1 Introduction

Friction is a well-known, but poorly understood, phenomenon that affects virtually all aspects of daily life. In some cases, friction is desirable, e.g., high friction in clutches leads to the effective transmission of forces between an automobile’s engine and its wheels, while in other cases friction is a significant drawback, e.g., friction between the piston and cylinder wall decreases the efficiency of automobile engines. Although macroscopic friction laws were introduced a few centuries ago,¹ and the existence of friction was recognized long before that, the underlying atomic-level mechanisms leading to friction have remained elusive. The identification of these mechanisms has emerged as a topic of significant interest, which has been driven by the miniaturization of mechanical devices, the peculiar behavior of condensed matter at the nanoscale, and advances in simulating chemically complex lubricants and surfaces with ever-increasing accuracy.²–⁴ Although a great deal of research has been directed towards elucidating the fundamental, atomic-level origins of friction in recent years, many key questions remain unanswered.

Atomic-level simulation has been used extensively in the study of friction, not simply as a means of supplementing experimental studies, but as a powerful tool for gaining unique insight into the relevant processes. Indeed, simulations allow one to study well-defined systems under a variety of conditions that may be difficult, or even impossible, to examine in real laboratory experiments. As such, simulations have shed much-needed light on fundamental aspects of friction, and in some cases have even overturned conventional wisdom regarding the origins of friction. In this chapter,
we discuss the key points associated with performing tribological simulations (tribology is the science of surfaces in relative motion, sometimes also defined as the science of friction, lubrication, and wear) and review representative studies in which such simulations have been applied.

Tribological simulations inherently involve studying systems that are far from equilibrium. As such, many general principles, such as the minimization of free energy, no longer apply. Along similar lines, the equivalence of ensembles in the thermodynamic limit does not apply to many tribological systems, and hence the proper use of boundary conditions is crucial. This is illustrated by the following example, where we consider two simulations of a system composed of two sliding surfaces performed with different boundary conditions. In one simulation, the kinetic friction force, \( F_k \), is determined at constant load and constant sliding velocity. In the second simulation, similar parameters are employed; however, instead of applying a constant load, the distance between the surfaces is constrained to a constant value. This corresponds to performing the simulation at constant separation and constant sliding velocity. These two simulations are likely to yield completely different values for \( F_k \), which demonstrates that implementing boundary conditions properly is crucial in non-equilibrium simulations if one wants to make reliable predictions. It also illustrates one of the many pitfalls that can diminish considerably the value of a largely well-designed tribological simulation. These pitfalls often result from convenience, where the unrealistic treatment, in this case constant separation, is easier to implement than the experimental condition, that is, constant normal load. Additional considerations relating to non-equilibrium simulations arise from the fact that imposing shear and load requires one to constantly pump energy into the system. This energy is converted into heat and must be removed during the simulation, which is achieved through the use of thermostats. However, non-equilibrium simulations can be more sensitive to temperature control than equilibrium simulations and applying thermostats in a naive manner may induce unrealistic velocity profiles in the sheared lubricant or lead to other undesired artifacts. The topic of non-equilibrium molecular dynamics has been reviewed previously in this series.\(^5\)

Another important issue associated with tribological simulations involves the definition of the
system to be studied. For example, a simple tribological system consists of two atomically-flat, defect-free surfaces sliding past one another. Due to computational convenience, it is common practice to orient the surfaces in a commensurate manner. That is, the surfaces are in perfect alignment and share a common periodicity. However, real engineering contacts rarely contain commensurate surfaces; indeed, it is exceedingly difficult to purposely devise such systems. Moreover, the frictional properties of systems that contain commensurate surfaces differ significantly from those of systems with incommensurate surfaces, i.e., surfaces that do not share a common periodicity in a systematic fashion. Thus, in order to obtain meaningful results from simulations, one must take care to ensure that the systems being studied do not contain commensurate surfaces, or other artificial symmetries. The design of realistic systems is further complicated by the presence of defects, surface curvature and roughness, lubricant molecules and surface contaminants, all of which must be treated appropriately. For instance, preventing lubricant molecules from becoming squeezed out of contacts, or even neglecting atomic-scale surface roughness, may lead to erroneous results in simulations. In some cases, it is even necessary to consider the chemical reactivity of the lubricant molecules.

