma > 1 \) and \( \alpha \ll 1 \), the motion is discontinuous. However, for regular atomic stick–slip to take place, a third condition should be satisfied, \( \beta \gg 1 \), so that thermal jumps do not randomize the sequence of slip events. If, as a first approach, we completely neglect thermal effects, we see that the tip resides, with probability unity, in the original well as long as the well exists, after which it moves to the adjacent well, etc. This is just the scenario of the zero-temperature Prandtl–Tomlinson model, discussed above (Section 3.2), see also arrow (A) in Fig. 15a. However, this approximation is never justified. Thermally activated slip events can never be neglected in the vicinity of the critical positions of the support, where the barriers \( U_0 \) are small and the jump rates (8) become large (arrow (B) in Fig. 15a). This is the origin of both a statistical distribution of slip positions in the stick–slip regime and a \( \langle \text{const} − (T \ln T)/V \rangle^2 \) scaling of the friction force, as discussed in the previous subsection.

In the opposite limiting case, \( \beta \ll 1 \), the tip executes many activated jumps, back and forth between the potential wells, see arrows (C) in Fig. 15a, in the time that the support traverses a single lattice spacing \( a \). We refer to this as the thermal drift regime, as the mean position of the tip \( \tau \) drifts continuously with support position \( X \). The difference \( X − \tau \) determines the mean lateral force. In contrast with the continuous character of the mean force, the instantaneous lateral force exhibits rapid variations of the order of \( k_B T \beta \) due to the frequent inter-well jumps. In the intermediate case, \( \beta \sim 1 \), the tip exhibits on average one activated jump per interatomic distance traversed by the support. As a result the tip performs irregular or stochastic stick–slip motion.

For \( \beta \ll 1 \), the friction force can be found from rate theory [80, 109]. To first order in \( \beta \), the friction force is close to

\[
\langle F \rangle |_{\beta \ll 1} \sim M_\alpha \eta V \frac{U_0}{k_B T} \exp \left( \frac{U_0}{k_B T} \right). \tag{12}
\]

Here, we have approximated the prefactor in Eq. (8) as \( r_0 \sim \nu^2 / \eta \), which is a reasonable approximation for the cases of moderate or strong damping. As the momentum relaxation rate, \( \eta \), is proportional to \( 1/M_\alpha \), the friction force (12) is independent of the macroscopic system mass, as expected.

In the thermal drift regime friction increases linearly with \( V \), and the increase is very steep. This is in contrast to the weak logarithmic increase characteristic for the stick–slip regime, discussed above. Furthermore, there is a strong, exponential dependence on \( U_0 / k_B T \). Consequently, even a small decrease in surface corrugation or a small increase in temperature will lead to a substantial decrease in friction. In this low-velocity regime, temperature acts as a lubricant.

Interestingly, the temperature dependence of (12) is close to the Arrhenius law [115]. In fact, this reflects the nontrivial physics behind the thermal drift motion of the tip. If the thermal jumps had enough time to establish the equilibrium distribution of tip positions around the support position, the mean friction force would be close to \( M_\alpha \eta V \), as a result of the motion with mean velocity of the order of \( V \). However, due to the lack of time at any nonzero velocity, the tip positions occupy a non-equilibrium distribution, with a systematic lag of the tip with respect to the mean equilibrium position that manifests itself as an increase in friction. This effect is inversely proportional to the effective jump rate that goes exponentially with \( U_0 / k_B T \), and so does the friction force.

According to Eq. (12), the friction force in the thermal drift regime can be comparable to or larger than that for continuous sliding \( (M_\alpha \eta V) \), but it remains significantly lower than in the stick–slip regime, when thermal activation plays only a modest role. The behavior of \( \langle F \rangle \) over the entire range of velocities and corresponding friction regimes are shown schematically in Fig. 16a. A steep linear increase of friction with velocity in the thermal drift regime is followed by a weak, logarithmic-like increase with a subsequent plateau in the stick–slip regime (see previous subsection), and finally there should be a linear increase reflecting continuous sliding.

