Recent advances in single-asperity nanotribology

This content has been downloaded from IOPscience. Please scroll down to see the full text.
(http://iopscience.iop.org/0022-3727/41/12/123001)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 158.130.107.37
This content was downloaded on 24/01/2014 at 21:28

Please note that terms and conditions apply.
TOPICAL REVIEW

Recent advances in single-asperity nanotribology

Izabela Szlufarska\textsuperscript{1}, Michael Chandross\textsuperscript{2} and Robert W Carpick\textsuperscript{3}

\textsuperscript{1} Department of Materials Science and Engineering, University of Wisconsin, Madison, WI 53706, USA
\textsuperscript{2} Sandia National Laboratories, Albuquerque, NM 87123-1411, USA
\textsuperscript{3} Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 23 November 2007, in final form 23 March 2008
Published 30 May 2008
Online at stacks.iop.org/JPhysD/41/123001

Abstract

As the size of electronic and mechanical devices shrinks to the nanometre regime, performance begins to be dominated by surface forces. For example, friction, wear and adhesion are known to be central challenges in the design of reliable micro- and nano-electromechanical systems (MEMS/NEMS). Because of the complexity of the physical and chemical mechanisms underlying atomic-level tribology, it is still not possible to accurately and reliably predict the response when two surfaces come into contact at the nanoscale. Fundamental scientific studies are the means by which these insights may be gained. We review recent advances in the experimental, theoretical and computational studies of nanotribology. In particular, we focus on the latest developments in atomic force microscopy and molecular dynamics simulations and their application to the study of single-asperity contact.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Nanotribology is the study of friction, adhesion, lubrication and wear at contacts of nanometre sizes. It is a fascinating field where physicists, chemists, mechanical engineers, materials scientists, biologists and others meet to unravel fundamental phenomena responsible for the production of forces and the dissipation of energy when two surfaces are in sliding contact. It is not only the basic science, but also the relevance to engineering applications that is fuelling the excitement surrounding nanotribology. In miniaturized devices, such as micro- and nano-electromechanical systems (MEMS/NEMS), surface forces and surface phenomena become dominant and they can enable or hinder functionality of a nanodevice. For example, silicon-based MEMS/NEMS exhibit undesirable stiction and high wear, which can render the devices completely non-functional \cite{1–4}. In biological systems, molecular interactions at interfaces control much of the function of the system as a whole, and atomistic-level insights into the mechanical interactions are important \cite{5,6}.

Macroscopic tribology often focuses on determining the friction coefficient and wear rate for the materials of interest. Neither the friction coefficient nor the wear rate is an intrinsic physical property, as both can be strongly dependent on the specific structure, chemistry and elastic/plastic properties of the surfaces, on the chemical environment in which the measurements are performed, and on the sliding history of the interface. These properties can also depend on the mechanics of the instrument itself that is used to make the measurement. Because of the complex nature of nanotribology, fundamental understanding requires experiments at well-defined interfaces. Therefore, single-asperity contact measurements have been a very useful tool in such studies. In particular, experiments with the scanning force microscope (SFM) provided well-defined interfaces for tribological studies \cite{7,8}. The materials and conditions studied continue to broaden to this day. The last two decades of developments in scanning force microscopy enabled measurements of forces in the sub-nanoNewton regime, and led to characterization of tribological properties of nanometre-scale contacts in various environments for a wide range of materials. Much progress in nanotribology has been also accomplished thanks to advances in computational methods and in computer hardware. Atomistic simulations based on the molecular dynamics (MD) technique have been
used to model single-asperity contacts as well as their dynamics during sliding.

Scientific progress in single-asperity nanotribology accomplished with the aid of SFM and MD techniques is the main focus of this paper. The state of the art in these experimental and simulation approaches is discussed in detail in section 2. In section 3 we review continuum mechanics of a single-asperity contact, we discuss its limitations and in doing so we demonstrate the necessity to study physical phenomena underlying friction. The current state of knowledge in the physics of friction is expanded on in section 4, where we review the atomic stick–slip phenomenon, superlubricity, the relationship between friction and other mechanical properties, the effects of adhesion and surface chemistry on friction, and the dependence of friction on temperature and scanning velocity. One specific method to design surfaces with reduced friction involves coating surfaces with monolayer lubricants. This topic is the subject of section 5. Section 6 deals with wear phenomena at the atomic scale. Finally, in section 7, we summarize remaining questions and future prospects in the field of single-asperity nanotribology. Excellent discussions of other techniques that have been employed to study friction and wear beyond a single-asperity contact, e.g. surface force apparatus (SFA), finite element simulations and analytical theories can be found elsewhere [9–11].

2. Experimental and simulation approaches to single-asperity nanotribology

Studying the physical origins of tribological phenomena requires examining well-defined interfaces. Thus, experiments and models often focus on single-asperity contacts, i.e. where there is a single, continuous contact area. This avoids ambiguities that can occur due to interactions of multiple asperities. This also facilitates the execution of meaningful comparisons between experiment, theory and simulation, since the tribological behaviour of an interface will depend not just on the material composition, but also on the contact geometry. Furthermore, if the behaviour of individual asperities can be understood, then models that account for surface roughness can make use of the single-asperity behaviour as a key ingredient to predicting the tribological behaviour of more complex, multi-asperity interfaces.

2.1. Scanning force microscopy

The atomic force microscope (AFM) is the most widely used tool for nanoscale single-asperity studies. It is used for studying tribology at a fundamental level because it provides a controllable, single-asperity contact between the tip and the sample, where forces and displacements can be measured with atomic-level precision and accuracy, and environmental conditions can be controlled over a wide range.

The general design of the AFM has been reviewed many times [12–15] and so we will only briefly summarize the essential components here. In particular, we will point out those aspects that are critical for single-asperity nanotribology studies, including some of its current limitations.

In the AFM, a sharp tip, with a radius typically between 10 and 100 nm, is integrated with a compliant cantilever near its free end (see figure 1). After the tip has been brought in close proximity to the sample’s surface, forces between the tip and the sample result in deflections of the cantilever. The cantilever bends vertically (i.e. towards or away from the sample) in response to attractive and/or repulsive forces acting on the tip. This vertical deflection of the cantilever from its equilibrium position is proportional to the normal load applied to the tip by the cantilever. By equilibrium, this will be equal to the forces applied to the tip by the sample. Lateral forces result in a twisting of the cantilever from its equilibrium position. The cantilever is slightly tilted to ensure that only the tip makes contact with the sample, and not other parts of the lever or the chip holder it is attached to. This tilt can in fact present complications for the measurements, and opportunities as well, as discussed in detail elsewhere [16–21]. AFM measurements can be performed in a variety of environments: ambient air, controlled atmosphere, liquids [22] or ultrahigh vacuum (UHV) [23–25].

Since the tip is attached to a compliant element (the cantilever), the technique is referred to as being load-controlled, which means that the load can be prescribed, but the actual displacement of the tip with respect to the sample cannot. The key manifestation of the load control is that the cantilever can exhibit snap-in and snap-out instabilities. Consequently, certain ranges of tip–sample separations cannot be stably accessed in these measurements.

In contrast, displacement-controlled techniques avoid such instabilities by effectively eliminating the compliance of the spring or holder. These techniques are already commonly used in macroscopic mechanical testing. Over the past 16 years developments have been made to incorporate displacement control into scanning probes, where control is achieved by displacing the tip through direct application of a force and using feedback to stabilize the displacement. Houston and coworkers [26,27] control the force electrostatically and refer to the instrument as an interfacial force microscope (IFM). Pethica and coworkers [28,29] use a magnetic coating on
the cantilever and external coils to apply forces to the tip. They refer to the instrument as a force-controlled microscope. Lieber and coworkers [30] use a variation on Pethica’s method, where a magnetic coil is used to apply a force to the cantilever, and the instrument has been adapted to work in solution. For generality, we will use the term scanning force microscopy (SFM) to refer to both load-controlled (AFM) and displacement-controlled (e.g. IFM) techniques.

Several SFM-based studies have reported that friction for various solid-solid nanocontacts, below the threshold of any observable wear, is proportional to the true contact area (i.e. number of interfacial atoms) [14, 31–36]. In other words, the friction force \( F_f \) for a single-asperity contact is given by

\[
F_f = \tau \cdot A,
\]

where \( A \) is the interfacial contact area, and \( \tau \) is the interfacial shear strength. Thus, \( \tau \) represents the frictional force per interfacial atom. The contact area \( A \) typically does not vary linearly with load; for example, in the classical theory of Hertz [37], the contact area is proportional to the load \( P \) raised to the 2/3 power (\( P^{2/3} \)). Thus, these observations are in stark contrast to the macroscopic observation of a friction coefficient, i.e. that friction is linearly proportional to applied load. This behaviour is referred to as interfacial friction, whereby the resistance to sliding is due to a pure shear resistance at the intimate contact interface between the two surfaces and wear is not occurring. Interfacial friction has been observed in both macroscopic contacts [38,39] and in SFA measurements [40–42].

The shear strength may be a constant, or it may have a dependence on the contact pressure, \( p \), or applied shear stress [43]. A linear dependence on contact pressure (\( \tau \propto P \)) combined with equation (1) is easily shown to result in a linear term connecting friction and load (\( F_f \propto P \)), thus matching the familiar macroscopic result, but now for a single-asperity contact. What remains unresolved is how the interatomic forces determine the value of \( \tau \) itself.

Another important interfacial parameter which can be measured is the interfacial work of adhesion (adhesion energy per interfacial atom), given by \( \gamma = \gamma_1 + \gamma_2 = \gamma_{12} \), where \( \gamma_1 \) and \( \gamma_2 \) are the tip and sample surface energies and \( \gamma_{12} \) the interfacial energy [44]. \( \gamma \) encompasses all interfacial forces, and is the work per unit area required to separate the surfaces from contact to infinity. If the tip is ‘round’, i.e. paraboloidal, and makes contact with a flat elastic surface, the behaviour spans a spectrum from the Johnson–Kendall–Roberts (JKR) model [45] (for large tips and compliant materials with strong, short range adhesion) to the Derjaguin–Müller–Toporov (DMT) model [46] (for small tips and stiff materials with weak, long-range adhesion). In a SFM experiment, \( \gamma \) is determined from the minimum (most tensile) force that occurs between the tip and sample. For a load-controlled AFM, as discussed above, the snap-out instability will occur very close to this point. Thus, this force is often referred to as the pull-off force or critical load, \( P_c \). For a tip of radius \( R \) this is given by

\[
\gamma = -\frac{P_c}{\chi \pi R^3},
\]

where \( \chi \) is a parameter that ranges monotonically from 1.5 (JKR) to 2 (DMT).

In many studies, continuum mechanics appears to provide an accurate description of the nanometre-scale contact area \( A \) and other contact properties [14, 31, 35, 47–49]. Key assumptions here are homogeneity, isotropy, linearity, and elasticity of the materials. Several modified continuum contact mechanical models describing other cases have been derived [50,51]. However, more fundamentally, Robbins and coworkers [52,53] have recently explored nanoscale contacts using atomistic simulations, and found cases where continuum mechanics breaks down altogether (see section 3.2). This presents a forefront challenge for properly analysing the contact properties in the atomistic limit.

Others have challenged the notion that equation (1) should apply at all. Using atomistic simulations to model contact between a curved elastic tip and a flat surface, Wenning and Müser [54] predict \( F_f \propto P \) for commensurate interfaces, but they also predict \( F_f \propto P^{2/3} \) for amorphous interfaces, in both cases assuming dry, i.e. unlubricated and uncontaminated conditions. They argue that friction is not being determined by the contact area itself. Rather, the frictional resistance experienced by each interfacial atom varies with the local normal stress it experiences, which will depend on the atomic structure of the interface. The net effect, integrating over the interface, thus produces a distinct power law dependence for different interfacial atomic arrangements. This work is discussed in more detail in section 3.2.2.

The process of wear, which is of critical importance for many applications, is more difficult to quantify in terms of fundamental physical parameters. So far, quantification of the loads and stresses to initiate wear, and characterization of changes inasperity shape, represent the extent to which SFM experiments have progressed. MD simulations that model bond breaking and atomic rearrangement are able to provide valuable insight into the nature of this process.

There are several other key challenges of SFM experiments. Experimental calibration of the normal and lateral forces is required for each cantilever, since force constants between nominally identical cantilevers can vary substantially. Several in situ experimental methods for performing cantilever force calibration have been demonstrated. These techniques can be readily implemented, but they require additional time and care from the experimentalist [21,55–62].

Additionally, control and characterization of the tip’s composition (i.e. at its surface, which can change due to contamination), and shape (which can change due to wear) are important if quantitative, physical insights into tribological mechanisms are to be derived since the tip represents half of the interface. Wearing and blunting of the tip are results of the substantial stresses (on the order of GPa), which are generated in small contact areas (nm²) at typical loads (nN) and which can exceed the strength of the tip. Such wear has indeed been observed in numerous studies [63–68]. Contamination is related to material transfer due to the adhesive nature of the surfaces [69–71]. It is not surprising that when researchers care to look, tip wear and contamination are often observed.

For many years SFM-based nanotribology studies have been primarily performed with silicon and silicon nitride tips because only such probes existed on the market.
However, a much wider range of tip materials are now commercially available, and more probes have been made in individual laboratories for the purpose of nanotribology studies. Experiments have been carried out with tips covered with a variety of coatings [32, 66, 72–78], and tips with attachments such as carbon nanotubes [67, 79] and colloidal spheres [80–83].

The stresses in the contact, particularly the average contact pressure, are critical to understanding tribological behaviour. Unfortunately, there is no direct measurement technique for determining the stress components. Rather, stresses can be inferred from the measured forces and displacements combined with contact mechanics models described above. Determining the contact pressure this way is far from straightforward. It requires knowledge of the tip radius throughout the experiment, calibration of forces, measurement of the work of adhesion, knowledge of the elastic constants of tip and sample materials, as well as trust in a particular model of contact mechanics (the applicability of continuum models to nanoscale contacts is discussed in section 3.2). The substantial challenges in interpreting SFM experiments prevent us from performing meaningful comparisons of the contact pressures reached in studies reported so far. This point illustrates an additional advantage of atomistic simulations, in which pressures and stress components can be quite easily evaluated.

Another challenge in interpreting SFM experiments in nanotribology is that characterization of surfaces in contact, if done at all, is usually carried out post mortem and ex situ, which makes it difficult to correlate the frictional response with dynamic phenomena that occur at the contact during sliding. In recent years, an exciting and important opportunity has been presented by the development of in situ loading stages for electron microscopes. In particular, nanoindentation experiments can now be carried out inside of the scanning electron microscope (SEM) and the transmission electron microscope (TEM) instruments, offering a previously inaccessible, live view of asperities during loading and unloading. Hybrid nanoindentation and TEM techniques have been pioneered by Minor and coworkers [84–88] and they have been reviewed elsewhere [89, 90]. Incorporating scanning probes into TEM has the potential to provide dramatic new insights into the structure of nanocontacts as they form and slide. For example, Rieheiro et al [91] used a diamond indenter coated with Au to slide against the (1 0 0) surface of Si. By using in situ TEM imaging, it was determined that the abrasive wear was initiated by cracks nucleating and propagating along the elastic strain contour formed by the indenter. Merkle and Marks [92] used tungsten probes 2–50 nm in diameter to slide against highly ordered pyrolytic graphite (HOPG). The in situ TEM studies confirmed that graphitic flakes transfer to the tip during sliding. Tribological studies using these techniques are still scarce but because of their great promise for unraveling nanotribological phenomena, they are worthy of further pursuit.

2.2. Atomistic simulations

There are multiple benefits to applying atomistic modelling techniques to understanding both nanotribology in general as well as single-asperity contacts in particular. For example, many of the difficulties encountered in sample preparation do not exist in the world of computer simulation, where virtually any system can be created, ranging from defect free to highly disordered ones. As a result, true experimental conditions would then be extrapolated from the wide range of initial conditions and system preparations covered by simulations. The most significant benefit of atomistic simulations, however, lies in the ability to track the motion of each individual atom. Such a level of detail can provide critical information to help interpret experimental results. We will discuss below a number of simulations of nanoasperity contacts starting from the introduction of SFM until today. While early models were inherently limited to very small systems, they were instrumental in determining the atomic origins of a number of experimental results. Later work with massively parallel computers has expanded the scope of simulations with attempts to directly model single-asperity contacts, resulting in enhanced understanding of, for example, contact mechanics at the tip–sample interface, and the effects of varying tip radius on measured properties (see figure 2 for more realistic representations of AFM tips from recent simulations [93]).