In this chapter, we discuss how to perform meaningful tribological simulations by avoiding the potential pitfalls that were mentioned above. In Section 2, some theoretical aspects of friction between solids will be explained. Section 3 contains an overview of algorithms that have been used in the simulation of tribological phenomena. Selected case studies will be presented in Section 4.

## 2 Theoretical Background

Every-day experience indicates that a finite threshold force, namely the static friction force, $F_s$, must be overcome to initiate lateral motion of one solid body relative to another. Experience also indicates that a finite minimum force, namely the kinetic friction force, $F_k$, must be applied in order to keep a solid body moving at a constant velocity. In plain terms, we must expend a certain minimum amount of energy to overcome friction if we want to move an object and keep it moving.
Figure 1: Force against time for a sliding system. Sliding is initiated once the threshold corresponding the static friction force is surpassed when $F/F_s = 1$.

This contrasts with the situation encountered when one pulls a solid through a fluid medium where there is no such threshold, and instead one only needs to overcome frictional forces that are linear in the final velocity, $v_0$. The notion of a threshold force is indicated in Figure 1, where the applied force is increased with time until the static friction force is overcome, i.e. $F/F_s = 1$ and sliding is initiated. In the sliding regime, the applied force is lower than that required to initiate sliding, and corresponds to the kinetic friction force, $F_k$.

In this section, we provide a short introduction to the topic of friction. Without such a background, it is difficult to ask meaningful questions and interpret the outcome of simulations. After all, our goals extend beyond simply reproducing experimental results. Moreover, a good understanding of the theoretical background will aid in determining which aspects of a simulation deserve particular focus and which details are essentially irrelevant.

The discussion begins in Section 2.1 with an overview of proposed energy dissipation mechanisms that lead to friction. This is followed by brief discussions of phenomenological friction laws that describe the dependence of friction upon normal load (Section 2.2) and sliding velocity (Section 2.3). The dependence of friction on the symmetry of the surfaces that are in contact is discussed in Section 2.4.
2.1 Friction Mechanisms

Given the ubiquitous nature of friction in our daily lives, it came as a surprise when Hirano and Shinjo suggested that friction between solids in ultra-high vacuum may essentially disappear.\textsuperscript{6,7} Although this proposal clearly conflicts with our intuition, it does not necessarily violate Newtonian mechanics. Consider the following case in which a slider is dragged over a surface. If the slider and substrate have homogeneous surfaces and if wear and plastic deformation are negligible, one may expect the same (free) energy at the beginning of the sliding process as at its end because of translational invariance. Thus, no work will be performed on the system, which opens up the possibility of ultra-low friction. This situation is commonly referred to as superlubricity and is currently an active area of research.

The demonstration that superlubricity is possible, at least from a theoretical standpoint, indicates that solids are not required to exhibit friction under all conditions. However, virtually all surfaces do exhibit friction as a result of various mechanisms that lead to deviations from the conditions necessary for superlubricity. In general, these mechanisms involve hysteresis and energy dissipation, which prevents the recovery of the energy that was expended during sliding. In what follows, we provide a brief overview of common friction mechanisms.

It has long been recognized that solid friction is intimately connected to hysteresis, as best illustrated by the model proposed independently by Prandtl\textsuperscript{8} and Tomlinson.\textsuperscript{9} In this model, a surface atom of mass $m$ is coupled to its lattice site via a harmonic spring of stiffness $k$. The lattice site, which moves at constant velocity $v_0$, is assumed to be located at the origin at time $t = 0$. In addition to the interaction with its lattice site, the atom experiences a coupling $V_0 \cos(2\pi x/a)$ to the substrate, where $V_0$ has units of energy and reflects the strength of the coupling, $a$ is the lattice constant of the substrate, and $x$ is the current position of the surface atom. Upon the introduction of a viscous damping term that is proportional to velocity $\dot{x}$ and a damping coefficient $\gamma$, the equation of motion for this atom is

$$m\ddot{x} + \gamma \dot{x} = k(v_0 t - x) + \frac{2\pi}{a} V_0 \sin(2\pi x/a).$$

[1]
Figure 2: Illustration of an instability in the Prandtl-Tomlinson model. The sum of the substrate potential and the elastic energy of the spring is shown at various instances in time. The energy difference between the initial and the final point of the thick line will be the dissipated energy when temperature and sliding velocities are very small.