Important is that Eq. (12) predicts friction to vanish in the limiting case of zero velocity,

\[
\lim_{V \to 0} \langle F \rangle = 0. \tag{13}
\]

The reason is that at nonzero temperature, for any given potential corrugation, the activated slips of the tip will have enough time to completely destroy the stick–slip motion, provided scanning is sufficiently slow. Although this result could be anticipated [4] on general grounds, it may seem to contradict earlier results. For example, the familiar, zero-temperature Prandtl–Tomlinson model produces velocity-independent friction, as discussed above. Also, numerical simulations at \( T = 0 \) [5, 116], e.g., based on equations of motion such as Eq. (4), produce nonzero friction in the low-velocity limit, in contrast to Eq. (13). Actually, there is no contradiction with Eq. (13), as we see from Eq. (12) that \( \lim_{V \to 0} (d \langle F \rangle / dV) = \infty \). Consequently, for \( T \to 0 \), the velocity dependence of friction (see gray line in Fig. 16b) transforms into a step function, and the zero velocity limit for the kinetic friction simply becomes indefinite. The case of static friction is somewhat analogous [4, 117]. Because of thermal activation, an appropriate definition should specify (i) how fast the externally applied shear force is ramped up and (ii) how long one is willing to wait to see the slider move by one lattice constant.

### 4.3 Thermolubricity

The specific friction regime in which a contact moves – stick–slip (\( \beta \gg 1 \)), intermediate (\( \beta \sim 1 \)), or thermal drift (\( \beta \ll 1 \)) – depends on \( U_0 \) and \( T \), and also on \( V, a, \) and \( r_0 \). Variable-temperature and low-velocity
low temperatures, high velocities and/or low prefactors, the prior to superlubricity. At higher temperatures and/or at lower temperatures, the thermal-drift regime, as thermolubricity friction at $\gamma < 0$ can produce substantial friction in the traditional, mechanistic theory. In analogy with the superlubricity that takes place at $\gamma > 1$, and hence it was due to thermolubricity, rather than superlubricity. Note, that the $\gamma$ values of the experimental data points in Fig. 17 were obtained from the measured force loops assuming that $F_{\text{max}}(\gamma)$ is determined by $\pi U_0/a$, while it is actually somewhat lower, due to thermal effects. Consequently, the experimental data should all be shifted to higher $\gamma$ values, making the thermal effect even more dramatic [80]. Comparing with the friction calculations, one finds a prefactor value, $r_0 > 1.6 \times 10^3 \text{s}^{-1}$, in order of the characteristic frequency $\nu_c$, which seems to be in correspondence with nearly critical damping.

Figure 17 Friction force as a function of relative surface corrugation $\gamma$ for $V = 30 \text{ nm s}^{-1}, a = 0.25 \text{ nm},$ and $k_{ab} = 1.8 \text{ N m}^{-1}$. Black curves are theoretical calculations [80] for $V/a_0 = 5.33 \times 10^{-n}$ with $n = 1, 2, 3, 4$ (from left to right). The gray curve is the prediction of the zero-temperature Prandtl–Tomlinson model. The data points are experimental data from the FFM experiment on graphite of Ref. [69].

FFM experiments, in which these predicted regimes are all addressed, are still lacking. Nevertheless, the effective surface corrugation $U_0$ has been varied over a wide range in the experiments of Refs. [57, 69]. To visualize the possible role of thermal effects for different surface corrugations, Fig. 17 shows calculated friction forces as a function of $\gamma$, for $T = 300 \text{ K}$ and for values of $T$, $V$, $a$, and $k_{ab}$ taken from the experiment [69]. The gray curve is the prediction of the Prandtl–Tomlinson model with thermal effects completely ignored, i.e., $T = 0$. The black curves represent the numerical solution to a rate theory equation [80, 109] over the entire range of $\beta$ for four values of the prefactor $r_0$, each time increasing by one order of magnitude. Even at low values of $r_0$, the friction force is decreased noticeably with respect to the zero-temperature case. Also, at high $\gamma$ values, the black curves do not approach the gray curve, thus illustrating the importance of thermal activation even in the stick–slip regime at high potential corrugations.