2.2.1. Challenges in atomistic simulations

Atomistic simulations of single-asperity contacts suffer from the same general difficulties that plague many areas of molecular modelling. A major issue for the simulation of any system is the availability of accurate interatomic potentials (also referred to as force fields) to describe the interaction between all the atomic sites in the system. Many reliable force fields exist in the literature. However, they must be chosen with discretion as they are generally parametrized to fit a given set or range of experimental data. For example, a classical force field that accurately predicts bonding structures of a given system may give a poor estimate of phonon frequencies. Furthermore, some MD simulations are performed with models that do
not allow for the breaking or formation of chemical bonds—all bonds are set at the beginning of the simulation and are modelled by a harmonic spring only [94]. This has obvious implications for simulations where one wishes to examine wear. Many groups do use so-called reactive potentials which can do an excellent job of describing chemical reactions for a small number of atomic species [95]. These techniques are much more computationally expensive and further limit the size and duration of simulations.

Even without chemistry, issues with system size and timescale in simulations are common. The size issue, as evidenced below, can be approached by parallelization of the computer code and the use of multiple processors simultaneously. Each processor tracks atoms within a given spatial area and it can exchange information with other processors (e.g. to compute interactions across the spatially divided areas). Consequently, larger systems can be modelled simply by increasing the number of processors. The timescale issue is more problematic because shear velocities in nanotribology simulations are essentially linearly related to processor speed. In order to reproduce the correct dynamics, the time step in atomistic simulations needs to be one to two orders of magnitude smaller than the time scale of the fastest dynamical process in the modelled system, e.g. the vibrational motion of atoms. Therefore the time step is often limited to 1 fs or less for fully atomistic simulations, and can be increased only to around 5 fs for coarse-grained models. Increasing the number of processors does not, in general, enable an increase in simulation velocity. Thus, unlike in the case of spatial extent, this issue cannot be solved through scaling arguments and qualitative comparisons to relate results to experiment. A very recent study by Mishin et al [96] uses a variant of parallel replica dynamics (in which multiple MD simulations are concurrently used on replicas of a single system to increase the likelihood of rare transition events) to study stick–slip motion of sliding grain boundaries in copper. The authors succeeded in simulating extremely low velocities of 500 µm s⁻¹ of Σ13 grain boundaries in Cu by using 1000 parallel simulations of about 10 000 atoms each. This and other accelerated MD methods [97] have the potential to advance the state-of-the-art in simulation by approaching experimental velocities. By construction, applicability of such accelerated MD techniques is limited to systems that exhibit rare-event behaviour (e.g. stick–slip). This and other time-accelerating techniques need to be explored and developed to model more general tribological phenomena. It is also desirable for faster SFM experiments to be conducted, and such efforts are indeed being pursued [98–101].

The limitations on simulations of single-asperity contacts were more severe in the mid-1980s when the AFM was first developed. Even with the limited computational power of that period, a number of modelling studies attempted to understand the details of then-current AFM experiments in both contact and non-contact mode. In the following we will briefly describe early simulations of single-asperity contacts to place nanotribological simulations of tips in the historical context.

2.2.2. Jump-to-contact phenomenon. One particular goal of early SFM simulations was to understand the origin of forces between tips and samples in the atomic scale contacts. It was shown that the attraction between the tip and the sample at small separations results in a jump-to-contact phenomenon [102]. This is to be distinguished from the snap-in that occurs in load-controlled AFM, which is due to the finite compliance of the cantilever. In the first MD simulations two rectangular blocks of fcc (001) crystals with flat surfaces were successively displaced towards one another until a jump occurred. The surfaces were then separated again to observe a hysteresis in the adhesive interaction. Landman et al [103–106] performed similar simulations, but with more realistic models of sharp pyramidal tips (4–100 atoms), and later of pyramidal tips with truncated apex (1400 Ni atoms and an effective radius of curvature of 30 Å). The authors modelled the jump-to-contact effect on Ni and Au samples with indentation velocities in the tens of m/s and they demonstrated the occurrence of material transfer between the sample and the tip. In the same papers, the authors reported how deformations in hard and soft materials are altered by changing tip–substrate interactions, e.g. compression and shear at the interfaces.

These early simulations demonstrated that the jump-to-contact phenomenon can affect the ultimate separation of tip and sample, thus limiting resolution of scanning probe techniques. Furthermore, and perhaps more importantly, these studies pointed out limitations of those continuum-level theories that do not account for the atomic-level attractive forces at finite tip–sample separations, which are in fact observed in experiments [107].

2.2.3. Atomic resolution with AFM. Another early goal of AFM simulations was to understand the atomistic details of the fascinating experimental results in order to both quantify the limits of the technique as well as verify claims in the literature. In particular, the possibility of atomic resolution of surface features with AFM proved to be a fertile ground for early simulation work. For example, Perez et al [108] showed that covalent bonds form across the interface between dangling bonds on the Si tip and substrate, and this bond energy dominates the van der Waals interactions, providing variation in the force gradients. It was concluded that it is the covalent bonding that is responsible for the atomic resolution in AFM experiments. In a different study, Koustos et al [109] modelled a variety of tip sizes from single-atom tips to 31-atom tips and found that atomistic resolution of a vacancy in contact mode could only be detected by an atomically sharp tip. A different conclusion was reached later from joint experimental and MD studies which showed that surface vacancies or adatoms can be resolved in non-contact mode with frequency feedback (known as the frequency modulation or FM technique) both with 34-atom pyramidal Si tips on Si(1 1 1)–7 × 7 [110] as well as 64-atom cube-corner MgO tips on NaCl islands on
Substrates beyond the initial plastic yield [119, 120]. MD simulations of nanoindentation with pyramidal [103, 121], spherical [122–125] and flat punch [126–128] tips brought a qualitative understanding of the load \( P \) versus normal displacement \( h \) curves in AFM and nanoindentation experiments. For example, discrete pop-ins in the \( P-h \) response were correlated with nucleation and propagation of dislocations [126, 129–131], and with structural transformations as in the case of the zinc-blende to rocksalt transition observed in MD simulations of GaAs [132]. Another type of solid-state transformation detected in the vicinity of the indenter is solid-state amorphization [133]. Kallman et al [134] reported such a transition in silicon at temperatures close to the melting point. Szlufarska et al [135] demonstrated that indentation-induced amorphization in zinc-blende silicon carbide takes place by defect-stimulated growth and coalescence of dislocation loops. In simulations by Walsh et al [128] amorphization has been identified as a primary deformation mechanism of silicon nitride. During the latter simulation, amorphization was arrested by cracking at the indenter corners and by piling up of substrate material along the indenter sides.

Pile-up has been a subject of study in a number of other MD simulations. For instance, Smith et al [136] reported joint experimental and large scale MD simulations of nanoindentation of single crystal iron, and showed anisotropic pile-up patterns on Fe(1 1 1), (1 0 0), and (1 1 1) surfaces formed by dislocation cross slip between different planes of the same family. Simulations by Chen et al [137] revealed a similar pile-up anisotropy in indented SiC. The authors showed that this anisotropy is caused by only \( \langle 1 1 1 \rangle \) and \( \langle 1 1 1 \rangle \) planes being active out of the \( \{ 1 1 1 \} \) family (see figure 3).

Formation of a pile-up is one of the mechanisms to dissipate energy when the load on the tip exceeds a critical value and therefore it is an important phenomenon to account for in studies of friction. Pre-stress in the film will affect the amount of pile-up as was shown in the MD simulations of Schall and Brenner [138]. The effect of pre-stress on nanoindentation was also studied experimentally [139] and by means of the finite element method [140] by Pharr and coworkers. It was demonstrated that both compressive and tensile stresses lead to contact areas larger than those typically assumed in the continuum-based elastic half-space models. The finding is particularly important if such continuum models are employed to interpret experimental SFM data on friction.

2.2.6. Monolayer lubricants. While the frictional behaviour of monolayer lubricants has been extensively studied with atomistic simulations, the use of tip-based approaches in such simulations has been limited. The early work of Landman et al [105] involved tip simulations on hexadecane coatings, but it was not until Bonner and Baratoff [141] that tip simulations of self-assembled monolayers (SAMs) were performed. In this latter study, a Au tip consisting of 285 atoms was modelled. The tip was attached to springs to simulate the compliance of the AFM system. The SAM structure was modelled with a united atom model, in which the CH2 and CH3 groups are treated as single, spherical units. Alkanethiols with 11 carbons
in the backbone (C11 chains) were attached rigidly to the Au(1 1 1) substrate and as a result of tip penetration the SAM structure became a host of gauche defects. Under shear, the tilt direction of the SAMs changed to become aligned with the shear direction. A similar effect was shown in later simulations with larger tips [93, 142].

Most of the tribological studies on SAMs have been performed for hydrocarbon chains, referred to as \( C_n \), where \( n \) stands for the number of carbon atoms in the chain. Ohzono and Fujihira [143, 144], who had originally applied a phenomenological model to study friction on SAMs, later employed MD simulations with united-atom force fields to determine changes in C8 and C16 monolayers attached rigidly to a substrate in response to sliding across rigid surfaces (sliders) with atomic-level features of varying shape. A number of differently shaped sliders were used in order to model the apex of an AFM tip. It was shown that incommensurate interfaces can lead to a reduction in friction, in agreement with theoretical predictions of Müser and coworkers [54, 145, 146]. These results also predict that very small tips with only a few atoms in contact can exhibit high friction. Small sliders will more easily form locally commensurate interfaces leading to stick–slip motion. Such local commensurability is more likely to occur at lower temperatures, i.e. when the sliding-induced disordering is suppressed. Leng and Jiang [147] and Zhang et al [148] used a similar united-atom model of the monolayer and larger tips in a hybrid simulation method that attempted to overcome the simulation velocity issue discussed in section 2.2.1. Both of these papers report the response of a monolayer subjected to sliding of a rigid 280-atom Au tip that is attached to lateral springs to account for lateral (shear) compliance of the experimental setup. For simulations performed at 0.1 K [147] and 300 K [148] the authors assume that the relaxation time of the SAM is fast enough so that its motion can be decoupled from the motion of the tip. The hybrid procedure employed in the study is to integrate the equations of motion for the tip sliding at 400 nm s\(^{-1}\) and subsequently to relax the SAM. While the concept of decoupling the motion of the tip and the substrate is intriguing, this method has only been shown to be effective at the extremely low temperatures (0.1 K) reported in the first of the two papers. On the other hand at 300 K, where most experiments and simulations are conducted (including those reported in [148]), the monolayer system will exhibit long time scale relaxation and dissipation, e.g. due to phonons, or recently proposed molecular plowing mechanisms [149], and/or other molecular relaxations. These mechanisms and the dynamic coupling between the motion of the tip and SAM are therefore not accounted for in the described hybrid method.

Very recent work by Chandross et al [93] has attempted to couple more closely with experiment by modelling fully atomistic SAMs in contact with amorphous AFM tips with radii in the tens of nanometres (figure 4). While this work did not use springs to account for the compliance of the AFM cantilever, it has been successful at reproducing a number of experimental results including the linear dependence of friction force \( F_f \) on applied load \( P \). Additionally, this study addressed the effects of the tip radius (ranging from 3 to 30 nm) on adhesion and friction. In friction simulations, an agreement with experiment was found [70], which shows that the tip radius has little effect on friction beyond the change in the pull-off force. Such change leads to a uniform shift of the friction force versus load curve, with larger tips resulting in larger friction forces. On
is the van der Waals radius of a generic atom. It was shown between polymer brushes and tips of radius 10–100\(\sigma\) as well as of Murat and Grest [142] who studied the interactions 2.2.7. Remaining questions. In the twenty years since the process, the system does not exhibit stick–slip motion, as it is difficult for coherent motion to occur. The physical picture that emerges from these simulations is that chains in the monolayer exhibit two kinds of responses to the indentation of the tip. Chains under the tip are deformed in the region at the leading edge of the tip, whereas such motion is more difficult in the case of larger tips, which show a crossover to flat-plate-like compressibility. The effect of adhesion was subsequently modelled by Hertz [37]. While studying interference patterns between glass lenses when pressed together, he determined that contact radius \(a\) for the circular contact between a flat plane and a spherical lens subjected to a normal load \(P\) followed the equation:

\[
a = \left(\frac{PR}{K}\right)^{1/3},
\]

where \(R\) is the sphere radius, \(K = \frac{1}{2} \left( \frac{1-v_1^2}{E_1} + \frac{1-v_2^2}{E_2} \right)^{-1} \), \(E_1, E_2\) are the sphere and flat plane Young’s moduli, and \(v_1, v_2\) are the sphere and flat plane Poisson’s ratios, respectively. This equation describes the area of a smooth macroscopic contact between homogeneous, isotropic, linear elastic materials, with no attractive forces (adhesion) between them. It assumes that the sphere radius \(R\) far exceeds the contact radius \(a\), which allows for the sphere to be approximated as a paraboloid. The effect of adhesion was subsequently modelled by two different groups. JKR [45] proposed a theory in 1971 to account for adhesion between elastic bodies. Using the large shear velocity imposed by short simulation time scales. While some simulations discussed above attempted to deal with this issue through time scale separation, it has not been clearly demonstrated that this approach is effective at finite temperatures. Due to the linear scaling of simulation velocity with processor speed, which is not sufficiently fast to bridge the time scale gap between simulations and experiment, the solution to this problem may necessitate a move beyond brute-force methods. A possible path involves continuum-level simulations with extensive input from atomistic simulations, but such approaches are still in their infancy. The parallel replica method discussed above is another route for overcoming this difficulty and it is applicable for systems where only rare events need to be considered [97]. With focused work from experiments and simulations on well-defined interfaces, this method may help clarify some of the issues pertaining to high simulation velocities.

There is an increasing body of evidence demonstrating changes in the chemistry and morphology of both the tip and the sample occur during sliding [69, 70, 149]. Models encompassing bond breaking and reformation are needed to help develop an accurate picture of the mechanics of the interface. While such models provide crucial information about the chemistry of the interface, increasing the complexity of the force field further limits the scale of achievable simulations. Results from large-scale simulations with less accurate force fields will therefore need to be combined with smaller simulations with more accurate chemistry to arrive at a complete description.

3. Mechanics of single-asperity friction

3.1. Continuum level theories of single-asperity contact

Contact mechanics is critical in tribology, as it provides quantitative descriptions of contact area, elastic indentation, contact stiffness and the stress and strain fields of a mechanically loaded asperity. The field of contact mechanics was pioneered by Hertz [37]. While studying interference patterns between glass lenses when pressed together, he determined that contact radius \(a\) for the circular contact between a flat plane and a spherical lens subjected to a normal load \(P\) followed the equation:

\[
a = \left(\frac{PR}{K}\right)^{1/3},
\]

where \(R\) is the sphere radius, \(K = \frac{1}{2} \left( \frac{1-v_1^2}{E_1} + \frac{1-v_2^2}{E_2} \right)^{-1} \), \(E_1, E_2\) are the sphere and flat plane Young’s moduli, and \(v_1, v_2\) are the sphere and flat plane Poisson’s ratios, respectively. This equation describes the area of a smooth macroscopic contact between homogeneous, isotropic, linear elastic materials, with no attractive forces (adhesion) between them. It assumes that the sphere radius \(R\) far exceeds the contact radius \(a\), which allows for the sphere to be approximated as a paraboloid.

The effect of adhesion was subsequently modelled by two different groups. JKR [45] proposed a theory in 1971 to account for adhesion between elastic bodies. Using
means that the free energy of the system is only reduced by
that the adhesive interactions are infinitely short range. That
contact to infinity. The key assumption of this theory is
the work done separating a unit area of the interface from
of adhesion, which is defined in section 2.1. It represents
critical load
load is often called the pull-off force, adhesive force or the
must be exerted to separate the surfaces. The minimum stable
would occur in that region. The adhesion induces a non-zero
the prediction of an infinite tensile stress at the contact area
of short-range adhesion forces, which is particularly suitable
forces (normalized per unit area) versus separation for the
JKR model is accurate. The DMT model is accurate in the
opposite limit (i.e. stiff materials, small sphere radii and weak,
long-range adhesion forces).