If $k$ is very large, i.e., $k$ is greater than the maximum curvature of the potential, $V''_{\text{max}} = (2\pi/a)^2V_0$, there will always be a unique equilibrium position for the atom of $x_{\text{eq}} \approx v_0t$ and the atom will always be close to $x_{\text{eq}}$. Consequently, the friction will be linear in $v_0$ for small values of $v_0$.

Things become more interesting once $V''_{\text{max}}$ exceeds $k$. Now there exists more than one stable position at certain points in time, as shown in Figure 2. The time dependence of the combined substrate and spring potential reveals that mechanically stable positions disappear at certain points in time due to the motion of the spring. Consequently, an atom cannot find a mechanically stable position at time $t + \delta t$ in the vicinity of a position that was stable a small moment $\delta t$ ago. As a result, at times slightly past $t$, the position of the atom becomes unstable and it must move rapidly towards the next potential energy minimum. After sufficiently many oscillations around the new mechanical equilibrium, most of the potential energy difference between the new and old equilibrium positions will be dissipated into the damping term. This process will repeat itself periodically as the atom moves between potential energy minima at finite sliding velocities. Consequently, the dissipated energy per sliding distance is independent of $v_0$ and $\gamma$ (in the present example of a bistable system) for sufficiently small values of $v_0$.

Despite its merits, the Prandtl-Tomlinson model should not be taken too literally. There simply is no reason why the inter-bulk coupling, reflected by $V''_{\text{max}}$, should be stronger than the intra-bulk
coupling $k$. But even if it were, one would have to expect more dramatic processes than elastic instabilities, such as cold welding and plastic deformation, so that the assumption of elastic coupling in the slider would break down completely. One could certainly argue that similar instabilities involving collective degrees of freedom may occur on longer length scales. However, it appears that elastic instabilities do not contribute considerably to dissipation.\textsuperscript{10} A notable exception to this rule is rubber, for which sliding friction is related to internal friction rather than energy dissipation taking place at the interface.\textsuperscript{11}

A traditional explanation of solid friction, mainly employed in engineering sciences, is based on \textit{plastic} deformation.\textsuperscript{12} Typical surfaces are rough on microscopic length scales, as indicated in Figure 3. As a result, intimate mechanical contact between macroscopic solids occurs only at isolated points, typically at a small fraction of the apparent area of contact. The net area of this intimate contact is called the real area of contact, $A_{\text{real}}$. It is assumed that plastic flow occurs at most microscopic points of contact, so that the normal, local pressures correspond to the hardness, $\sigma_h$, of the softer of the two materials that are in contact. The (maximum) shear pressure is given by the yield strength, $\sigma_y$, of the same material. The net load, $L$, and the net shear force, $F_s$, follow by integrating $\sigma_h$ and $\sigma_y$ over the \textit{real} area of contact, $A_{\text{real}}$. That is, $L = \sigma_h A_{\text{real}}$ and $F_s = \sigma_y A_{\text{real}}$. Hence, the plastic deformation scenario results in the following (static) friction coefficient

$$\mu_s = \frac{\sigma_y}{\sigma_h}$$

where $\mu_s$ is defined as the ratio of $F_s$ and $L$. Although this explanation for a linear relationship between friction and load has been used extensively in the literature, Bowden and Tabor, who originally suggested this idea, were aware of the limitations of their model and only meant to apply it to contacts between (bare) metals.\textsuperscript{12} There are two important objections to the claim that plastic deformation is generally a dominant friction mechanism. First, friction between two solids does not typically depend solely upon the mechanical properties of the softer of the two materials in contact, but rather on the properties of both of these materials and the lubricant between them.
Second, theoretical calculations of typical surface profiles have shown that plastic flow should occur at only a very small fraction of the total number of contact points, and hence it is unlikely that plastic deformation contributes significantly to friction in real engineering contacts.\textsuperscript{13}