Figure 17 demonstrates that friction can be negligibly low even when the surface corrugation is large enough to produce substantial friction in the traditional, mechanistic theory. In analogy with the superlubricity that takes place at $\gamma < 1$, we refer to the effect of substantial suppression of friction at $\gamma > 1$, due to the transition from the stick–slip to the thermal-drift regime, as thermolubricity [80]. When the surface corrugation is reduced, thermlubricity should occur prior to superlubricity. At higher temperatures and/or at lower velocities, the $\gamma$ range of thermolubricity will be wider. At low temperatures, high velocities and/or low prefactors $r_0$, the system will switch nearly directly from stick–slip motion to continuous sliding without a distinguishable, intermediate role of thermal effects.

We will return to these thermal effects below, following an approach that explicitly takes into account the low effective mass and the related, rapid dynamics of the nanocontact; this will add further complexity to the phenomenon of thermolubricity.

4.4 FFM observations of thermal effects Various atomic-scale friction experiments have been reported that have been directly or indirectly sensitive to thermal effects. Here, we restrict ourselves to a limited selection. Returning once more to Fig. 17, we see that, in addition to calculated friction forces, the figure also shows data from the experiment of Dienwiebel et al. [69]. Like the calculations, the experimental data deviate significantly from the gray, zero-temperature calculation, suggesting a pronounced role of thermally activated jumps. Nearly vanishing friction was reached in this experiment at $\gamma > 1$, and hence it was due to thermolubricity, rather than superlubricity. Of the direct measurements of the temperature- and velocity dependence of friction, we specifically mention careful FFM experiments by Schirmeisen et al. [118] for silicon tips sliding over a silicon surface and over a graphite surface [106], the latter experiment displaying well-resolved atomic stick–slip patterns. In both experiments, a surprising, non-monotonous dependence has been observed on temperature. The first of these experiments [118] was explained in terms of the thermally activated formation and rupture of multiple bonds in the tip-sample contact [119–121]. Further temperature-dependent friction measurements have been reported, e.g., on MoS$_2$ [115, 122, 123] and diamond [124].

Experiments by Riedo et al. [125] have demonstrated that capillary condensation can lead to a logarithmic decrease of friction with increasing velocity. This has been interpreted as the consequence of the thermally activated nucleation of water bridges between tip and sample asperities, which results in a dynamics that somewhat resembles the velocity weakening observed in macroscale contacts in, e.g., [17]. Measurements of the friction of surfaces modified so that they are sensitive to thermal effects even more dramatic [80]. Comparing with the friction calculations, one finds a prefactor value, $r_0 > 1.6 \times 10^3 \text{s}^{-1}$, in order of the characteristic frequency $\nu_c$, which seems to be in correspondence with nearly critical damping.

Of the direct measurements of the temperature- and velocity dependence of friction, we specifically mention careful FFM experiments by Schirmeisen et al. [118] for silicon tips sliding over a silicon surface and over a graphite surface [106], the latter experiment displaying well-resolved atomic stick–slip patterns. In both experiments, a surprising, non-monotonous dependence has been observed on temperature. The first of these experiments [118] was explained in terms of the thermally activated formation and rupture of multiple bonds in the tip-sample contact [119–121]. Further temperature-dependent friction measurements have been reported, e.g., on MoS$_2$ [115, 122, 123] and diamond [124].

Experiments by Riedo et al. [125] have demonstrated that capillary condensation can lead to a logarithmic decrease of friction with increasing velocity. This has been interpreted as the consequence of the thermally activated nucleation of water bridges between tip and sample asperities, which results in a dynamics that somewhat resembles the velocity weakening observed in macroscale contacts in, e.g., [17]. Measurements of the friction of surfaces modified so that they can form hydrogen-bonding networks also show a reduction of the friction force with increasing sliding velocity [126], which has been explained analogously, in terms of the formation and rupture of these bonding networks.