These limits are described quantitatively by a non-
dimensional physical parameter which is often referred to as
Tabor’s parameter \(\mu_T\) [150, 151]:
\[
\mu_T = \left( \frac{16R}{9K^2z_0^3} \right)^{1/3}. \tag{8}
\]
Here \(z_0\) is the equilibrium separation of the surfaces, or in
an atomistic picture, the equilibrium bond length for the two
materials. If we consider an atomistic interaction like the
Lennard–Jones (LJ) potential, then the spatial range of the
attractive force scales directly with \(z_0\). In other words, \(z_0\) is
the only length scale in the definition of the potential. Tabor’s
parameter is equivalent to the ratio between the normal elastic
deformation purely caused by adhesion only (not by the applied
load), and the spatial range of the adhesion forces denoted by
\(z_0\). Figure 6 schematically illustrates the assumed interaction
forces (normalized per unit area) versus separation for the
Hertz, JKR and DMT models, and an intermediate and more
realistic interaction such as the LJ potential [152].

To quantitatively describe these intermediate cases,
Maugis [118] considered a square-well (Dugdale) potential to
describe attractive forces between the two surfaces, which is
schematically shown in figure 7. This potential assumes that a
constant adhesive stress \(\sigma_0\) acts over a separation range \(\delta_1\),
and then drops to zero at larger separations. Thus, the work
of adhesion is \(\gamma = \sigma_0 \cdot \delta_1\). Maugis defines a transition parameter,
\(\lambda\), which is similar to \(\mu_T\), given by
\[
\lambda = 2\sigma_0 \left( \frac{R}{\pi \gamma K^2} \right)^{1/3}. \tag{9}
\]
If \(\sigma_0\) is set to be equal to the minimum adhesive stress
in the LJ potential, it follows that \(\delta_1 = 0.97z_0\), and thus

\[
a = \left( \frac{R}{K} (P + 3\pi \gamma R + \sqrt{6\pi \gamma R P + (3\pi \gamma R)^2}) \right)^{1/3}. \tag{4}
\]
Here, \(\gamma\) is called the work of adhesion or the Dupré energy
of adhesion, which is defined in section 2.1. It represents the
work done separating a unit area of the interface from
contact to infinity. The key assumption of this theory is that
the adhesive interactions are infinitely short range. That
means that the free energy of the system is only reduced by \(\gamma\)
units of energy for each unit of area in intimate contact, and
thus there is no reduction in energy (no attractive interaction)
from regions where the surfaces are at any greater separation.
While unphysical, it represents the asymptotic case in the limit
from regions where the surfaces are at any greater separation.

Separately, Derjaguin, Muller and Toporov (DMT) [46]
derived their own expression to account for adhesion in elastic
contacts. Their key assumption was that the deformed contact
profile remains the same as in the Hertz theory, but with a higher
load overall due to adhesion. In contrast to the JKR model,
the DMT model is equivalent to assuming that the attractive
interaction acts at all separations between the two surfaces.
The expression for contact area then becomes
\[
a = \left( \frac{R}{K} (P + 2\pi \gamma R) \right)^{1/3}. \tag{6}
\]
The form of this equation was first presented in the work of
Maugis [118]. The DMT paper itself focused on the pull-off force, which is given by
\[
P_{c(DMT)} = -2\pi \gamma R. \tag{7}
\]

The DMT model is often referred to as the Hertz-plus-offset
model because of the shift in the load axis by the pull off force.
In the DMT framework, the contact area is equal to zero at
the pull-off force, and there is no predicted singularity in the
contact stresses.

The discrepancy between these two theories was resolved
by the realization that they are both valid, but in completely
opposite limits of contact behaviour. When the induced elastic
deformations in the materials are large compared with the range
of the attractive forces (which occurs for compliant materials,
large sphere radii and strong, short-range adhesion forces), the
JKR model is accurate. The DMT model is accurate in the
opposite limit (i.e. stiff materials, small sphere radii and weak,
long-range adhesion forces).

Figure 6. Schematic illustration of the interaction between AFM tip and a sample as a function of their separation according to (from left to right) Hertz, JKR and DMT models and for the actual model represented here by the Lennard-Jones potential. Reprinted with permission from [152]. Copyright 2005 Koninklijke Brill NV.
\[ \lambda = 1.1570 \mu \text{T}, \text{i.e. they are nearly equivalent. The JKR model applies to cases when } \lambda > 5, \text{ and the DMT model applies when } \lambda < 0.1. \] Intermediate values correspond to the \textit{transition regime} between the JKR and DMT limits. The conventions used for defining this transition parameter are summarized by Greenwood [150]. The Hertz theory obviously corresponds to the case of no attractive surface forces \((\gamma = 0)\). The contact area predictions for these models are plotted as a function of load in figure 8. Relations to calculate the normal displacement in terms of load or contact area are also provided by the Maugis–Dugdale theory.

While the assumption of a square-well potential may seem overly simplistic, the predicted contact behaviour is relatively insensitive to the details of the shape of the potential, as long as only one length scale is involved [153]. However, if an additional length scale of the interaction is introduced, the behaviour can vary significantly [154]. A multiple length scale description may be appropriate when studying simultaneous effects of short-range solid–solid adhesion and longer range electrostatic, van der Waals or solvation forces.

The Maugis–Dugdale equations lack a single expression relating only \(a\) and \(P\), and so they are somewhat challenging to use. A key practical concern is that the value of the pull-off force in terms of \(R\) and \(\gamma\) must be determined through iteration [155] if \(\lambda\) is not known (which is the case with any experimental measurement). An approximate general equation for easily describing the contact area was provided by Carpick, Ogletree and Salmeron (COS) and is described in detail elsewhere [156]. In a similar vein, Piétrement and Troyon subsequently presented approximate general equations for the normal displacement and the normal contact stiffness [157]. Schwarz then showed that these models could be derived from physically based arguments by combining an infinitely short-range attractive interaction (essentially a JKR-type interaction) with a long-range force [158] of adjustable relative weight.

The models described above are all defined for a contact between a sphere and a plane (or, equivalently, two spheres) where the contact radius is much smaller than the sphere radius (allowing it to be described mathematically as a paraboloid), and the loading is purely in the normal direction. The materials are assumed to be homogeneous, isotropic, linear and elastic. However, deviations from these assumptions may be significant in practice. Solutions for many such cases have been derived. Maugis solved the adhesive contact problem for large contact radii (i.e. using spheres, not paraboloids) [159]. The solution for adhesive contact for an axisymmetric power-law indenter shape was presented by Carpick \textit{et al} [31, 160] and recently by Borodich [161]. Johnson discussed the effect that lateral forces may have on the contact [43]. Viscoelastic effects have been discussed in a number of cases [162–167] although not in as much detail.

Models accounting for the presence of a thin layer on either of the two surfaces have also been considered. These are important as many systems including MEMS, hard disks, and other devices of interest involve ultrathin coatings or other layers such as native oxides. A model for a thin stiff layer on a compliant substrate in the presence of adhesion was presented by Johnson and Sridhar [173]. They found that the JKR model works well for a sufficiently thick or sufficient thin layer, but intermediate thicknesses violate the conditions of the JKR theory if the contact radius is small compared with the layer thickness. The deviations are evaluated numerically. A separate model accounting for thin compliant coatings developed recently by Reedy, referred to as Thin Coating Contact Mechanics (TCCM), is more generally applicable [174]. For cases where the ratio of the coating thickness to the radius of the indenter is less than 0.1, the Poisson ratio of the coating is less than 0.45, and the strain is less than 0.2,
this work found that contact area varies with the square root of the applied compressive load. The model treated values outside these ranges through dimensional analysis, where the dependence of contact area upon the load, tip radius, work of adhesion, and elastic properties of the substrate, tip and coating were reduced to a set of analytical power-law relations with unspecified exponents. The exponents were then determined by fitting the analytical relations to the results of a series of precise finite-element analysis (FEA) simulations of adhesive contact. Consequently, this work provided a flexible and FEA-validated set of analytical relations that are relatively easy to use for fitting to data. Examples will be discussed in section 3.2.

Continuum mechanics models have been also applied to study plastic loading characteristics of a single-asperity contact. For instance Etsion et al [175] employed FEA to determine the elastic–plastic behaviour of a sphere loaded on a rigid smooth flat surface. The authors introduced a so-called elastic—plastic loading (EPL) index, which is a parameter that quantifies the amount of plastic deformation in the sphere. It was shown that for high values of the EPL index, a secondary plastic flow may occur in the sphere during unloading.

In summary, contact mechanics models can be powerful tools for extracting fundamental parameters from studies of materials in contact. However, they must be used cautiously, with attention paid to the appropriate limits and assumptions they entail. Key challenges that remain include developing reliable and accessible descriptions of adhesive contacts for arbitrary geometries in the presence of shear stresses and plastic deformation, and further considering the limits of applicability of the continuum models themselves.

3.2. Breakdown of continuum level contact theory

3.2.1. Experimental evidence of issues. Despite the remarkable successes of continuum models, they are also plagued by some serious limitations. At the nanometre length scale, where the discreteness of atoms often has a direct effect on physical properties, there is no a priori reason to believe that continuum level models will be capable of reproducing the tribological behaviour. In fact, a number of atomic-scale phenomena have been observed during sliding that are not accounted for in any continuum theories. For instance, Socoliuc et al [176] found that the lateral contact stiffness at the lowest accessible loads for AFM tips with radii nominally below 15 nm cannot be accounted for by a continuum approach, as it was found to be nearly load independent. This may suggest that a fixed number of atoms are interacting across the interface, and that this contact size does not change with the load. Other tribological phenomena that require atomic level analysis include atomic stick–slip, non-linear elastic deformations of the interface, dependence of friction on atomic-scale surface inhomogeneities, and the contributions of plasticity to energy dissipation during sliding, e.g. dislocation nucleation and motion, or the creation or diffusion of vacancies, interstitials, impurities, and electric and phononic dissipation.

Semi-continuum models have been developed to address some of these problems. The Tomlinson model [177] uses arguments based on the one-dimensional, zero-temperature energy landscape to explain atomic stick–slip behaviour. This model has been recently extended to include thermal activation of slip events at the interface [15, 178], and this approach is now consistent with some of the experimental findings such as the velocity and load dependence of friction. Stick–slip phenomenon and the Tomlinson model are discussed in detail in section 4.1.4. The velocity and temperature dependence of friction are discussed in section 4.4.

However, most of the atomistic phenomena that accompany sliding cannot be described even by semi-continuum models. An example of such phenomena is dependence of adhesion and friction on surface chemistry. In a recent paper by Gao et al[117], AFM measurements of friction on polycrystalline diamond were reported, in which substantial variations in the dependence of friction versus load curves have been observed between different locations on the same crystal lattice and for the same sliding direction. These variations were suggested to be due to surface inhomogeneities, which could include adsorbed contaminants, dangling C bonds, C–O moieties, dihydride moieties, vacancies and steps. In this study, the Maugis–Dugdale model was employed to fit the variation of friction with load for different surfaces of diamond. While the quality of the individual fits was quite good, discrepancies were reported between the Tabor parameters from fits of the data and those estimated using realistic values for the equilibrium separation and tip elastic properties. It was concluded that the discrepancy was either due to shear forces present at the interface during sliding as discussed in other AFM studies [31, 43, 155] or to a breakdown of continuum mechanics at the nanometre length scale.

A definite limit for continuum and semi-continuum theories occurs when the contact contains only tens of atoms as evidenced by the ambiguity in determination of contact radii at this scale. At the atomic level, it is not clear exactly where to define the edge of the contact zone, i.e. the contact radius. For real materials, the contact zone edge can be envisioned to correspond to the point where interfacial bonds are no longer formed, although defining whether or not a bond has formed can itself be ambiguous for ionic and van der Waals interactions. In a continuum model, one can define the edge of contact to be where the gap between the continuum surfaces first displays an infinitesimal increase beyond the equilibrium separation, or possibly somewhere further out from the centre of the contact zone to a point where adhesive interactions are below a specified (and possibly arbitrary) value. For parameters in some AFM contacts, one may need to span radially outwards a few nanometres to find a change in the gap of only 1 Å (i.e. less than an interatomic distance). Continuum mechanics cannot provide a definition of the contact zone when the atomistic nature of the material dominates the physical behaviour. This challenging issue is discussed thoroughly by Greenwood [179] and Luan and Robbins [53].

In AFM experiments, the contact radius r or the contact area A cannot be measured directly. One way they can be estimated is from measurements of contact stiffness k_{cont} through the continuum relation:

\[ k_{cont} = 8aG^* \]  

(10)
In this equation $G^*$ is the reduced shear modulus:

$$\frac{1}{G^*} = \frac{2 - \nu_{\text{sub}}}{G_{\text{sub}}} + \frac{2 - \nu_{\text{tip}}}{G_{\text{tip}}},$$

(11)

where $G_{\text{sub}}$ and $G_{\text{tip}}$ are the shear moduli and $\nu_{\text{sub}}$ and $\nu_{\text{tip}}$ are the Poisson ratios for substrate and tip, respectively. Equation (10) assumes homogeneous, isotropic, linear elastic paraboloids in contact, with any adhesion relation (i.e. it applies to Hertzian contacts as well as to adhesive contacts with any value of $\mu T$). In some cases this method leads to estimates of contact radii of atomic or even smaller size, which is in disagreement with the minimum contact size inferred from adhesion forces [180]. Plausible explanations of this discrepancy have been proposed based on MD simulations from adhesion forces [180].

**3.2.2. Atomistic models.** As discussed in section 2.2, atomistic simulations are very well suited to unravel molecular level mechanisms that govern tribological behaviour. In recent years, MD simulations have been employed to explore the limits of continuum theories and to explain some of the experimentally measured deviations from continuum approximations.

One of the assumptions of continuum models is that within the bulk, discrete displacements of atoms can be replaced by continuous strain fields, which in turn can be related to stress fields through elastic constants. This assumption has been tested by Luan and Robbins [52, 53] in MD simulations based on the LJ potential for spherical and cylindrical tip geometries. It was demonstrated that the dependence of normal displacement on load for non-adhesive contact in the elastic regime is quite accurately described by the Hertz model [50]. Cha et al [122] employed MD with a more realistic embedded atom method (EAM) potential for gold to study deformation at a single-asperity contact. It was shown that the JKR theory is applicable to the unloading part of the load versus normal displacement curve, for which the deformation is primarily elastic (dislocations were created during the loading portion of the simulation). These and similar results imply that the discreteness of atoms within a bulk has little effect on mechanics in the elastic regime, even at small length scales. Such a conclusion would not necessarily be true when plastic deformation is taking place. For example, in the aforementioned simulations on Au it was shown that JKR could not be applied to describe asperity loading during which considerable dislocation activity had been observed.

While load versus normal displacement curves are governed by the bulk elastic constants of the substrate and the tip, friction is much more sensitive to the surface structure and its atomistic details. For example, in the aforementioned simulations by Luan and Robbins [52, 53] it was shown that friction and lateral stiffness can deviate from continuum predictions by as much as one order of magnitude, and contact area may differ by a factor of two. Contact pressures can exhibit substantial fluctuations when tip and sample atoms are not in registry (figure 9). Such dramatic discrepancies were attributed to a breakdown of the continuum models (see table 1).

Recently, Wenning and M"user [54] have shown with MD simulations that for small contacts, atomic-scale contamination and lattice commensurability are critical factors. The authors studied the effect of tip curvature combined with the degree of commensurability of the tip and sample, and predicted different scaling laws of friction with load, some of which are consistent with experimental data. In those simulations, atoms in the solids were coupled elastically to their ideal lattice positions and the LJ potential was used to describe interactions of atoms across the interface. It was shown that friction is proportional to load for commensurate tips while amorphous tips lead to a power law dependence.
with exponent $\sim 0.63$, similar to the 2/3 exponent determined in some friction force microscope experiments. In the experimental work this sub-linear dependence of friction on load was attributed to interfacial friction according to equation (1), with the contact area described by the Hertz or DMT models [188]. In simulations, the sub-linear behaviour was explained by the fact that increasing the load leads to an increase in atomic misfit between the tip and the substrate. Interestingly, these results were in contradiction with more recent MD simulations of sliding of diamond tips over diamond samples performed by Gao et al [117], who observed linear behaviour for both commensurate and incommensurate contacts. Gao’s simulations were performed with a more realistic force field than the LJ interaction used by Luan and Robbins [52, 53] and Wenning and Müsser [54], but with much smaller tips. These results indicate that both the system size and the interatomic potentials will have a significant effect on the friction versus load behaviour and ultimately both need to be accounted for when exploring the limits of continuum theories.