So far, we have not considered lubricants added intentionally, such as oils, or unintentionally in the form of contaminants such as airborne hydrocarbons. It is known that adsorbed molecules can alter the behavior of sliding contacts dramatically as long as these molecules remain at the microscopic points of contact.\textsuperscript{14,15} From an engineering point of view, such molecules prevent surfaces from making intimate mechanical contact, thereby reducing plastic deformation and wear. However, they also prevent surfaces from becoming superlubric. Under sliding conditions where there exists a sufficient quantity of lubricant, the last one or two layers of that lubricant remain in the contact, where they solidify due to the typically large pressures at the microscopic scale. In this regime, one generally talks about boundary lubrication, and because the interactions between lubricant particles are relatively weak, the adsorbed atoms and molecules will try to optimize their interactions with the confining walls. This can lock the surfaces geometrically, as illustrated schematically in Figure 4, and when sliding one surface relative to another, an energy barrier has to be overcome, which generates a static friction force.

There are many other mechanisms leading to energy dissipation, although they may be less universal than those related to boundary lubricant-induced geometric frustration. Chemical changes in lubricant molecules, reversible or irreversible, produce heat. Examples are configurational
changes in hydrogen-terminated diamond surfaces\textsuperscript{16} or the terminal groups of alkane chains\textsuperscript{17}, as well as sliding- and pressure-induced changes in the coordination numbers of surface or lubricant atoms.\textsuperscript{18,19} Although the microscopic details of these processes differ significantly, they all exhibit molecular hystereses which are similar to that described by the Prandtl-Tomlinson model.

There also exist many irreversible tribological phenomena, such as cold-welding, scraping, cutting, or uncontrolled, catastrophic wear. Characterizing these phenomena is often tedious because many of these processes are system specific and, in addition to being far out of equilibrium, they lack a steady state. For these reasons, we will not elaborate on these processes here.

\section{2.2 Load-Dependence of Friction}

Many macroscopic systems show an almost linear relationship between friction, $F$, and load, $L$,

$$F = \mu L,$$

where the friction coefficient, $\mu$, does not depend on the apparent area of contact. This linear dependence, which is routinely taught in introductory physics classes, is referred to as Amontons’ law.\textsuperscript{1} The origin of this relationship is the subject of a great deal of controversy.

Initial attempts to explain Amontons’ law focused on the nature in which surface asperities slide past one another. The basic argument put forward was that if an asperity protrudes from the surface at some angle, $\theta$, the force required to move past that asperity will be $F = L\tan(\theta)$. When one averages over all of the asperities on a surface, $\tan(\theta)$ will attain some constant value and the friction will be proportional to the load. Although this explanation seems to account for Amontons’ law at first glance, it was quickly pointed out that once an asperity reaches the top of
another, it will slide down the other side and the applied energy will be recovered. Since, in a statistical sense, the average value of \( \theta \) will be constant irrespective of the sliding direction, there is no reason to anticipate any net loss of energy, and friction should essentially be non-existent in such a scenario. In fact, similar arguments have been used to support the possibility of low-friction situations, such as superlubricity.

More recently, arguments for the origin of Amontons’ law have arisen that are based on experimental studies demonstrating that the shear stress, \( \sigma_s \), varies with the local pressure, \( P \), according to Eq. [4]

\[
\sigma_s = \sigma_0 + \alpha P \tag{4}
\]

where \( \sigma_0 \) and \( \alpha \) are constants.

The reason for the linear relationship between shear and normal pressure, even in the presence of adsorbed atoms, can be rationalized qualitatively by considering Figure 4. For the top wall to move to the right it must move up a slope, which is dependent upon how the adsorbed atom is interlocked between the substrate (bottom) and the slider (top). If the system remains rigid, this will lead to \( F = L \tan(\theta) \), and hence a linear relationship between friction and load as discussed above. Of course, this argument is highly qualitative, because it assumes implicitly that non-bonded atoms behave like hard disks in areas of high pressure. Moreover, this argument must be modified if curved surfaces are considered. However, it appears to be a reasonable approximation for many systems.