4.5 Mechanical versus thermal excitation Analogous to thermal excitation as a means to “jump” the energy barriers rather than “climb them,” also mechanical excitations can set the system in motion. It seems natural to direct this excitation along the sliding direction, as has been pro-
posed in Refs. [127, 128], but strong effects are also observed when the excitation is oriented perpendicular to the sliding interface. In micro- and macro-contacts, the amplitude of such excitations is increased to the point where the two surfaces are briefly separated during every oscillation cycle, which leads to arbitrarily large reductions in the lateral (friction) force required to make the system slide, as has been demonstrated in MEMS devices in [129, 130]. On the smaller scale of FFM experiments a more refined possibility exists, that has been nick-named “dynamic superlubricity” [77, 131]. Here, the normal force modulation is kept small enough that the two surfaces remain in permanent contact.

Nevertheless, it is large enough to temporarily lower the loading force significantly during each cycle and, hence, to lower the amplitude of the lateral variation of the tip–surface interaction potential $U_0$. Since this amplitude appears in the exponent of the rate of thermally excited jumps between neighboring potential energy wells, see Eq. (8), the jump rate is periodically enhanced by a dramatic factor. In other words, genuine thermolubricity is switched on briefly during each cycle of the normal force modulation, which brings the effective friction force down to nearly zero. This effect has been demonstrated in an elegant experiment by Socoliuc et al. [132–134]. One should realize that, no matter how carefully set up, every FFM experiment displays some level of excitation by external mechanical vibrations. Labuda et al. [135] have conducted a careful analysis of this effect. They have demonstrated that it not only lowers the measured friction force, but also that the typical reduction of the damping of the instrument vibrations at low temperatures can even lead to a non-monotonous temperature dependence of the observed friction.

5 Rapid dynamics of nanocontacts and new mechanisms of dissipation In the previous sections, we have avoided an important question. We know that thermal activation is a natural phenomenon for the motion of atomically small objects, e.g., for surface diffusion of atoms [136, 137] and atomic clusters [85, 86]. However, in the thermally activated Prandtl–Tomlinson model, we have assumed thermal motion to be of serious importance for an object with a macroscopically large mass, namely the combination of a large part of the cantilever and the tip. Of course, a large mass should be associated with a low “attempt frequency”, i.e., a low pre-exponential factor $r_0$ (see expression (8)). In fact, the $r_0$-values extracted from the analysis of different experiments scatter by orders of magnitude, from the kHz to the MHz range [80, 106], sometimes being as low as the characteristic cantilever frequency, and hence in seeming correspondence with the prediction of transition state theory, but often much larger. In principle, the prefactor could be high, due to an entropic effect concerned with the number of atoms in the contact [86], but for a three-dimensional object (the tip) interacting with a solid surface this effect should be weak. Rather, the high prefactor values suggest the presence of some rapid modes of the system motion. Further indications for some hidden type of motion can be recognized in the nontrivial slip dynamics (e.g., slips with intermediate state) observed in the high-resolution experiment of [91].

Incorporating an extra type of motion, requires us to go beyond the traditional single-spring Prandtl–Tomlinson model. Such excursions have been made by several authors [62, 91, 138–140] and a systematic investigation has been conducted by the authors of this article [79, 141–144]. As will be discussed below, the flexibility of the tip apex brings in an additional – rapid – mode of the system motion, which leads to the appearance of new friction regimes and a reinterpretation of the familiar ones.