Deviations from standard continuum models have been also demonstrated in MD simulations of SAMs. Chandross et al [93] used realistic models of curved AFM tips to study friction on amorphous and crystalline silica substrates coated with alkylsilane SAMs. It was shown that for the case of a 10 nm tip on a C11 SAM, the behaviour of the calculated contact radius lies outside of the limits set by JKR and DMT models. This is not a breakdown of continuum mechanics per se, but at the very least it illustrates the need to apply more appropriate continuum treatments to the monolayer system. Because the contact radius and all material properties can be calculated in a series of simulations, these are true comparisons of the predictions of the models without fitting. The authors have shown that the simulated data were instead accurately predicted with the TCCM model [174] discussed in section 3.1. The contact radius scales as the square root of applied load for TCCM, as opposed to the 2/3 power in the JKR and DMT models.

Continuum models can be extended to account for some of the new phenomena observed in atomistic simulations and in SFM experiments. Suggestions include development of the TCCM model for SAMs or extension of the Maugis–Dugdale model to include normal interfacial compliance. However, even if such extensions are possible, they need be based on a detailed understanding of fundamental physical parameters that can only come from comparisons of single-asperity contact measurements and atomistic simulations.

### 4. Physics of single-asperity friction

Friction is correlated with the energy dissipation that takes place during sliding of two surfaces over each other. Each real contact at the macroscale consists of a large number of micro- and nano-contacts, and friction will correspond to the total energy dissipated in all these small contacts. Understanding the origins of friction is challenging even for a single-asperity contact because the dominant energy dissipation mechanism will be material and environment dependent [146, 190–193]. The multitude of phenomena underlying friction include phonon excitations [194], stick–slip behaviour [195], surface roughness [196, 197], geometric interlocking and interlocking mediated by so-called third bodies [198, 199], wear [200], dislocation mediated plasticity [116], rupture of bonds [201], excitations of charge-density waves in metals [202, 203] and electron scattering [204]. In this section we choose those phenomena that are well established in the literature and that are applicable to more than one class of material. Specifically, in section 4.1 we discuss physical origins of stick–slip behaviour and so-called superlubricity. In section 4.2 we describe known relationships between atomic-scale friction and other mechanical properties. Section 4.3 addresses the dependence of friction on surface structure and composition, and section 4.4 deals with velocity and temperature dependence of friction. Discussion of the current state of the art in the electronic and phononic contributions to friction can be found in [15, 194].

#### 4.1. Stick–slip versus smooth sliding

##### 4.1.1. History of atomic lattice stick–slip

Atomic-lattice stick–slip friction was first uncovered in Mate et al’s pioneering work at IBM Almaden, measuring friction with AFM for a tungsten tip on graphite (0 0 0 1) [205]. The lateral force exhibited stick–slip behaviour with the periodicity of the graphite lattice. Since then, atomic-lattice stick–slip behaviour has been
observed on a wide range of materials: from soft materials like stearic acid crystals with silicon nitride tips [206] to a diamond tip on a diamond surface [23]. Typical atomic-lattice stick–slip behaviour is shown in figure 10 for a silicon nitride tip on the surface of muscovite mica (0 0 0 1). The image exhibits a periodic lattice. The line trace shows that the lateral force starts from zero and builds up to some maximum force. The tip is sticking to the surface throughout this portion of the measurement and there is essentially no relative slip (although there may be some lateral deformation of the tip and sample). The arrow indicates the occurrence of the first slip event. The tip then sticks again until the maximum lateral force is reached once more, and the next slip occurs, and so on. The periodicity of the slip events is equal to 0.52 ± 0.02 nm, which is equal to the lattice constant of the mica surface. The well-defined force, \( F_f \), at which the tip slips, is the static friction force.

AFM images actually track the change in the slope of the end of the AFM cantilever, which bends or twists due to forces parallel to the surface, as explained in more detail in figure 11. Morita et al [207] have carried out a systematic study of atomic-lattice stick–slip on a range of materials, demonstrating precise determination of the slip motions that take place. As seen in figures 11 and 12, both torsional and buckling rotations at the end of cantilever occur, due to frictional forces acting either transverse (\( F_x \)) or parallel (\( F_y \)), respectively, to the projection of the long axis of the cantilever’s projection onto the surface plane. The data in figure 12, and many other experiments, demonstrate that on an ordered sample, the tip, whose surface atoms are not necessarily ordered, prefers to reside in positions in registry with the sample lattice (more on the importance of interfacial commensurability will be discussed below). This periodic interaction is responsible for all atomic-lattice contrast images obtained with contact-mode AFM. One must not imagine the AFM tip smoothly tracing out atomic corrugations as with a scanning tunnelling microscope (STM), but instead realize that the relative tip–sample motion is discontinuous.

The first few observations of this phenomenon were acquired with highly anisotropic samples, such as graphite and mica, which exhibit strong covalent bonding within each layer, but weaker van der Waals or electrostatic forces between the layers. These materials exhibit easy cleavage to expose their basal planes. It was suggested that the periodic forces occurred because a flake of the layered material had become attached to the tip. Thus, the tip and sample structures were commensurate, and therefore a periodic interaction would be expected. However, further measurements reported stick–slip on materials that did not possess such bonding anisotropy, such as NaCl, gold and diamond. Atomic-lattice stick–slip can thus occur between the tip itself and the sample.

4.1.2. Stick–slip and contact size. Initially, several researchers misunderstood data such as those in figure 10, thinking that true atomic resolution was achieved. One aspect that contributed to this misunderstanding is that there is no way to distinguish between the buckling and bending deformation modes of the cantilever (see figures 11(c) and (d)). Therefore, atomic lattice stick–slip behaviour was misinterpreted as being a topographic signal from the corrugation, as seen in STM images. However, this was not the case. The lack of true atomic resolution in contact AFM can be understood in light of the contact mechanics. When the tip is in contact with a given sample, for typical tip radii, loads and elastic constants, the contact is larger than a single atom. For example, a 20 nm radius silicon nitride tip exerting a 1 nN load on a mica sample produces a contact area involving nearly 15 mica unit cells as estimated using the Hertz theory, which neglects adhesion. Including the effect of tip–sample adhesion makes the contact area even larger and can ensure a substantial contact area even...
Figure 10. (a) 7.5 × 7.5 nm² lateral force image of the muscovite mica (0001) surface. The fast scan direction is from left to right. The black dots represent the repeat units of the mica lattice, whose periodicity coincides with the lateral forces. (b) Line trace of the section indicated in (a). The lateral force exhibits ‘stick–slip’ behaviour, where the lateral force builds up to some well-defined maximum value, and then quickly relaxes (first arrow). During the relaxation, the tip slips by one unit cell. This behaviour repeats itself with the lattice periodicity. Reprinted with permission from [343]. Copyright 2004, The Minerals, Metals & Materials Society.

Figure 11. (a) Force components which act on the tip apex of an AFM cantilever. Fₓ, Fᵧ and Fz are the forces across, along, and normal to the cantilever, respectively. These forces cause torsion (b), bending (c) and buckling (d) respectively. Note that in the AFM the cantilever is actually tilted out of the x–y plane, i.e. rotated with respect the x axis. Regardless, the force Fz along the projection of the cantilever’s long axis onto the sample will result in a buckling deformation of the lever. Reprinted from [207] Morita S, Fujisawa S and Sugarawa Y Spatially quantized friction with a lattice periodicity Surf. Sci. Rep. 23 1–41, Copyright 1996 with permission from Elsevier.

at the lowest possible applied loads. Atomistic models confirm this argument. As a result, contact-mode AFM cannot possess single-atom resolution as an STM does. In fact, Mate’s original paper [205] presented similar calculations which showed the contact area to be far greater than a single atom contact. This observation has several consequences, e.g. point defects are not imaged and the lateral resolution of features is limited by the contact area. It therefore remains to be explained why, despite having a multiple atom contact and (most likely) a non-commensurate tip structure, the interaction between the tip and sample possesses the periodicity of the sample’s atomic lattice.

4.1.3. Comparison with conventional stick–slip motion. The term stick–slip must be used with caution: historically, stick–slip refers to a macroscopic behaviour involving multiple contact asperities. A creaking door hinge, a bowed violin string, screeching tires and earthquakes are all examples of macroscopic stick–slip. Furthermore, stick–slip in micrometre-scale single-asperity contacts has been frequently observed in SFA experiments [8]. A rich variety of phenomena are involved in these examples [208], but the unifying principle is that the instability results from the dependence of friction upon the sliding velocity, specifically when friction during sliding is lower than in the static case, or more generally when friction reduces with velocity. When a force applied to surfaces that are stuck together exceeds the static friction force, sliding occurs and thereby friction is lowered. Initially, this leads to increasingly faster relaxation of the applied force until it is no longer large enough to maintain sliding. The system then sticks again and the cycle repeats. The behaviour is influenced to varying degrees by factors such as the roughness of the surfaces, creep strengthening of the interface during sticking and velocity-dependent effects particularly evident in viscous or viscoelastic materials. In contrast, here we are exclusively discussing atomic-lattice stick–slip. Unlike macroscopic stick–slip, the interface is atomically smooth, wear does not occur, and the contact involves only solid materials that are often largely elastic, although the behaviour is also seen in viscoelastic materials.

4.1.4. Commensurability at the interface. Commensurability at the interface is not a necessary condition for the occurrence of stick–slip. For example, atomic-lattice stick–slip has been observed with AFM tips made out of amorphous silicon nitride and oxide [207]. Even if the tip atoms are ordered, they will not necessarily be in an arrangement that is commensurate with the sample’s lattice. Without a commensurate interface, a sufficiently large tip would have no preferred relative positions to reside in, and therefore smooth sliding could be expected. In light of this argument, the regular appearance of atomic-lattice stick–slip for a wide range of tip materials and sizes is surprising.
Figure 12. FFM data of lateral lever twisting ($f_x/k_x$) and buckling ($f_y/k_y$) due to frictional forces parallel to a MoS$_2$ surface acting on a Si$_3$N$_4$ tip. In (a), the lever was scanned along the $x$ direction indicated in (c); i.e. perpendicular to its long axis. Stick–slip behaviour resulted in periodic lateral twisting of the lever, and no appreciable back-and-forth longitudinal buckling of the lever. In (b), the lever was scanned along the $y$ direction indicated in (c); i.e. parallel to its long axis. This time, the lever buckled back and forth as it stuck (top) and twisted back and forth (bottom) as it was scanned. Individual stick points and the path of the tip were mapped out, as indicated in (c), which corresponded to the MoS$_2$ unit cell. Reprinted from [207] Morita S, Fujisawa S and Sugarawa Y Spatially quantized friction with a lattice periodicity Surf. Sci. Rep. 23 1–41. Copyright 1996 with permission from Elsevier.

Not only does this phenomenon occur for many tip and sample materials, but it has also been observed in wet and dry air, liquid and vacuum, and from cryogenic to elevated temperatures. While it is often observed with crystalline samples, it is not always observed. A given tip can vary between conditions where stick–slip is and is not observed. This behaviour is much-discussed amongst experimentalists, but no systematic study of the specific conditions that govern the occurrence of stick slip friction has been carried out.

4.1.5. Recent theoretical approaches. Several theoretical efforts to explain and model atomic-lattice stick–slip behaviour, specifically in the context of force microscopy, have appeared in the literature. These studies can be divided into semi-continuum simulations, analytical models [209–221] and MD simulations [103, 116, 191, 213, 222–225]. The semi-continuum approaches primarily address the mechanics of stick–slip behaviour, i.e. a potential is assumed and the resulting behaviour studied. Most of the semi-continuum approaches build on the Prandtl–Tomlinson model, which was proposed more than five decades ago [177, 226]. Some of these recent models represent the tip as a single atom or a single entity without internal degrees of freedom [209–211, 216, 217, 219], although multi-atom (tens of atoms) tips have also been considered [214, 215, 218]. Scanning is simulated by increasing the lateral displacement between the fixed end of the lever and the sample. The tip initially resides in a potential minimum that is determined by the tip–sample interaction. Because finite static friction due to tip–sample interactions inhibits sliding of the tip, elastic energy is built up in the cantilever and in elastic deformations of the tip and sample themselves [211]. The total energy of the system comprises the interaction energy, and the elastic energy stored in the lever and the deformed contact (see figure 13(a)). Eventually a critical point is reached where the elastic strain energy becomes sufficient to move the system out of the potential minimum. As a result slip between tip and sample takes place. In this slip stage, the lever and the contact quickly relax, the previously stored energy is released, and the motion is brought to a stop as the tip finds a new potential minimum, the closest of which is located at the next nearest lattice site.

The stick–slip motion generates vibrations both in the sample and the cantilever. The phonons excited in this process carry energy away from the interaction region, the vibrations are damped and thereby the energy is dissipated. Since phonon frequencies are much higher than typical AFM scanning frequencies (by a factor of $\sim 10^{11}$), this relaxation occurs very quickly. The collective results of the semi-continuum models can be summarized as follows:

(1) The stick–slip instability can be interpreted as the system (tip and sample) residing in or searching for potential energy minima, where the energy is the sum of the tip–sample interaction and elastic energy stored in the cantilever and the contact.

(2) Sufficiently small stiffness values of the cantilever springs and the contact itself, and a sufficiently strong tip–sample interaction are required to produce the stick–slip instability. If this is not the case, then the stick–slip instability can be prevented and near-frictionless sliding can occur [209, 227–229]. This phenomenon is discussed in more detail in the following section.

(3) The energy stored and then dissipated will be distributed amongst the lever, tip and cantilever depending on their relative stiffness values and damping [211].

(4) Friction decreases with increasing temperatures due to thermally activated hopping across potential barriers [230]. Increased scanning velocity will lead to increased friction, because of the reduced amount of time given to
direction with respect to crystallographic directions. Landman et al. [103, 106, 233] simulated a Si tip/surface pair and observed a wearless atomic stick–slip for low applied loads. Interestingly, a periodic lateral force was observed even if the tip was disordered, but this effect was not discussed in any detail. Sørensen et al. [116] simulated Cu tips on Cu surfaces. Wearless stick–slip occurred for a (1 1 1)-terminated tip sliding on a (1 1 1) surface. The bottom layer of the tip (9 × 9 atoms) was shown to slip via a dislocation mechanism. The tip atoms initially resided in surface fcc positions and during sliding, tip atoms began shifting from fcc to hcp sites, then back to fcc sites, to relieve lateral strain. The slipped and unslipped atoms were separated by a dislocation that propagated through the contact.

These important simulations provide atomic-level descriptions of vibrational motion and energy dissipation mechanisms active during stick–slip motion, revealing that excitations are highly localized in the contact zone. However, such approaches suffer from several major limitations:

1. In all cases, the modelled tip is at least 10 times smaller than those used in AFM experiments.
2. Scanning velocities are several orders of magnitude faster than what is achieved in AFM experiments. Typical MD simulation velocities are 10^6–10^7 m s^{-1} versus typical AFM experimental velocities of 10^{-7}–10^{-5} m s^{-1}.
3. Simplifying assumptions are often made regarding the interaction potentials, including, in some cases, the use of a very generic LJ potential.

4.1.6. Stick–slip transitions: smooth sliding and multiple slip. Recently, Socoliuc et al. [176] used AFM to observe the existence of smooth sliding with no stick–slip when the load was sufficiently low, which corresponded to extremely low energy dissipation. As the load increased, a transition occurred to stick–slip behaviour. The basic idea for this transition goes back to Prandtl [226] and Tomlinson [177], who both showed that stick–slip instability occurs when surface corrugation is sufficiently weak or the cantilever spring is sufficiently stiff. Specifically, Tomlinson introduced a parameter γ_T = (2π^2 V_0)/(k_{exp} a_i^2), which describes the relation between the lateral corrugation amplitude of the tip–substrate interaction V_0, the substrate lattice parameter a_i, and the experimental lateral stiffness of the system k_{exp}. The experimental lateral stiffness can be evaluated from the expression: k_{exp} = k_{lever} + \frac{1}{k_{tip}} + \frac{1}{k_{cont}}, which includes the effect of the lateral stiffness of the cantilever k_{lever}, of the tip structure k_{tip}, and of the tip–sample contact k_{cont} [33] as given in equation (10). Atomic stick–slip behaviour is observed only if γ_T > 1, i.e. when the system is sufficiently compliant, or the interfacial corrugation is sufficiently strong.