Integrating Eq. [4] over \( A_{\text{real}} \) yields the friction, which can then be divided by the load to give the following relationship for friction coefficient

\[
\mu = F/L = \alpha + \sigma_0/P, \tag{5}
\]

where the local pressure, \( P \), is given by \( P = L/A_{\text{real}} \). Based on this equation, it is clear that Amontons’ law will be satisfied if \( P \) is constant and/or \( \sigma_0 \) is small.

Numerous simulations of boundary lubricants have shown that \( \sigma_0 \) is indeed small, even at
pressures close to the yield strengths of solids. Thus, it seems that Amontons’ law should hold in a wide range of cases. However, exceptions are observed when adhesive interactions are strong, which leads to large values of $\sigma_0$.

Conditions that lead to constant $P$ can be understood from macroscopic contact mechanics. Even highly polished surfaces are rough on many different length scales and when two macroscopic solids are brought into contact, only a small fraction of these surfaces will be in microscopic, mechanical contact. It can be shown that the pressure distribution averaged over these real contacts is surprisingly independent of the externally imposed load, $L$, provided that the surfaces are not too adhesive or too compliant. Basically, the number of points in microscopic contact increases with the load in such a way that $A_{\text{real}} \propto L$, and hence $P$, which is given by $L/A_{\text{real}}$, will remain relatively constant. Thus, Amontons’ law can also be understood in terms of macroscopic contact mechanics. This interpretation indicates that this law should hold irrespective of the local relation between normal and shear pressure. However, it is important to note that the independence of the pressure distribution on the normal load is not valid when conditions are less ideal and adhesion and plastic deformation play a role.

### 2.3 Velocity-Dependence of Friction

In general, solid friction is relatively independent of the sliding velocity $v_0$, with corrections on the order of $\ln(v_0)$. This finding, also known as Coulomb’s law of friction, can be rationalized nicely in the Prandtl-Tomlinson model when the spring constant, $k$, is sufficiently small. Recall, that it was noted in Section 2.1 that friction is relatively independent of the sliding velocity in this regime due to the occurrence of instabilities, which cause the surface atoms to undergo rapid transitions between minima and lead to energy dissipation. A certain number of instabilities will occur per sliding distance, $\Delta x$, with each instability producing a similar amount of heat, $\Delta Q$. Thus, in the steady state, one may associate the kinetic friction force, $F_k$, with the quotient

$$F_k = \frac{\Delta Q}{\Delta x} \quad [6]$$
Once temperature comes into play, the jumps of atoms between minima may be invoked prematurely, i.e., prior to the formation of instabilities, via thermal fluctuations. These thermally-activated jumps decrease the force that is required to pull the surface atom, which leads to a decrease in the kinetic friction. The probability that a jump will be thermally activated is exponentially related to the energetic barrier of the associated process, which can be understood in terms of Eyring theory. In general, the energetic barriers are lower when the system is not at its thermal equilibrium position - a scenario that is more prominent at higher sliding velocities. Overall, this renders $F_k$ rate or velocity dependent, typically in the following form

$$F_k \approx F_k(v_{\text{ref}}) + c \left( \ln \frac{v_0}{v_{\text{ref}}} \right)^\gamma$$

[7]

where $c$ is a constant, $v_{\text{ref}}$ is a suitable reference velocity, and $\gamma$ is an exponent equal to or slightly less than one. Of course, this equation will only be valid over a limited velocity range. In many cases, $F_k$ becomes linear in $v_0$ at very small values of $v_0$, i.e., when one enters the linear response regime, in which the system is always close to thermal equilibrium.