5.1 Flexibility and effective mass of the tip apex At first glance, the silicon, silicon nitride or metal tips, typically used in FFM experiments, seem very rigid objects. However, their spring constants are usually much smaller than the stiffness $K$ of the cantilever. From a measurement of the restoring lateral forces at small, sub-atomic support displacements, the effective spring coefficient can be determined of the FFM instrument, which is usually in the order of a few $N m^{-1}$. We write it as $k_{	ext{eff}} = (K^{-1} + k^{-1})^{-1}$, to emphasize that it is low due to a flexible element in series with the cantilever. This element has a low stiffness $k$, again in the order of a few $N m^{-1}$, i.e., of the order of the stiffness of atomic bonds. There is no other way than to associate $k$ with the flexibility of the tip, which hence turns out to be even softer than most cantilevers [56, 57, 69, 102]. The change from $K$ to $k_{	ext{eff}}$ has long been believed to be the only consequence of this inherent, local compliance.

Tip flexibility can introduce a wealth of new dynamics in the FFM. A straightforward way to capture this is via a two-mass-two-spring model (Fig. 18), with one mass ($M$) accounting for the combined cantilever + tip inertia, and the other – effective – mass ($m$) associated with the bending motion of the tip. Let us try to estimate the effective mass $m$.

It is straightforward to estimate [141] that the bending deformation of an atomically sharp, conical, or pyramidal tip is associated with at most a few hundred atomic layers at its apex, so that the effective mass $m \sim 10^{-20}$ kg. The typical mass $M$ of the tip–cantilever combination is 9–12 orders of magnitude higher. If we combine the small effective mass with the typical value for the associated spring coefficient $k$, we find that the characteristic frequency of the tip apex bending vibration $(\nu_{\text{t}})$ should be in the order of several GHz, while the characteristic cantilever frequencies $(\nu_{\text{c}})$ fall in the kHz to MHz range. The extreme hierarchy between the effective masses and frequencies can have dramatic consequences. First, in FFM experiments one always measures the low-frequency response of the cantilever, which can be very different from the rapid motion, performed by the ultra-low effective mass that is actually probing the surface. Second, thermal effects can be much stronger than in a single-spring system.

5.2 The FFM as a two-mass-two-spring system A full numerical description of the two-mass-two-spring sys-
able potential wells, over almost the full range of support positions (Fig. 19d), due to its rapid, thermally activated motion between the accessible surface wells. This regime, that we refer to as “stuck-in-slipperiness” [141], is explained as follows. If the rate \( r \) of thermally activated tip apex jumps between neighboring potential wells strongly exceeds the characteristic frequency \( v_c \) of the cantilever, as is typical for \( m \ll M \), the cantilever cannot follow the rapid thermal apex motion. Instead, it experiences the effective “surface” interaction that remains after averaging over this rapid motion. This effective potential still exhibits a corrugation with the period of the substrate lattice. If the cantilever is sufficiently soft, it advances in this effective potential by “ordinary” stick–slip motion. Obviously, the shape and amplitude of the effective potential differ significantly from the originating tip–substrate interaction potential, which would disqualify the analysis of the observed force variations with the traditional Prandtl–Tomlinson model.

In spite of the extremely chaotic motion of the tip apex, the stick–slip pattern of the cantilever in Fig. 19c is extremely regular, in contrast with the irregularity mentioned already for the stick–slip motion in Fig. 19a. This is because the effective potential, obtained by averaging over the rapid apex motion between the available wells, is completely regular and reflects the substrate periodicity (Fig. 19d). The stick–slip motion of the cantilever in this regular potential is not affected by thermal excitations, as a direct consequence of the macroscopic mass of the cantilever and its correspondingly low “attempt” frequency.

### 5.3 Experimental evidence for rapid apex dynamics

The natural timescale of the tip apex motion is much shorter than the best time resolution of present-day FFM instruments. In addition, the slow response of the cantilever seems to completely obscure the rapid motion of the tip apex. At first sight, these limitations may seem to keep the tip apex motion completely hidden in FFM experiments. Fortunately, the two-mass-two-spring calculations show [142] that there should still be a signature of the apex motion in “fine structure” in the slip dynamics of the cantilever that can be accessed when the time resolution of the FFM is made good enough, e.g., 1 \( \mu s \), to follow the cantilever in full detail.