When γ_T < 1, sliding occurs without stick–slip instabilities. This phenomenon has been termed superlubricity [234]. The term is misleading because this is not a quantum coherence phenomenon, and there can still be remaining dynamic dissipation. However, so far the friction force observed in these cases has been lower than the detectable limit of the AFMs used, and correspondingly the friction loops have no observable hysteresis within the experimental uncertainty.
The transition to smooth sliding has been accomplished by a variety of ways. For example, Dienwiebel et al [235] observed superlubricity for a graphite flake attached to the tip sliding on a graphite surface. Smooth sliding was only observed when the flake and the surface were oriented to be incommensurate, in agreement with the idea that incommensurability renders the corrugation of the interfacial potential sufficiently weak to allow smooth sliding. Interestingly, the same group subsequently showed, with both experiment and theory, that as sliding proceeded the flake exhibited a tendency to rotate and settle into a stable registry with the substrate, which destroys the superlubric state [236]. According to the proposed model, this rotation is induced by a net torque that exists naturally between the flake and surface.

Smooth sliding has also been observed for sharp AFM tips sliding over atomically flat surfaces at sufficiently low loads [176,237]. In these experiments, the superlubricity was enabled by the sufficiently low interfacial corrugation resulting from the low applied loads and, as shown by Socoliuc et al [176], it was correlated with low values of the interfacial lateral stiffness. The demonstrated agreement with the predictions of the Tomlinson model is impressive (see figure 14).

Recently Socoliuc et al [238] reported another method that can be more readily applied in practical situations. When the tip is oscillated at frequencies corresponding to normal resonances of the system, one can intermittently achieve low loads and thus low interfacial corrugations with very little actuation required. Such an oscillating tip is able to stably slide at the instants where the corrugation (load) is low, resulting in a lower net friction.

Another type of transition has been predicted by Johnson and Woodhouse [239] who showed that under certain conditions, slip may occur over an integer number of lattice spacings. This phenomenon is called a multiple slip. In fact, multiple slip was observed in the original letter reporting atomic lattice stick–slip by Mate et al [205], but since then it was barely discussed until recently. Johnson and Woodhouse identified the relationships between the lateral (i.e. torsional) cantilever stiffness, the lateral stiffness of the elastically deformed contact itself, and the corrugation of the lateral force interaction, as key parameters controlling the transition to multiple slips. An adjustable damping factor was introduced, which represents the dynamic energy dissipation in the tip or sample materials, or in the cantilever itself. The transition from single to double slips occurs when high-frequency fluctuations in the lateral force, triggered by the slip instability, overshoot the corrugated lateral tip–sample interaction force. The possibility of an overshoot reduces with increased damping.

Further models have been employed to describe transitions from single to multiple slip in atomic-scale friction. For example Nakamura et al [240] employed dynamical simulations of the Tomlinson model to investigate transitions between single and double slip modes and their dependence on damping, sliding velocity and finite temperature. Conley et al [232] used a combination of continuation techniques, perturbation methods and numerical simulations to deal with issues of the complex dynamics in atomic-scale friction. The authors considered a quasistatic limit and transitions between multiple slip modes by solving the equation of motion numerically. An analytical solution of the transition between different slip regimes for the simple case of the one-dimensional Tomlinson model in the quasistatic limit was recently reported by Medyanik et al [237].
analysed the energy landscape and showed how the number of local energy minima increases with increasing interfacial corrugation (see figure 13(b)). Single and multiple slips correspond to sliding of the tip to the nearest, the next nearest, the next-next nearest etc., local minimum. Slipping to further minima can occur only with sufficiently low energy dissipation during slip. Transition from single to multiple slip occurs with increasing load which indicates that corrugation increases with load. Specifically, the existence of multiple slip regimes is governed by the Tomlinson parameter $\gamma_T$ reaching specific values. In other words, $\gamma_T = 1$ is the transition from smooth sliding to slips by one lattice site. The possibility of slips of higher multiplicity occurs for larger critical values of $\gamma_T$. In the same paper, the authors reported experimental observation of the dependence of stick–slip behaviour on load. The experiments were performed on an HOPG sample and, as shown in figure 15, the system exhibited superlubricity at the lowest applied load. At higher loads, stick–slip instabilities occurred with periodicity of the HOPG lattice, while increasing load even further leads to slips over integer multiples of the lattice spacing as predicted by the model.

4.1.7. Remaining questions. There has not yet been any clear conclusion indicating under exactly what conditions stick–slip behaviour occurs. Often, beautiful images such as the one shown in figure 10 are not obtained. It is possible that under the same loads, with the same sample and with the same cantilever, some unknown change in the tip occurs and stick–slip is suddenly observed. The reasons for this are not established. Furthermore, no one has studied whether friction varies with load in the same manner in the presence and absence of stick–slip.

Another unresolved question pertains to stick–slip periodicity. Most accounts so far report one stick–slip event per surface unit cell, even when the unit cell contains more than one atomic species, such as alkali halide surfaces, including KBr [241] and NaF [207, 242]. One exception is the large unit cell of Si(1 1 1)7 $\times$ 7 measured in UHV with tips coated with polytetrafluoroethylene [243] where multiple stick–slip events per unit cell were resolved. With KBr, Giessibl and Binnig [244] ‘resolved’ both K$^+$ and Br$^-$ ions in the normal force signal in UHV at 4.2 K, so two stick–slip events per unit cell took place. On the other hand Lüthi et al [241] observed only one stick–slip event per unit cell at room temperature with KBr in UHV. What then determines the periodicity of stick–slip behaviour? Are all the interfacial atoms participating in periodic motion, or only a fraction of them?

As discussed in section 4.1.4, energy released during sliding is carried away by phonons excited in the sliding process. Phonon frequencies are eleven orders of magnitude higher than AFM scanning frequencies and therefore the relevant dissipation processes may occur quickly. However, it has been found recently in experiments involving Si tips on KBr samples that slip times in atomic stick–slip can be as long as 10 ms [245]. Such time scales are currently not accessible to conventional atomistic simulations. However, the accelerated simulation techniques, such as in [96], may be a route to addressing this challenge.

Overall, a tremendous gap between experiment and modelling exists. The gap spans materials, time scales and length scales. The lack of knowledge of the tip is also a problem. To resolve these issues, a tight-knit effort that focuses on tractable systems, new quantitative approaches and creative ideas is required.

4.2. Relationship between mechanical properties and friction

For engineering applications it is often not one particular mechanical property that is desired, e.g. low friction, but a combination of such properties. For example, design of micro- or nanoresonators will require good tribological properties of the surfaces as well as high Young’s modulus of the material. For high-performance coatings, good wear resistance needs to be accompanied by high hardness and high fracture toughness. Because of the diverse nature of physical mechanisms underlying energy dissipation during sliding, there are currently no straightforward rules to describe the relations between tribological and other mechanical properties. Here we review the correlations that have been so far demonstrated in the literature and we discuss their limitations.

Elastic moduli $E_{\text{sub}}$ and $E_{\text{tip}}$ of the substrate and the tip, respectively, are related to the continuum contact modulus $K$ introduced in section 3.1 through a reduced modulus $E^* = \frac{1}{2} K$. In the JKR and DMT models, the contact area $A$ and contact radius $a$ scale as $A \propto a^2 \propto (E^*)^{-2/3}$. According to the model of interfacial friction, the friction force is proportional to $A$. Therefore, stiffer materials are, in general, expected to exhibit lower coefficients of friction. Reduced Young’s modulus (or normal stiffness) $E^*$ enters the Tabor’s parameter $\mu_T$ (see equation (8)) with the same exponent of $-2/3$. Tabor’s parameter defines the applicability of the JKR, DMT or intermediate models. As a result, systems with high normal stiffness will be better described by the DMT model, while the JKR model will be appropriate for more compliant systems, all other parameters (adhesion, range of interfacial forces and tip radius) being held constant.

The shear modulus of the bulk is another important parameter in tribological applications. The reduced modulus
Consequently, unless based on experimental lateral stiffness through equation (10). This relation is derived from continuum theory, and it assumes a contact between two homogeneous, isotropic, linear elastic materials. Equation (10) involves the experimental lateral stiffness $k_{\text{exp}}$, which is the slope of the friction force versus tip displacement curve. As discussed in section 4.1.5, this stiffness constitutes an important parameter in the Tomlinson model of stick–slip friction. However, the true lateral stiffness $k_{\text{true}}$ of the system can differ from $k_{\text{exp}}$ due to the finite lateral stiffness of the interfacial potential. This issue becomes critical when the Tomlinson parameter $\gamma_T \sim 1$ or less [176]. Researchers have indeed observed small values of $\gamma_T$ for nanoscale contacts [176]. The true stiffness can be expressed as:

$$k_{\text{true}} = \frac{\gamma_T + 1}{\gamma_T} k_{\text{exp}}. \quad (12)$$

Consequently, unless $\gamma_T$ is large, the contact area determined based on $G^*$ will be underestimated since $G^*$ does not include the interfacial stiffness [177].

It has been proposed that relative motion of two contacting bodies during sliding can be assisted by dislocation motion [246–249]. Dislocation motion is a vehicle for plastic deformation and, because it is normally irreversible, constitutes an efficient mechanism for energy dissipation during friction. Dislocations are generated in a substrate and propagate under load when the shear stress resolved on a given slip plane reaches a critical value of the Peierls stress $\Gamma_T$ for nanoscale contacts [250]. Gliding of a dislocation along the interface will lead to a relative displacement of the two bodies in contact [251]. This phenomenon is known in tribology as a microslip. For contact radii between tens of nanometre and fewer micrometres, dislocation-assisted sliding has been shown to be energetically more favourable than concurrent slip when all adhesive bonds are broken simultaneously [43,246]. Merkle and Marks [252] built on established theories of dislocations and interfaces to develop an analytical model for friction. The proposed model is capable of reproducing the velocity, temperature and orientation dependence of friction, including the superlubricity phenomenon and contamination effects. This approach has a substantial potential as it leverages the well-established theory of dislocations.

It is not clear how important dislocation-assisted slip is in nanometre-size single-asperity contacts, since it is known that decreasing asperity size increases resistance to plastic deformation [253]. However, there is evidence based on MD simulations that at least in certain cases dislocations will play a significant role at these length scales. As mentioned in section 2.2.4, Sørensen et al [116] performed simulations of a number of tip–surface contacts for copper. It was shown that stick–slip behaviour is related to nucleation and subsequent motion of dislocations at the interface. There is also experimental evidence [254] of dislocation nucleation in potassium bromide (KBr) substrates under AFM tips of curvature radii as small as ~7–30 nm. Dislocation activity and its effect on friction will be material specific and will depend on a particular surface orientation. If for geometrical or statistical reasons there are dislocations concentrated at the surface, the stress at which the plastic yield occurs at the interface will be lower than yield strength of the bulk substrate. Care needs to be exercised when predicting onset of permanent deformation during sliding. For example, in friction force microscopy experiments on KBr carried out by Luthi et al [241] wear appeared at a load as low as 3 nN. Atomic scale wear phenomena are discussed in section 6.

4.3. Relationship between surface atomic composition and friction

Friction is a measure of how much work done on the system is converted into another less ordered form of energy. For strongly adhering systems, this work may be converted into plastic deformation within the bulk, e.g. the formation of shear planes [255]. Weaker adhesive forces will result in localization of the damage near the surface, for example by forming wear debris or transferred particles [256,257]. For still weaker adhesive forces, no damage is done to the surface and we refer to this case as wearless or interfacial friction [258,259]. There is therefore evidence for strong correlation between adhesion, friction and wear, and the details of this correlation depend on the specific system. For example, while SAM coatings can dramatically reduce adhesion between Si substrates, energy dissipation and wear in these films can still be severe [3]. That reduced adhesion can lead to a decrease in friction was demonstrated in AFM experiments and in MD simulations of diamond (1 1 1) single crystal surfaces [260,261]. It was shown that hydrogen passivation of dangling (reactive) bonds at the diamond surface leads to a reduction in friction by more than two orders of magnitude. The high friction, adhesion and wear, which were observed for only partially passivated surfaces, are due to bridging of the interface by unterminated (dangling) bonds. Consistent results were reported for the underside of ultrananocrystalline diamond (UNCD) films investigated with tungsten carbide and diamond AFM tips [262,263]. As shown in figure 16, fully hydrogenating the surface removes sp$^2$-bonded carbon and oxygen, and correspondingly yields a reduction of both adhesion and friction.

The dependence of adhesion on the chemistry of the surface has been modelled by Qi et al [264]. The authors employed DFT-based ab initio calculations to determine the effect of surface termination on the work of adhesion, interface geometry, bond character and adhesive transfer at interfaces. The work of adhesion calculated for the (1 1 1) diamond surface terminated with H was lower (8 mJ m$^{-2}$) than for the OH-terminated surface (20 mJ m$^{-2}$). These results are at the same order of magnitude as the experimentally determined work of adhesion between diamond tip and H terminated UNCD (10.4 ± 4 mJ m$^{-2}$). However such comparisons should be made with caution because of the limitations of DFT calculations. Specifically, the standard DFT method does not account correctly for the van der Waals interactions, which can have a non-negligible contribution to adhesion.

Another way by which surface termination can affect friction is by changing the vibrational properties of the surface. Cannara et al [265] performed AFM studies of
hydrogen- and deuterium-terminated single crystal diamond and silicon surfaces and have shown that changing the mass of the chemisorbed species changes the rate at which the tip’s kinetic energy is dissipated. Deuterium-terminated surfaces exhibit reduced friction, which confirmed the hypothesis that adsorbate excitations can substantially contribute to energy dissipated during sliding.

The term tribochemistry refers to the changes in the chemistry of the contacting surfaces induced by sliding. Tribochemical reactions can be induced by frictional heating, as evidenced by enhanced chemical reactivity of sheared regions as compared with unsheared regions in SFA measurements of Ag/metal substrate interfaces [266]. Chemical reactions observed in these experiments, e.g. formation of silver oxide, could not be explained solely by the chemistry of the individual species, but required some knowledge of tribology, such as the local stresses, temperature rise and type of wear produced by shearing surfaces.

Tribochemical properties can be also affected by environmental species physisorbed to the surfaces. For example, friction between DLC films and Si$_3$N$_4$ AFM tips was shown to increase monotonically and reversibly with relative humidity [267]. Because the measured adhesion showed no dependence on humidity, it was concluded that the increase in friction arises from the increase in shear strength due to physisorbed water, in contrast to a conventional interpretation of meniscus formation. Schwarz et al [49] have shown that diamond and amorphous carbon exhibit lower friction in dry Ar environment as compared with air with 40–60% humidity, which was attributed to the presence of water on the surface in the latter environment.

It is worth noting that relationship between adhesion and friction is still a matter of debate. Recently, Zhang et al [268] employed MD simulations of sliding of Al$_2$O$_3$ and Al slabs in various configurations involving commensurate and incommensurate interfaces to show that static friction is actually not related to adhesion, but to the change of potential energy along the sliding direction. The lack of correlation between adhesion and friction found in this study is in contrast to the popular belief that both of these properties increase or decrease concurrently.

Very recently, the lateral forces required to move a single atom and a single molecule across a crystalline surface have been measured [269]. Using the highly sensitive FM technique mentioned in section 2.2.3 in ultrahigh vacuum at a temperature of 4 K, the authors were able to move individual adsorbates laterally on single crystal substrates using a metal tip. Specifically, they found that 210 ± 30 pN of lateral force was required to move an individual Co atom on a Pt(1 1 1) substrate, but only 17 ± 3 pN for a Co atom on Cu(1 1 1). A CO molecule on Cu(1 1 1) required 160 ± 30 pN.

The lateral force measurement in this case is indirect. In the FM experiment, a shift in the flexural resonance frequency of an oscillating cantilever is measured. This shift is proportional to the vertical force gradient $\frac{\partial F_z}{\partial z}$ of the tip–sample interaction force $F_z$. The authors measured this gradient as a function of the average tip–sample separation $z$ and lateral coordinates ($x$, $y$), integrated twice over $z$ to evaluate the tip–sample potential, then differentiated with respect to $x$ to get the lateral force. A potential energy landscape is mapped out in the process, and it compares favourably with energy barriers estimated from DFT calculations. The observation that the identities of the adsorbate and the substrate atoms critically affect the force needed for lateral motion confirms, at the atomic level, that surface composition affects lateral forces. This remarkable measurement extends studies of nanoscale friction truly to the scale of individual atoms.