To our knowledge, Maier et al. [91] have performed the only FFM experiment, in which a slipping cantilever has been followed with sufficient time resolution. The measurements show amongst others (i) cases in which the cantilever rapidly slips over a single atomic spacing, (ii) cases in which the cantilever performs occasional backward jumps, within a short time after a forward jump, and (iii) cases in which the cantilever appears to make a continuous and very slow transition from one lattice position to the next. Maier et al. [91] have invoked a special configuration of simultaneous contact via several “nanotips” in a commensurate geometry with the substrate lattice in order to explain their results. In contrast, we regard these unique experimental findings as an intimate view on the underlying tip apex dynamics. Strong potential corrugations, for which the cantilever follows the...
simple stick–slip motion of the tip apex, correspond to trajectories of type (i). When the corrugation is somewhat lower and the apex performs forward and reverse jumps with a low frequency, the cantilever can follow this stochastic, thermally induced behavior, leading to observations of type (ii). At low corrugations, when the tip apex executes a high frequency of forward and reverse jumps, the cantilever follows the slowly moving average apex position, resulting in the peculiar, ultra-slow slipping of type (iii).

An excellent match can be obtained between two-mass-two-spring calculations [142] and the experimental observations [91], when the effective mass of the tip apex is chosen as \( m < 10^{-18} \) kg, in support of the above estimate of \( 10^{-20} \) kg [141] and in contrast with earlier expectations [91, 138].

5.4 Extra friction regimes

The two-mass-two-spring system reveals rich dynamics that can be categorized in a multitude of sliding regimes, each corresponding to a different scenario of frictional energy dissipation. The determining factors are the corrugation of the interaction potential, the temperature, and various other quantities [143, 144], which can be combined into four dimensionless system parameters. We have already encountered two of these, namely the familiar friction parameter \( \gamma \) (7), and the parameter \( \beta \) (11), in this case referring to the relative rate of thermally activated jumps of the tip apex. A complete description can be found in Ref. [143].

Figure 20 illustrates how the different friction regimes are traversed as a function of the corrugation of the interaction potential. These calculations were again performed for parameter values, typical for FFM experiments, both for a low and for a high cantilever stiffness, each mimicking an actual FFM experiment. Detailed dynamics of both the tip apex and the cantilever for each regime can be found in figures 4 and 5 of [143]. We see from Fig. 20 that between the
familiar stick–slip (SS) regime at high potential corrugations and the superlubricity (SL) regime at the lowest corrugations, there is a wide range of corrugations where friction is heavily suppressed by thermal activation. This range extends to corrugations as high as 0.6 eV, i.e., well above the thermal energy $k_B T$. This thermolubricity effect is much more pronounced than that within the single-spring model, because of the high “attempt” frequency, introduced by the tip apex motion.

Depending on the cantilever stiffness, two types of strong thermal suppression of friction occur, namely true thermolubricity (TL) for hard cantilevers and the “stuck-in-slipperiness” motion (SinS), described above for cantilevers that are soft enough to experience mechanical instabilities in the effective potential energy landscape, obtained after averaging over the rapid motion of tip apex.

At the transition between thermolubricity and stick–slip motion, where the tip apex makes a limited number of thermally activated jumps, back and forth, during the time $(a/V)$ required for travelling one lattice spacing, the system exhibits a variety of stochastic types of behavior. This is where we find the stochastic stick–slip (SSS) motion of the cantilever, encountered already in the previous section, with structured slips and frequent backward jumps, as was observed in Ref. [91].

One further regime appears when the corrugation is above but close to the level for superlubricity, which we refer to as the passive apex (PA) regime. At these corrugations, the tip apex cannot slip by itself, but dynamical instabilities in the combined, two-mass system can take place. In this case, the system will exhibit stick–slip motion, in which the apex moves together with the cantilever on the timescale characteristic for the combined mass. This is the only regime where the single-spring description of the original Prandtl–Tomlinson model would be truly appropriate.