In summary, a thorough study of the complex relationship between surface atomic composition, adhesion and friction is needed. On the simulation side, there is a need to develop models that can simultaneously capture chemical reactivity of the surfaces and the geometrical effects related to the size and shape of realistic single-asperity contacts. Experimentally, careful characterization of the surface chemistry and atomic-scale roughness need to accompany measurements of tribological properties so that systematic understanding and a thorough knowledge base can be built and used in intentional surface preparation.

4.4. Velocity and temperature dependence

While the classic laws for dry macroscopic contacts between solids state that friction is independent of velocity, for contact sizes in the nanometre range friction can increase, decrease, or be independent of sliding velocity. A logarithmic increase of friction with scanning velocity was found by Bouhacina et al [270], who studied friction of polymer layers grafted on silica over a range of velocities and with four different probes. A thermally activated Eyring model was employed to analyse these data, resulting in estimates of the interfacial shear stresses and energetic barriers to sliding. While some uncontrolled approximations were used in the analysis (leading to large variations in calculated contact areas), the fits for the
above quantities are consistent for the different probes. Shear stresses were found to lie mostly between 200 and 300 MPa for bare silica and between 400 and 600 MPa for the grafted layers, with barrier heights in the range 1–2 × 10⁻¹⁹ J for both systems. FFM studies of diamond, graphite and amorphous carbon by Zwörner et al. [271] exhibited no velocity dependence as long as the scanning velocity was much lower than a characteristic slip velocity of the tip (∼60 μm s⁻¹). Consequently, no velocity dependence of friction was observed for scanning speeds of up to about 10 μm s⁻¹, and a logarithmic dependence was found for larger velocities. A different trend was found in atomic scale studies by Bennewitz et al. on Cu(1 1 1) [272] and Gnecco et al. on NaCl(1 0 0) [231], who reported a logarithmic dependence of friction on scanning velocity for velocities less than about 1 μm s⁻¹. A thermally activated slip model and a modified Tomlinson model, respectively, were used to interpret data in these two studies. A similar activation model was used to describe another set of experiments by He et al. [191] on hexadecane and octamethylyclosoterasiloxane, where a logarithmic dependence of friction was observed for scan velocities ranging from 0.1 to 100 μm s⁻¹.

Surface chemistry and reactivity have been shown to have an important effect on the velocity dependence of friction. For instance, Liu et al. [273] measured friction by sliding a sapphire ball over Si(1 0 0) coated with different films: a native oxide, diamond-like carbon (DLC) and SAMs. The experiments revealed an increase of friction with increasing velocity in all the cases except for the native oxide coating, for which friction decreased with increasing velocity. The native oxide of Si is naturally hydrophilic, and a capillary water neck is expected to form at the contact. The decrease in friction in this case was attributed to the decreased ability of the system to maintain a stable capillary neck as the sliding velocity increases. Hild et al. [274] found similar results for oxide-, and SAMs-coated Si(1 0 0). AFM experiments by Riedo et al. [275] on CrN and DLC surfaces revealed a logarithmic increase of friction with sliding velocity for partially hydrophobic surfaces of DLC and a logarithmic decrease of friction with increasing velocity for partially hydrophilic surfaces of CrN. A model based on the kinetics of capillary condensation was developed, which fully reproduced the experimental data. Specifically, two competing mechanisms that contribute to the velocity dependence were proposed. A thermal activation model, which is appropriate for stick–slip motion, leads to a logarithmic increase in friction with increasing velocity. On the other hand, capillary condensation results in a logarithmic decrease in friction with increasing velocity, in agreement with the aforementioned experiments by Liu et al. The hydrophilicity of a surface and the humidity (when experiments are performed in ambient conditions) are therefore major factors in determining the final velocity dependence. In a subsequent set of experiments, Riedo et al. [276] proposed an analytical model of hopping dynamics on a corrugated effective potential surface. Experimental results presented in the same report confirmed predictions of the model, i.e. it was observed that friction increases logarithmically with increasing velocity up to a critical velocity, beyond which friction remains constant. Another example of how surface chemistry affects the velocity dependence of friction was recently reported by Chen et al. [277]. The authors demonstrated that friction decreases with increasing velocity in systems that are capable of forming cross-linked structures across the interface with linkage energies that are easily overcome at the applied loads. The opposite trend is true if surfaces are terminated with species that are not capable of forming such networks.

Concurrently with the experimental studies, modelling efforts were reported that aimed at understanding the effects of velocity on friction. For example Sang et al. [178] argued that earlier models [231,270], which were based on the concept of thermal activation and resulted in a logarithmic velocity dependence, described in fact the case of linear creep between surfaces (i.e. when a force acting on a substrate produces a small constant potential bias) rather than the experimental situation in which a tip is dragged across a surface. The latter phenomenon is known as ramped creep, which means that a potential bias is continually ramped up until the support is moved. The proposed model predicts that friction is proportional to the logarithm of velocity raised to the power of 2/3. The velocity dependence of friction was also studied by means of atomistic simulations. The MD study by Sørensen et al. [116] revealed two different mechanisms underlying this dependence: (i) excitation of phonons which promotes future slip events and (ii) the delay between motion of the bottom of the tip, and the top of the tip (i.e. where the force is applied). Note that the sliding velocity in these simulations (1–20 m s⁻¹) far exceeds what is typically achieved in AFM experiments (at most 100 μm s⁻¹). As mentioned in section 2.2.1, this time discrepancy is one of the outstanding challenges for making meaningful comparisons between MD and AFM.

The conclusion to be drawn is that, in the absence of a dominant environmental and/or chemical adhesion mechanism across the tip–substrate interface, friction increases logarithmically with scan velocity due to reduced opportunities for thermally activated slip at higher velocities. This theoretical framework has explained a number of experimental results. In the presence of capillary necking or hydrogen bond formation, the dependence becomes more complicated since both these effects have been shown to result in a logarithmic decrease of friction with a scan velocity. There is thus a competition between thermal and adhesive effects, and their balance will determine the final velocity dependence.

Investigations of the temperature dependence of nanoscale friction have been less common, but recently there have been a few extremely interesting results published. As with any system in a metastable state, thermal excitations can provide sufficient energy to overcome local barriers and enable slip. Thus, a general prediction is that interfacial friction should reduce with temperature provided no other surface or material parameters are altered by the change in temperature. Additionally stick–slip behaviour should be suppressed as the temperature increases. This effect has been termed thermolubricity [230].

Krylov et al. [230] considered the effect of finite temperature on the Tomlinson model. In this study thermally induced hopping from neighbouring stable tip positions was modelled analytically, and the authors introduced a non-dimensional parameter μ = V/t exp (β/μ), where V is the
scans, \( a \) is the lattice spacing, \( E_0 \) is the corrugation of the interfacial potential, \( k \) is Boltzmann’s constant and \( r_0 \) is an undetermined pre-exponential factor for the probability per unit time of the thermally activated slip event. Similarly to the stick–slip transition parameter \( \gamma_T \), the parameter \( \beta \) determines a transition from a high-friction stick–slip regime to a low-friction smooth sliding regime. Values of \( \beta \gg 1 \) correspond to the low-temperature stick–slip regime where thermal energy does not substantially affect friction, and the case of \( \beta \ll 1 \) corresponds to an elevated-temperature thermal drift regime where the tip exhibits many back-and-forth jumps between potential wells. When \( \beta \approx 1 \), approximately one activated jump occurs for every interatomic site traversed at the scan velocity \( V \), which case can be regarded as a stochastic or irregular-slip regime. In the low-temperature regime, friction is predicted to still depend logarithmically on velocity, as in the zero-temperature Tomlinson model, but at higher temperatures the dependence would become linear. Another important prediction is that even in the low-temperature regime, friction is appreciably lower than that estimated from the Tomlinson model. The authors support their theory by making comparisons to room-temperature friction data from two distinct experiments [176,278]. Indeed, the variation of friction with the Tomlinson parameter \( \gamma_T \) follows the theoretical prediction of the authors, and for the data from [278] it notably deviates from the predictions of the athermal Tomlinson model. This behaviour is consistent with the presence of thermally induced slips observed in individual friction loops and it indicates that thermal effects are critical to consider during stick–slip motion. Interestingly, the prefactor derived from fits to the experimental data yields attempt frequencies \( r_0 \) that are surprisingly small (from 0.2 to 1.6 kHz). These values are much smaller than any atomic lattice vibration frequency, but may correspond to vibrations of the cantilever, or of the contact interface itself, which due to its low stiffness may have a low vibrational resonance frequency.

Temperature dependence was also investigated by Zhao et al [279] who measured friction between a silicon nitride AFM tip and HOPG in UHV in the temperature range of 140–750 K. Dramatic reductions in friction force and friction coefficient with temperature were observed and the data were very well approximated by a simple Arrhenius relation, which indicates that frictional sliding is indeed assisted by thermal activation. Regular stick–slip motion was not observed in this experiment, and therefore a comparison to the finite-temperature Tomlinson model could not be conducted. However, the results were remarkably consistent with macroscopic friction measurements reported for a range of solid lubricant systems, including polytetrafluoroethylene (PTFE) [280].

Schirmeisen et al [281] measured friction between a Si AFM tip and a Si(1 1 1) wafer coated with a native oxide in UHV in temperatures between 50 and 273 K. The authors observed a peak in both adhesion and friction coefficient at \( \sim 100 \) K, after which a sharp decrease in both properties occurred. The friction behaviour above 100 K is similar to the dependence seen in the experiment of Zhao et al and may in fact be another example of a thermally activated friction. A transition in the velocity dependence of friction was also observed: friction increased logarithmically with velocity below \( \sim 150 \) K, but was independent of velocity above 150 K.

Complete understanding of the temperature and velocity dependence of friction both for nanoscale single-asperity contacts and for macroscopic interfaces requires further studies. However, these very recent results can arguably be said to consistently indicate that the thermal behaviour of low-wear systems (nanoscale asperities, or macroscopic sliding of solid lubricants, both of which exhibit interfacial sliding) is generally governed by a thermally activated behaviour. If these conclusions are confirmed by future studies, it would be a significant advance in our ability to predict friction since there is a paucity of friction data over a wide range of temperatures even for macroscopic contacts.

5. Friction of solids coated with monolayer lubricants

The study of nanotribology of monolayer lubricants began in the late 1980s following the introduction of AFM [282] and the demonstration of the first MEMS [283, 284]. Monolayer lubricants are interesting from a nanotribology standpoint because they can provide boundary lubrication in cases where thin coatings are needed but the use of liquid lubricants is contraindicated due to the domination of viscous drag forces under nano-confinement. Additionally, monolayer lubricants can be attached to a number of different substrates ranging from ceramics to metals, and provide well-defined interfaces with tailorable chemical activity. An example of a specific successful commercial application is hard drives, where read/write heads fly at distances of a few nanometre above the platters. In this case, effective boundary lubricants have allowed for the increase of data storage, which is directly related to fly heights.

5.1. Boundary lubrication and MEMS

The study of nanotribology of SAMs is intimately related to the development of MEMS, where the high surface to volume ratio makes the interfacial interactions a dominant factor in the wear and lifetime of devices. While MEMS can be fabricated out of a number of different materials, the most popular by far is silicon or, more specifically, polycrystalline Si. A major advantage of Si over other materials is the massive infrastructure and knowledge base for design, fabrication and testing of Si devices that exists in the microelectronics industry. Current Si technology allows combining logical and actuation functions on the same chip, thereby saving fabrication time and resources, and greatly reducing the number of interconnects which are a major source of device failures [285]. An additional advantage of Si lies in its intrinsic properties, which among other things allow for highly selective etching, relatively high processing temperatures and deposition of films with low residual stresses, all while being compatible with integrated circuit processing lines [285].
Polysilicon rapidly oxidizes in air, resulting in a formation of a highly reactive, hydrophilic surface layer. If not removed during the release process (e.g. by supercritical drying techniques), residual water will cause components to adhere, rendering devices useless. After release, even brief contact of untreated parts can lead to undesired adhesion, which effect is generally referred to as stiction [286]. SAMs are very useful in reducing stiction, as their head groups can be tailored for attachment to the oxidized silicon substrate, while their hydrophobic tail groups can greatly reduce adhesion and interfacial friction. While SAMs have shown a great promise for use as boundary lubricants in MEMS, a number of questions still remain unanswered, ranging from fundamental properties of SAMs to stability and wear of coatings subjected to both mechanical and environmental stresses.

In order to be effective as a boundary lubricant, there are a number of properties that a given SAM must exhibit. The most obvious one is a reduction of adhesion and friction as compared with an uncoated substrate. Additionally, stringent requirements are placed on the ability of the coating to withstand processing environments, such as high temperatures. It is further desirable that the monolayer be able to fully coat the devices with reasonable packing densities, including occluded regions such as those found between a micro-gear and the substrate. Further requirements include the ability to withstand both the mechanical stresses due to contact (either through impact or rubbing), as well as the environmental stresses from packing and ageing. Oftentimes a trade-off has to be made with respect to these different requirements. For example, in the SUMMiT™ process used to manufacture MEMS at Sandia National Laboratories, two different precursors are used for molecules in the SAMs: perfluorodecyltrichlorosilane (C_8F_17C_2H_4SiCl_3, FDTS) and its hydrocarbon counterpart octadecyltrichlorosilane (C_18H_37SiCl_3, ODTS). Both FDTS and ODTS form hydrophobic films, with water contact angles of 115° and 110°, respectively, in contrast to less than 30° angle exhibited at the native oxide. While FDTS displays higher friction and has not been as widely studied as ODTS, it is favourable due to its ability to withstand higher processing temperatures of the SUMMiT™ process.

5.2. Details of SAMs as boundary lubricants

The nanotribology of SAMs has been widely studied, both because of their applications in MEMS as well as for their generally desirable properties as model systems for boundary lubrication. In general, a SAM comprises long or short chain molecules with a head group that binds to a surface (either through physisorption, chemisorption or both), and a tail group that delivers desirable chemical qualities to the exposed interface. The majority of the nanotribological studies have been performed on hydrocarbon SAMs (similar to ODTS), referred to as C_n, where n stands for the number of carbon atoms in the backbone. Some work has also been reported on the fluorocarbons CF_n, as well as on other molecules such as phosphonic acid SAMs [70, 287–289] and alkoxyl monolayers [187]. Mixtures of monolayers have also been studied [290], including both mixtures of C_n and CF_n as well as mixtures of C_n with different chain lengths [290].

When a thiol group is used as a head group, chains of SAM can be attached to metal substrates such as gold or silver and such alkanethiol SAMs are known to form ordered structures [291]. Other materials, e.g. diamond [255] amorphous carbon [255], and aluminium [70] have also been explored as substrates for SAM coatings. We will focus here on alkysilane-type SAMs on both SiO_2 and gold substrates, as these are the two most extensively studied systems.

SAMs are generally formed on silica through solution-based [292] or vapour phase deposition methods [293] in which the precursor trichlorosilanes react with water and become hydrolyzed before bonding to the surface. Covalent Si–O–Si linkages can potentially form between the head groups of neighbouring molecules and/or between the head groups and hydroxylated sites on the silica surface. However, steric considerations demonstrate that it is impossible for the SAM to be bonded to the surface and simultaneously cross-linked to its neighbours over large areas [294]. This is because the chain–chain separation is far too large to accommodate the rather short Si–O–Si bond lengths. Furthermore, the relative amounts of physisorption (e.g. H-bonding or electrostatic interactions) and chemisorption (Si–O–Si linkages) between the SAM molecules and the surface are unknown [295]. Order in the SAM is strongly dependent on chain length, with longer chains generally leading to more ordered monolayers. A simple calculation of van der Waals cohesive energy between chains due to CH_2 groups provides a reasonable physical basis for this effect [296]. The effects of ordering and of chain length on friction measured in AFM experiments can be decoupled from one another and the dependence on ordering has been proven to be much stronger. Both of these effects are discussed in the subsequent sections.

5.3. Early experimental studies of SAM lubricants

Tribological studies of SAMs were preceded by friction measurements of Langmuir–Blodgett films of fatty acids on mica [41] and metals [297] performed in the early 1980s. Friction of alkysilane-type SAMs was first measured in 1988 in a pin-on-disc experiment, where among other systems 7-octenyltrichlorosilane (CH_2=CHCH_2CH_2CH_2CH_2CH_2SiCl_3) and heptadecafluoro-(1,1,2,2-tetrahydrodecyI)-1-trichlorosilane (CF_3CF_2CF_2CF_2CF_2CF_2CF_2CF_2CF_2CH_2SiCl_3) molecules on glass were investigated [298]. These molecules reduced friction coefficient µ to 0.35 and 0.28, respectively, from the bare glass value of > 0.61. The values of µ for these molecules are higher than those reported in later studies, as discussed below, but this effect could be due to the presence of a double bond in the tail group. Another alkysilane-type molecule was studied in 1989 when µ of a monolayer of mono-octadecyltriethoxysilane (C_18H_37Si(OC_2H_5))_3, a molecule similar to ODTS but with a bulkier headgroup, was measured to be approximately 0.1.