### 5.5 A critical view on FFM results

What is evident from the discussion in the previous sections is that many FFM measurements may require a more refined interpretation than what is suggested by their seemingly uncomplicated appearance. This can be difficult to recognize, because the slow cantilever hides most information about the rapid dynamics of the tip apex and it still often performs what appears to be straightforward stick–slip or continuous-sliding motion.

Perhaps most relevant to the theme of this article is that a re-analysis [79] of the two experiments with near-zero friction [57, 69], discussed in Section 3.6, has shown, that both of them were really not in the superlubricity regime of $\gamma < 1$. Instead, each had reached one of the two forms of extreme thermal suppression of friction, namely the “stuck-in-slipperiness” mode in second case and true thermolubricity in the first.

A related word of caution is that the familiar single-spring model fatally underestimates the corrugation of the tip–substrate interaction potential. It is customary (see, e.g., [57]) to use the Prandtl–Tomlinson model to estimate $U_0$ on the basis of the observed maximum $F_{\text{max}}$ in the lateral forces in the stick–slip regime, $U_0 = a F_{\text{max}} / \pi$. Even though one may recognize that due to thermal activation (precritical slips) the actual corrugation should be somewhat higher, the actual error can easily be as large as a factor 3 [79]! Under-estimating an energy barrier by a factor 3 or more is really dramatic since this leads to an underestimate of the thermally activated rates by many orders of magnitude.

In general, we can conclude that thermal activation plays a much more pronounced role in friction than has been anticipated. If we are allowed to view the FFM tip as an appropriate model for each of the asperities that constitute the contact between macroscopic sliding bodies, we might even speculate that thermally assisted dynamics could also play a dominant role on the macroscopic scale. This might provide interesting opportunities for low-friction applications.

### 6 Concluding remarks

Atomic-scale lateral-force measurements have become routine on a wide variety of materials and under an expanding range of conditions. In parallel, simple models and straightforward, atomistic computer simulations appear to capture the essence of the experimental observations. Nevertheless, we are only beginning to unravel the intricate details of atomic-scale energy dissipation and its connection to friction. Such insight will prove necessary to ultimately reach full predictability of friction forces of natural and man-made interfaces.
At present, we have to admit that our understanding of atomic-scale friction remains at a semi-phenomenological level. In most cases, the discussion of possible microscopic dissipation mechanisms is decoupled from experiment. Formal reference may be made to possible phononic or electronic mechanisms, but the questions that are key to friction, namely (i) how, where and at what timescale the energy and momentum losses become irretrievable and (ii) how, where and at what timescale the transition takes place of mechanical energy to heat, have not been addressed seriously. In fact, it is not generally recognized that these really are separate issues.

We have already touched upon the necessity to assume damping to be critical (Section 3.3), in order to consistently describe typical, atomically resolved FFM patterns. As we have recently found [145], this peculiar element is present not only in the context of the oversimplified, single-spring Prandtl–Tomlinson model, but also in the more realistic two-mass-two-spring model, in which the rapid motion of the tip apex appears to be close to critically damped, while damping on the slow motion of the cantilever is weak. These observations may provide us with a stepping stone toward a genuine, microscopic theory.

Related to the previous two points, friction is a nonequilibrium statistical phenomenon. The non-equilibrium character is frequently implied by the use of Langevin-type equations in theoretical calculations or computer simulations. But a serious justification of the inherent approximations remains to be given, e.g., whether, when and why memory effects and noise correlations can be neglected.

Also on the experimental side, much remains to be done, as many of the theoretical predictions still remain to be put to the experimental test. For example, at present we cannot routinely tailor FFM tips, in order to control the precise contact area and the precise mechanical properties of the very apex of the tip. Once we have such parameters under full control, we can explore and manipulate the tip–surface interaction potential and the rapid tip apex dynamics in a fully quantitative manner. Even in supposedly dry-friction experiments, the cleanliness of the interface is often not controlled sufficiently well. As we have seen, a sub-monolayer density of adsorbed atoms or molecules can already ruin superlubricity and dominate friction. In view of the confusing combination of observations that have been reported so far, many more FFM measurements are necessary of the temperature and velocity dependence of friction, in order to identify the rules and exceptions over a wide range of material combinations and conditions.