Also in 1989, De Palma and Tillman [299] used a glass slider in a pin-on-disc experiment to compare adhesion and friction of ODTS, undecyltrichlorosilane (C_11H_23SiCl_3,
UDTS) and FDTS monolayers on freshly cut silicon wafers. The friction force $F_I$ for all systems was found to be linearly dependent on the applied load $P$, with friction coefficients $\mu$ (given by the slopes of the $F_I(P)$ lines) of 0.07 ± 0.01 for ODTS, 0.09 ± 0.01 for UDTS and 0.16 ± 0.02 for FDTS. In this study the ratio $\mu_{\text{FDTS}}/\mu_{\text{UDTS}}$ was found to be close to 2, in contrast to the value of 3 that is generally quoted in the literature [290]. Note also that there is a hint of chain-length dependence here in that $\mu_{\text{UDTS}}$ is slightly larger than $\mu_{\text{ODTS}}$. Factors influencing the chain length dependence of the frictional response will be discussed in detail in section 5.5.

The adhesion of alkanethiol SAMs on gold substrates was first studied in 1992 with a bare tungsten tip compressing a $n$-hexadecanethiol (C16H33SH) monolayer in an IFM experiment [300]. This work found little interaction between the tip and sample. In 1994 Frisbie et al performed AFM experiments that measured the interactions between tips and substrates that were both coated with monolayers with either the same (COOH/COOH or CH$_3$/CH$_3$) or different (COOH/CH$_3$) functional groups [301]. The strongest interaction was found between COOH/COOH coatings, followed by COOH/CH$_3$, with the weakest interaction between CH$_3$/CH$_3$ coatings. One of the first friction studies of alkanethiols themselves was reported in 1995, where frictional force microscopy experiments of other tip/substrate termination combinations were performed, including a bare tip against a C17 alkanethiol monolayer [302]. This work measured $\mu = 0.23$, in contrast to experiments with a CH$_3$ modified tip, which found $\mu = 0.07$.

5.4. Early simulations of SAM lubricants

The first MD simulations of SAM friction were performed in 1993 by Glosli and McClelland [303], who studied opposing pairs of $6 \times 6$ arrays of chains on rigid substrates. Each chain contained six monomers and was treated with the united atom method. By varying the interaction strength across the chain, a mechanism in which accumulated energy dissipation during shear leads only to a change in the tilt angle of the monolayer chains. The authors found that energy dissipation during shear (i.e. friction simulations) at 100 m s$^{-1}$ occurs through oscillations in the collective chain tilt angles similar to the plucking motion of UDTS and FDTS monolayers. By comparing the results for C16 and C21 chains, Tupper et al found little effect of chain length on the frictional properties, with a stronger effect on the structure of the monolayer itself under the applied load. The high shear rates here are a common issue in atomistic simulations of friction, as discussed in section 2.2.1.

5.5. The effect of chain length on friction

One of the most fundamental questions regarding the nanotribological properties of alkyllsine SAMs is related to the effects of the length of the carbon backbone on the adhesion and friction. Because order in the monolayer is intimately tied to the chain length of the precursor molecule, these effects can be difficult to disentangle experimentally. We will begin by focusing on work aimed at understanding the effects of chain length, and then move on to separate attempts at revealing the effects of monolayer disorder.

Early work by Salmeron’s group [306] was performed to study the differences between the frictional response of Cn ($n = 3, 6, 8$ and 18) alkyllsine chains on mica and probed with a Si$_3$N$_4$ tip. For chains with $n < 8$, friction forces were found to be larger than that measured for bare mica surface at the same load. On the other hand, for C8 friction was lower than that of bare mica and even lower for C18 chains. These data demonstrated that the chemical nature of the interface is not the only surface property that determines frictional response. However, as all data presented were taken with the same tip there are inherent issues in the interpretation of the results with regards to both blunting of the tip during shear as well as contamination from the substrate coating. Some of these issues are demonstrated in an extended friction curve for C18 chains (with a fresh tip) that extends beyond 300 nN of applied load. This plot shows distinct regimes of the friction response corresponding to an elastic response of the coating, plastic deformation and displacement of the coating, contact of the tip with the substrate and substrate damage (see figure 17). Nevertheless, it can be still concluded from this work that the increasing order that corresponds to increasing chain length promotes reduction of friction.

Building on this work, Salmeron et al [296] reported another study of the chain length effects on friction for both
alkylsilanes on mica and alkanethiols on gold. The alkyllsilane system in general showed similar trends to the previous work although the C12 chains (which were new to this study) showed slightly smaller friction than the C18 monolayer. The alkanethiol/gold system demonstrated another interesting trend. Friction for the C6 thiol was high, but decreased for C8. Friction of the C11 monolayer was nearly the same as that of C8. The C12 thiol showed lower friction still, but when the length of C18 was reached the friction increased to be nearly the same as C8 and C11. The authors used these results to again indicate increased disorder in the monolayer as being the main cause of higher friction. By separately comparing thiols and silanes at each chain length, it became clear that for \( n < 10 \) silanes show higher friction, while such systems are nearly identical for \( n > 10 \). The thiols show long-range order, while the silanes do not. Therefore the implication is that as the order in the silanes increases (due to saturation of inter-chain van der Waals interactions at around \( n = 8 \)–10) the friction decreases to meet that of the more ordered thiols. As for the lower friction seen in the C12 chains in both systems, the authors offer the possibility that this effect is due to sample purity, but they also admit that the observed differences are not yet understood. In light of recent work, it is likely that these differences are, in fact, due to the native chain length dependence rather than sample purity (see below). Note that these papers refer to the effect of chain length on friction force, not friction coefficient. Extraction of slopes from the reported figures indicates that, at least for alkanethiols on gold, the friction coefficient changes non-monotonically as a function of chain length, with smaller values corresponding to intermediate lengths.

The same C12 systems (i.e. alkanethiols on Au(1 1 1) and alkyllsilanes on oxide covered Si(1 0 0)) were later studied with IFM in order to compare a third type of SAM (hereafter alkoxyl) in which alcohol precursor molecules chemically bind to Cl terminated Si(1 1 1) [287]. The friction force and coefficient were significantly larger for the silane chain than for the thiol one, an effect that was ascribed to better ordering in the thiols. The alkoxyl system, designed to enhance both reproducibility in film formation and stability in humid environments, shows a frictional response essentially identical to the silane system. Further studies of the alkoxyl SAMs with IFM examined the chain length dependence of the friction [187]. A comparison of C18 thiols and alkoxyls revealed little difference in the friction, indicating (along with Fourier transform infrared spectroscopy (FTIR), water contact angles and x-ray photoelectron spectroscopy (XPS)) that the latter monolayer was highly ordered. When chain length was varied, C6 alkoxyls showed the highest friction, with C12 being lower, and C18 the lowest. These results were consistent both in experiment with a bare tungsten tip and a tungsten tip modified with a C8 silane. The authors indicate that these results are due to order in the monolayer, and fit their results to contact mechanics models, finding it necessary to fit the C6 results with a JKR model, while using a DMT model for the C12 and C18 results. This discrepancy is possibly due to a difference in material properties as a function of chain length, but more likely points to limitations of the continuum level models (see section 3.2).

Similar studies of the chain length on the friction of alkanethiols on gold were reported in [307], where it was found that increasing \( n \) leads to decreasing friction force and friction coefficient. This effect was correlated with increasing disorder through the bandwidth of the methylene stretching mode. Later work from this same group [308] found decreasing \( \mu \) with increasing \( n \) in alkanethiols on gold, but with the twist that a strong odd–even effect was seen with the friction increasing as \( \mu_{C13} < \mu_{C16} < \mu_{C13} < \mu_{C14} < \mu_{C12} \). The odd–even effect is discussed in section 5.8. In 1999 Li et al [309] also found \( \mu \) decreasing with increasing \( n \) while calibrating lateral and normal forces in situ.

More recent work on the chain length effect has produced slightly different results. Sambasivan et al [310] used a combination of the near edge x-ray absorption fine structure (NEXAFS) and FTIR spectroscopies to examine the structure of alkyllsilane films on Si with \( n \) varying from 5 to 30. In particular, the authors claim that for \( n = 12, 16, 18 \) their monolayers are highly ordered. The authors find friction coefficients that correlate well with the inverse of the dichroic ratio (via NEXAFS), but interestingly find the lowest \( \mu \) at intermediate chain lengths, with \( n = 12 \) showing the lowest friction. This work matches well with some previous experiments [306] and predictions from recent simulations on well-ordered systems [311].

Another recent paper reports the effects of chain length studied by using two-dimensional asperity arrays on Si wafers [312]. While the authors state their results as indicating \( \mu \) decreasing with increasing \( n \), their figure 7 indicates that the measured \( \mu \) for \( n = 18 \) is likely higher (within error bars) than that for \( n = 14 \), as seen in previously discussed experiments [306, 310] and simulations [311].

5.6. Chain length effect in simulations

While experiments reveal the nature of actual systems, the almost unavoidable conflicts between deposition quality and chain length make it extremely difficult to determine the true relationship between chain length and frictional response. Computer simulations, while suffering from many limitations of their own, have no such inherent difficulty and are prime candidates for unraveling the chain length dependence of friction. The phenomenal increase in available computational power over the past two decades has allowed for the growth of simulations of SAM friction from the initial work studying pairs of 6 × 6 arrays of non-specific united-atom chains [303] to all-atoms simulations of realistic AFM tips in contact with SAM coated substrates 50 nm on a side.

While Tupper and Brenner’s early simulations were performed on alkanethiols of two different chain lengths [304], the first work specifically aimed at comparing the frictional response of differing chain lengths came from Harrison et al in 1995 [255, 313]. In these papers, the authors reported the effects of hydrocarbons with \( n = 1, 2, 3 \) bonded to one-eighth of the sites on a diamond (1 1 1) surface in contact with another flat diamond (1 1 1) surface. Remaining sites on the surfaces were terminated with hydrogen atoms. Friction coefficients (here arrived at by dividing friction force by normal
force—a procedure which does not take into account shifts due to adhesive interactions) were compared as a function of normal load for the various chain lengths. In the simulations, longer chains exhibited lower friction coefficients, an effect that was ascribed to the increasing ability of longer chains to avoid collisions by manoeuvring around hydrogen atoms on the opposing surface. Continuation of this work was reported in 2000 [314] where the role of defects in the monolayer was studied by simulating 36 chains sliding against a diamond surface. This work, while finding a weak dependence of friction force on applied load, agreed with Salmeron’s assertion [306] that the formation of gauche defects in the monolayer is a major energy dissipation mechanism during sliding friction. Furthermore, the authors showed that at a given load shorter chains develop more defects than the longer ones, and offer this phenomenon as an explanation for the chain length dependence of the friction. Later, in a more detailed study, it was shown that the contribution to total energy from these defects reaches a plateau as the applied load increases, lessening their significance [315] as a dissipation channel. This conclusion was echoed in other simulation work [316] on well ordered, fully packed alkylsilanes on crystalline SiO2. The latter work was performed at lower pressures and, through energy decomposition, was able to demonstrate that while gauche defect formation is strongly correlated with slip-events, the dominant contribution to the total energy is due to Coulomb interactions.

Fujihira et al performed a series of simulations based on Glośli and McLelland’s model studying, among other things, the chain-length effect [317–319]. Additionally, the authors performed calculations using a phenomenological model that describes a system sliding at infinitesimally small velocity at 0 K. In this model SAMs are represented by two potentials, an interfilm potential $V_c$ and a deformation potential $V_d$. The two potentials $V_c$ and $V_d$ were determined by assuming their functional form and fitting free parameters to simulation results. In the chain length studies, it was found that $V_d$ did not vary for chains between C6 and C18 when the monolayer was assumed to be perpendicular to the substrate. As the chains were tilted by varying the lattice spacing, differences in $V_c$ arose for the different chain lengths. The effect of chain tilting due to change in the lattice spacing has also been observed in other systems, e.g. alkanethiols on gold and silver surfaces [291]. The above results lead to at least a qualitative explanation of the observed chain length dependence, but the authors caution that more experimental and theoretical work must be completed to fully understand the phenomena.

Hybrid simulations were performed using a small model of an AFM tip to study the chain length dependence of united-atom alkanethiols on gold at low temperature (e.g. 0.1 K) [147] and at 300 K [148] both with a velocity of 400 nm s$^{-1}$. The hybrid scheme was justified based on differences in the resonant frequencies of the tip and the SAM and effectively the tip was moved with a large time step followed by a short relaxation of the SAM. At both temperatures larger friction force is found for shorter chains and in both cases there is a non-linear dependence of the friction force on the applied normal load.

An all-atom study of friction between pairs of fully packed, well-ordered alkylsilane SAMS on crystalline silica found a different chain length dependence than what is generally accepted [311]. This work found lower $\mu$ for the intermediate chain lengths of C8 and C12, with higher $\mu$ for the shorter and longer lengths of C6 and C12, respectively. These authors claim that this is an indication of multiple dissipation mechanisms at work, each of which has a different chain length dependence. Such interplay of different mechanisms can result in high dissipation at intermediate chain lengths at all loads, leading to a low friction coefficient even though the friction force itself is high.

With this large body of work taken as a whole, it is difficult to arrive at a single, satisfying conclusion as to the effect of chain length on friction. The majority of researchers, both from experiment and from simulation, find that friction force decreases with increasing chain length and many find the same behaviour for friction coefficient as well. Some more recent work, again both from experiment and simulation, disagrees with this conclusion and finds a native chain length dependence that is non-monotonic. In general, researchers from both camps agree that (i) defects increase with decreasing chain length and (ii) an increase in the number of defects leads to an increase in friction. One of the reasons for the discrepancies in reported studies is the inherent difficulty in making high quality monolayers of short chains (or truly of most chain lengths) and it is clear that the differences in monolayer quality have a major effect on friction measured in experiments. Interpretation of simulations is not much easier. All simulations suffer from some deficiencies, some of which are unavoidable and some of which are not. The issue of simulation size has been shown to have a negligible effect at least in some simulations, such as those that compared friction of 10 $\times$ 10 chain systems with 20 $\times$ 20, and 40 $\times$ 40 [316]. While a reduction in stick–slip behaviour was seen in larger systems (due to the difficulty of moving such system as a whole), the major effect from increasing system size was a reduction in noise of measured quantities. For smaller systems noise reduction can be accomplished by means of appropriate time-averaging methods. The issue of simulation speed and related shear velocity is much more serious. AFM experiments are performed at velocities that generally peak under 10 $\mu$m s$^{-1}$. MD simulations, on the other hand, are generally now performed with shear velocities around 1 m s$^{-1}$, with occasional forays into speeds as low as 1 cm s$^{-1}$ [311, 320]. In fact, many older simulations have quoted velocities in the hundreds of m/s. The hybrid simulations [147, 148] have attempted to overcome this limitations, but their success is open to debate. The issue of velocity in friction simulations is one that will likely not be overcome in the near future, as simulation time essentially depends on processor speed. A 10$^4$ increase in processor speed is, unfortunately, not expected in the near future. While it is often observed that friction increases in some way with increasing shear velocity [41, 270], it is not necessarily true that all chain lengths are affected equally, and the physical origins of this phenomenon are yet to be fully explored. As an
example, the melting point of alkanes, with chain lengths similar to the molecules studied in SAM friction, is close to room temperature. The addition of heat through shear-induced vibrations could potentially lead to ‘melting’ of SAMs made of shorter chains and not the longer ones. Thus it is not always clear how high speed simulations relate to slow speed experiments. Despite their limitations, the ability to track individual molecules and energetic contributions/pathways still make atomistic simulations extremely useful for this type of study.

The final major limitation of simulations is that most studied thus far have been performed for infinite flat plates in contact. Such geometries are in stark contrast with experimental setups, where the tip is often curved or pyramidal. This difference between simulations and experiments introduces additional complications into the interpretation of results, such as the exact dependence of contact area on applied load. Some work has already been performed in attempt to model realistic tips, and such simulations will allow for more realistic comparisons with experiments.