In order to support the quest for the fundamental aspects of friction and energy dissipation with compelling experiments, new approaches will be required, for example, enabling the direct exploration of the microscopic time domain, covering the entire range from the femtosecond timescale characteristic for electronic excitations, via the vibrational timescales of picoseconds, to the nanosecond and microsecond timescales associated with, e.g., phonon lifetimes and thermalization. “Seeing” and quantifying the signature of the atomic lattice in the lateral force patterns may be inspiring, but it is not enough!

We have restricted this article to only a single part of nanotribology, namely atomic-scale, dry friction. Dry friction in ensembles of larger contacts, e.g., on a μm-scale, cannot be regarded as a trivially scaled-up version of the nano-scale friction, as new behavior takes over on the larger scale, such as plasticity and the dominating role of lattice defects and material inhomogeneities. Also nano-lubrication, i.e., friction in the presence of thin interfacial films of lubricants, introduces a rich collection of extra phenomena related to the spatial organization of the lubricant molecules in the confinement between the surfaces and their motion under shear. Also these practically very relevant subjects are increasingly enjoying attention from a fundamental physics perspective.

Acknowledgements The authors gratefully acknowledge support from the European Research Council through an ERC Advanced Grant and from the Netherlands Foundation for Fundamental Research on Matter (FOM), through the FOM-Program on Fundamental Aspects of Friction.

References
[1] D. Dowson, History of Tribology, second ed. (Longman, London, 1998).
[2] B. N. J. Persson, Sliding Friction: Physical Principles and Applications (Springer, Berlin, 1998).
[3] R. W. Carpick and M. Salmeron, Chem. Rev. 97, 1163 (1997).
[4] M. H. Müser, M. Urbakh, and M. O. Robbins, Adv. Chem. Phys. 126, 187 (2003).
[5] M. O Robbins and M. H. Müser, in: Modern Tribology Handbook, edited by B. Bhushan (CRC Press, Boca Raton, FL, 2001).
[6] J. Krim, Phys. World 18(2), 31–34 (2005).
[7] J. Krim, Adv. Phys. 61, 155 (2012).
[8] O. M Braun and A. G. Naumovets, Surf. Sci. Rep. 60, 79 (2006).
[9] M. Urbakh, J. Klafter, D. Gourdon, and J. Israelachvili, Nature 430, 525 (2004).
[10] G. V. Dedkov, Phys.-Usp. 43, 541 (2000).
[11] M. Hirano, Surf. Sci. Rep. 60, 159 (2006).
[12] E. Gnecco, R. Bennewitz, T. Gyalog, and E. Meyer, J. Phys.: Condens. Matter 13, R619 (2001).
[13] C. Mate, G. McClelland, R. Erlandsson, and S. Chiang, Phys. Rev. Lett. 59, 1942 (1987).
[14] L. Prandtl, Z. Angew. Math. Mech. 8, 85 (1928).
[15] G. A. Tomlinson, Philos. Mag. 7, 905 (1929).
[16] F. P Bowden and D. Tabor, Proc. R. Soc. Lond. A 169, 391 (1939).
[17] T. Baumberger, P. Berthoud, and C. Caroli, Phys. Rev. B 60, 3928 (1999).
[18] R. P. Feynman, Lectures on Physics I, sixth ed. (Addison-Wesley, Reading, MA, 1977).
[19] M. Dienwiebel, E. De Kuyper, L. Crama, J. W. M. Frenken, J. A. Heimberg, D. J. Spaanderman, D. G. van Loon, T. Zijlstra, and E. van der Drift, Rev. Sci. Instrum. 76, 043704 (2005).