5.7. Packing effects on friction

Because of the difficulty in separating chain length and disorder in experimental systems, Perry [321] developed another method to probe the effects of disorder on friction in SAMs. In this approach a mixture of standard alkanethiols \((n = 17)\) with spiroalkanedithiols \((n_1 = 17, n_2 = 1)\) was used on a gold substrate. In spiroalkanedithiols two chains with separate head groups and tail groups are bonded just above the thiol and \(n_1\) and \(n_2\) refer to the lengths of the two tail groups. As a result, the chain length was kept constant, but the disorder was varied. The order of the pure \(n = 17\) system was the highest, followed by the mixed \(n = 17\) with \(n_1, n_2 = 17\), and finally by the mixed \(n = 17\) with \(n_1 = 17, n_2 = 1\). This work demonstrated that both the friction force and coefficient increase with increasing disorder in the films. A more detailed set of experiments was performed subsequently [290], in which \(n_1 = 17, n_2 = 2\) or 12. It was shown that for \(n_1 = 17\), the value of \(n_2\) is essentially irrelevant, i.e. disorder added by the second chain had no effect on friction versus load curves. This conclusion was later confirmed by detailed simulations [311], in which friction was studied among pairs of ordered monolayers on crystalline silica in which 0%, 10%, 30% or 50% of chains had been removed (0% corresponds to fully packed, well-ordered SAM). In the same work similar simulations were carried out on amorphous silica substrate with 30% and 50% chains removed and the results were comparable to those on crystalline silica. No chain length dependence of friction force and coefficient was found for systems with 10% or more defects. The friction force increased with increasing defect density, while the coefficient remained nearly constant. Similar trends were found in a different set of experiments, where friction of higher and lower density phases of C18 chains on silica was measured [77]. The main conclusion both from the experiments and the simulations is that it is the disorder present at the sliding interface that determines the frictional response. Another simulation was carried out to address the dependence of friction on defects in hydrocarbon chains bonded to diamond surfaces [315]. In this work higher friction was found in a system with 30% of the chains removed than in the pristine system. This effect was attributed to more energy dissipation through bond-length fluctuations, which in turn was enabled by the lower packing density.

Packing effects have also been studied by comparing tribological properties of hydrocarbon films with those in which hydrogen atoms are either partially or entirely replaced with fluorine atoms. Beyond shedding light on the molecular origins of friction, this type of study has direct applications in MEMS, where perfluorinated chains are used as friction modifiers because of their ability to withstand higher processing temperatures. Early experimental work demonstrated that CF3 termination alone leads to a factor of three increase in the frictional response [322]. This work was extended to systems with mixed CF3 and CH3 terminated chains which demonstrated that even the addition of a small amount of CF3 chains results in an increase in friction, and that this increase was unrelated to the percentage of CF3 chains. This result again is echoed in the simulations, which show that after the change in friction due to adding initial defects, further increasing the density of defects does not affect coefficient of friction [311]. The parallel is drawn here because, similarly to the addition of defects, the addition of bulkier CF3 chains lead to an increase in disorder.

In order to understand the increase in friction in pure CF3 films (as compared with pure CH3 ones), Ohzono et al [318] applied their phenomenological model, which had been earlier applied to study chain length effects, to the CF3 systems. It was found that the increased friction was possibly a result of increased interfilm interactions when CF3 molecules are present. Based on a different set of simulations, Park et al [323] postulated that it is more likely the tighter packing density in purely fluorinated films on gold that leads to higher friction. This is more in agreement with the earlier experimental results [324], as well as later experiments by Li et al [325] on mixed CH3 and CF3 terminated systems in which the chain lengths were varied as well. Li et al however, identify the size of the terminal CF3 group as being the major cause of higher friction. More recent experiments on alkanephosphonic acid SAMs on alumina agree with the packing density argument [70].

The effect of differing end groups has also been studied in the context of hydrophilic systems with either COOH or OH end groups. In early work, Frisbie et al [301] found higher friction between self-mated pairs of monolayers (i.e. COOH/COOH or CH3/CH3) as compared with a non-mated pair (i.e. COOH/CH3). The difference was particularly notable in the adhesive interactions. Similar results were seen in later experiments [326, 327] and simulations [328] (which did not study friction). The general conclusion seems to be that the increased intrafilm bonding leads to higher friction, although given the unclear relationship between adhesion and friction, more work in this area is needed.
The energy barrier for bond breaking is reduced by the contact stress so that the number of defects can be described as

\[ N_{\text{def}}(L) = t_{\text{res}} n_0 A(L) v \exp \left( - \frac{E_{\text{act}} - \Delta E(L)}{k_B T} \right), \]

where \( t_{\text{res}} \) is the residence time of the AFM tip, \( n_0 \) is the surface density of atoms, \( A(L) \) is the contact area which depends on the applied load \( L \), \( v \) is the attempt frequency to overcome activation energy \( E_{\text{act}} \) to break the Si–O bond, and \( \Delta E(L) \) is the change in activation energy due to the applied stress. \( \Delta E \) can be evaluated from the contact pressure and by assuming Hertzian contact mechanics. It was argued that when the density of defects \( [N_{\text{def}}/n_0 A(L)] \) reaches a critical value \( \eta_{\text{crit}} \), the defects grow to form visible wear scars \(~2 \text{ Å}~\) deep. With these assumptions, the number of AFM scans \( N_{\text{scans}} \) multiplied by the defect density has to equal \( \eta_{\text{crit}} \) and the following relationship can be derived for the number of scans necessary for visible damage to appear:

\[ N_{\text{scans}} = A_0 L^{-2/3} \exp \left( - B_0 L^{2/3} \right), \]

where \( A_0 \) and \( B_0 \) are constants. Figure 18 shows that such relationship can be fitted to represent the experimental data. Similar study for mica in aqueous environment was performed by Helt et al. [333].

A beautiful study of atomic scale wear has been reported by Gnecco et al. [334] who investigated KBr (0 0 1) surface with Si SFM tip in UHV. As a result of sliding, the lateral force increased continuously with the number of scratches and it was demonstrated that the wear mechanism is due to removal and rearrangement of single ionic pairs. The debris generated during scratching rearranged in an epitaxial process and formed regular terraces with the same periodicity and orientation as the underlying substrate. Studies of nanoscale wear were performed also for polymers. Gotsmann et al. [335] used a heated cantilever tip to investigate a thin film of polystyrene. The type and amount of wear were strongly dependent on temperature. Below the glass transition temperature surface ripples were generated during sliding. Ripple formation was attributed to the system moving between metastable states of the glass with activation energy for this process estimated to be \(~0.4 \text{ eV}~\). The formation of ripples is largely enhanced in the glass transition regime. At higher temperatures, no ripples were observed, but the polymer was transported along with the moving tip.

If wear is accompanied by material transfer between the tip and the substrate, the mechanisms and amount of transferred material will depend on the relative humidity and applied load, and will differ from material to material [69]. The amount of wear may also depend on bulk chemistry of the materials in contact, which can be understood by the fact that wear often occur when strength of the bonds across the interface is large relatively to the bonding in the bulk. Bulk chemistry has been shown to have a dramatic effect on wear of DLC films, which are amorphous films with a mixture of sp² and sp³ bonding stabilized with H atoms [36, 336]. Interpretation of such experiments is challenging because it is often difficult to decouple the sp²/sp³ ratio and the H content in the sample.
Initial AFM experiments by Jian et al. [336] showed that H concentration leads to more wear generated in DLC. The H concentrations considered in this study varied between $\sim 2\text{at}\%$ to 40 at\%, however the $sp^2/sp^3$ ratios were unfortunately not reported. More recent studies [337] indicated that an optimal H concentration may exist. There is a need for further comprehensive studies that would consider varying H concentration and $sp^2/sp^3$ ratio separately and that would establish individual contributions from these parameters to the mechanical and tribological behaviour.

Recently, the development of *in situ* TEM nanoindenters, discussed briefly in section 2.1, has created an exciting opportunity for understanding wear phenomena at the nanoscale. Ribeiro et al [91] used this *in situ* technique to directly observe abrasive nanoscale wear of a Si substrate probed with a gold-coated diamond indenter. It was determined that wear (material detachment) was associated with cracks forming under the indenter. The authors hypothesized possible mechanisms for wear initiation: (i) propagation of voids, (ii) fracture and (iii) nucleation and motion of dislocations. Based on the analysis of energetics involved in the three aforementioned mechanisms, it was suggested that most likely it is the formation of voids along the elastic contour under the indenter that initiates wear. Subsequently, these voids coalesce into cracks and provide nucleation sites for dislocations. This mechanism was consistent with high-temperature nanoindentation results of Schuh et al [338], who proposed that vacancies and vacancy clusters precede dislocation nucleation during loading.

Systematic studies of wear need to be combined with state of the art imaging techniques so that atomistic mechanisms of wear can be fully characterized. On the simulation side, the MD technique has been the primary computational tool in studies of wear because material transfer at the interface cannot be described by means of continuum mechanics. However, most of the simulations of wear have been performed with small tips or flat slabs sliding past one another, which makes comparison with experiments difficult. More simulations, particularly those with realistic tip geometries and sizes, are therefore needed in order to unravel the complex phenomena of wear and its relation to friction. Additionally, for understanding macroscopic wear it is essential to extend single-asperity studies to multi-asperity contacts as wear processes will continually change the landscape of the interface and they will affect subsequent wear behaviour as well as friction [339–341].

**Figure 18.** (a) Number of consecutive scans necessary to cause visible damage to the mica surface (2 Å holes). Data from two series of experiments, measured with the same tip (force constant 0.4 N m$^{-1}$) on two consecutive days, are shown. (b) The same data in a semilog plot (see axes labels). Reused with permission from [332] Kopta S and Salmeron M 2008 *J. Chem. Phys.* **113** 8249. Copyright 2008 American Institute of Physics.

7. Future prospects

Despite the remarkable progress that has been made in nanotribology, there is still much work to be done. For example, when the size of a single-asperity reaches a few nanometres, processes underlying friction and wear cannot be fully understood without help from atomistic simulations such as those based on the MD technique. However, in order to be able to compare results obtained in computer simulations and in SFM experiments, the gaps in length- and time scales that currently exist between the two approaches needs to be closed. While the so-called spatial decomposition of simulated systems reaches across the length scale gap, there has been no universal remedy proposed to address the time scale and shear rate limitations of MD simulations. Much faster sliding speeds in SFM experiments would help to close this gap as well. Experimental techniques have other challenges when it comes to single-asperity contact measurement and bridging the length scale gap. Specifically, AFM experiments are in need of better characterization and control of the atomic level topography and composition of the tip, since atomic scale roughness and chemistry have been shown to have a large effect on the contact stresses, adhesion and friction. Therefore it is difficult to draw general conclusions unless the tips used in experiments are fully characterized at the atomic level before and after the experiment, or even better, *in situ*, i.e. during the experiment.

Ultimately, one would like to be able to understand and predict friction, adhesion and wear based on the bulk and surface properties of the two materials in contact. For this goal to be attained, more simulations and experiments need to be performed at well-defined interfaces to establish relevant energy dissipation mechanisms and their relative contributions to friction. Physical mechanisms that have so far received
little attention in the literature yet need further investigations include electronic and phononic contributions to friction, and irreversible processes such as bond breaking, defect creation, material mixing and tribochemical reactions [342]. Even for mechanisms that have been studied in greater detail there are still many fundamental questions to be addressed, such as why stick–slip occurs with the lattice periodicity of the substrate for incommensurate surfaces with multiple-atom contacts? Following the pioneering work by Robbins and others [52, 53], who employed simulations based on the LJ interatomic potential to demonstrate a break-down of continuum theories contact, the limits of these theories need to be determined for more realistic tips in various chemical environments. Prediction of friction between two materials will only be possible if fundamental rules are established for relationships between friction and other mechanical properties.

Connecting nanotribology to macroscopic behaviour presents a particularly exciting and so far largely unexplored area of research. There is a need for experiment, theory and simulation to study the sliding of surfaces with well-defined roughness at different length scales for specific materials and environments, so the gap between single-asperity contact and macroscopic tribology can be bridged. Given that all macroscopic sliding interfaces involve some degree of wear, it is crucial that atomic studies further explore the origin and progression of wear to establish the underlying mechanisms and ways to control it. Clearly, connecting nanoscale and micro-/macroscale friction behaviour still remains a challenge, and this provides ample motivation for much further work, which will require creative, quantitative and thoughtful studies.

Acknowledgments

IS gratefully acknowledges support from the National Science Foundation grant DMR-0512228 and from the Air Force of Scientific Research grant FA9550-07-1-0126. RWC gratefully acknowledges support from the Air Force Office of Scientific Research under contract # FA9550-05-1-0204, and the National Science Foundation under contract # CMMI-0700351 and CMMI-0733585. Some of this work was performed at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy’s National Nuclear Security Administration under Contract DE-AC04-94AL85000.

References

[3] Flater E E, Corwin A D, de Boer M P and Carpick R W 2006 In-situ wear studies of surface micromachined interfaces subject to controlled loading Wear 260 580
[16] Hutter J L 2005 Comment on tilt of atomic force microscope cantilevers: effect on spring constant and adhesion measurements Langmuir 21 2630


[37] Hertz H 1881 On the contact of elastic solids J. Reine Angew. Math. 92 156


[40] Christenson H K 1993 Adhesion and surface energy of mica in air and water J. Phys. Chem. 97 12034


[46] Derjaguin B V, Muller V M and Toporov Y P 1975 Effect of contact deformations on the adhesion of particles J. Colloid Interface Sci. 53 314


[52] Luan B and Robbins M O 2005 The breakdown of continuum models for mechanical contacts Nature 435 929

[53] Luan B and Robbins M O 2006 Contact of single asperities with varying adhesion: comparing continuum mechanics to atomistic simulations Phys. Rev. E 74 026111


[61] Sader J E and Green C P 2004 In-plane deformation of cantilever plates with applications to lateral force microscopy Proc. R. Soc. Lond. A 463 303


[64] Bloo M L, Haintjema H and Pril W 0 1999 Deformation and wear of pyramidal, silicon-nitride AFM tips scanning micrometer-size features in contact mode Measurement 25 203


[66] Tao Z and Bhushan B 2006 Surface modification of AFM tips for adhesion and wear reduction Tribol. Lett. 21 1


[70] Qian L M, Xiao X D and Wen S Z 2000 Tip in situ chemical modification and its effects on tribological measurements Langmuir 16 662


[72] Miyake S and Asano M 2000 Nano processing by hard film coating and movement of processed nanosider supported by van der Waals forces J. Japan Soc. Precis. Eng. 66 1275


[89] Schuh C A 2006 Nanoindentation studies of materials Mater. Today 9 32


[100] Pietrement O and Troyon M 2002 Quantitative study of shear modulus and interfacial shear strength by combining modulated lateral force and magnetic force modulation microscopies Surf. Interfacial Anal. 31 1060


[105] Landman U, Luedtke W D and Ringer E M 1992 Atomistic mechanisms of adhesive contact formation and interfacial processes Wear 153 3


[113] Foster A S 2004 Realistic model tips in simulations of nc-AFM Nanotechnology 15 S60


[118] Maugis D 1992 Adhesion of spheres—the JKR-DMT transition using a Dugdale model J. Colloid Interface Sci. 150 243


[120] Szlufarska I 2005 Atomistic simulations of nanoindentation Mater. Today 9 42

[121] Tomagnini O, Ercolessi F and Tosatti E 1993 Microscopic interaction between a gold tip and a Pb(110) surface Surf. Sci. 287–288 1041


[151] Tabor D 1977 Surface forces and surface interactions J. Colloid Interface Sci. 58 2


[298] Suzuki M 1988 Characterization of monolayer and bilayer (polymer monolayer) structures for their use as a lubricant Thin Solid Films 160 453


[318] Fujihira M 1999 Bases of chemical force microscopy by friction: energetics and dynamics of wearless friction between organic monolayers in terms of chemical and physical properties of molecules Japan. J. Appl. Phys. part 1 (Regular papers & short notes) 38 3918


[320] Lee S 2000 The influence of packing densities and surface order on the frictional properties of alkane-thiol self-assembled monolayers (SAMs) on gold: a comparison of SAMs derived from normal and spiroalkanedithiols Langmuir 16 2220


[335] Gotsmann B and Durig U 2004 Thermally activated nanowear modes of a polymer surface induced by a heated tip Langmuir 20 1495