An example of the velocity dependence of friction is given in Figure 5 for a boundary lubricant confined between two incommensurate surfaces. For the given choice of normal pressure and temperature, one finds four decades in sliding velocity for which Eq. [7] provides a reasonably accurate description. In the example shown in Figure 5, $c$ is positive and the exponent $\gamma$ is unity; however, neither of these statements are universal. For example, the Prandtl-Tomlinson model can best be described with $\gamma = 2/3$ in certain regimes, while confined boundary lubricants are best fit with $\gamma = 1$. Moreover, the constant $c$ can become negative, in particular when junction growth is important, where the local contact areas can grow with time as a result of slow plastic flow of the opposed solids or the presence of adhesive interactions that are mediated by water capillaries.
Figure 5: Typical velocity relationship of kinetic friction for a sliding contact in which friction is due to adsorbed layers confined between two incommensurate walls. The kinetic friction $F_k$ is normalized by the static friction $F_s$. At extremely small velocities $v^*$, the confined layer is close to thermal equilibrium and consequently $F_k$ is linear in $v^*$, as to be expected from linear response theory. In an intermediate velocity regime, the velocity dependence of $F_k$ is logarithmic. Instabilities or “pops” of the atoms can be thermally activated. At large velocities, the surface moves too quickly for thermal effects to play a role. Time-temperature superposition could be applied. All data were scaled to one reference temperature. From reference 25.

2.4 Role of Interfacial Symmetry

The two surfaces that comprise a contact can be oriented in any number of specific ways; however, for crystalline surfaces, interfacial symmetries correspond to either of two broad classifications. The first type of orientation is called the *commensurate* case, and is found when two identical surfaces are perfectly aligned. The term *incommensurate* corresponds to the case in which two crystalline surfaces are mis-oriented or have different periodicities. An example of a commensurate systems is given as structure A in Figure 7, while structures B through D are incommensurate. Interestingly, the orientation of the surfaces within a contact has a tremendous influence on the frictional properties of the system. Because commensurate surfaces are rarely found in real engineering contacts, it is important to avoid incorporating such artificial symmetries into calculations. In what follows, we illustrate the importance of this point by considering the effect of interfacial symmetry on static and kinetic friction.

Commensurate surfaces have a tendency to exhibit much larger static friction than incommen-
surate surfaces. This can be understood through the following example. Imagine two egg cartons sitting on top of one another. If the cartons are perfectly aligned, the peaks of one carton will each sit in a valley on the other, and a large force will be required to simultaneous lift all of the peaks out of the valleys in order to initiate lateral motion. On the other hand, if the cartons are brought out of registry, e.g., through rotation, only a few of the peaks and valleys will be aligned and less energy will be required to initiate motion. This scenario holds if the cartons are separated by eggs, which introduce new peaks and valleys with the same periodicity as the underlying carton. The rationale behind this simple example applies to surfaces that are composed of atoms and may be separated by other atoms and molecules. In this case, the periodicity of the surface atoms defines the peaks and valleys, and any confined atoms and molecules will attempt to adopt this periodicity.

It may come as a surprise to some that two commensurate surfaces withstand finite shear forces even if they are separated by a fluid. But one has to keep in mind that breaking translational invariance automatically induces a potential of mean force $F$. Due to the symmetry breaking, commensurate walls can be pinned even by an ideal gas embedded between them. The reason is that $F$ scales linearly with the area of contact. In the thermodynamic limit, the energy barrier for the slider to move by one lattice constant becomes infinitely high so that the motion cannot be thermally activated, and hence static friction becomes finite. No such argument applies when the surfaces do not share a common period.

The kinetic friction, $F_k$, is also affected by commensurability. If two crystalline surfaces are separated by one atomic layer only, $F_k$ may actually be reduced due to commensurability, although static friction is increased. The strikingly different behavior for commensurate and incommensurate systems is demonstrated in Figure 6.

Therefore, whenever we introduce symmetries into our systems, we risk observing behavior which is inconsistent with that observed when these symmetries are absent. Because opposing surfaces are almost always dislike, unless they are prepared specifically, it will be important to avoid symmetries in simulations as much as possible. Unfortunately, it can be difficult to make two surfaces incommensurate in simulations, particularly when the interface is composed of two
Figure 6: Kinetic friction $F_k$ as a function of the stiffness $k$ of the spring pulling the upper wall at constant, small velocity. The inset shows a part of the simulated system. At large values of $k$, the slider moves at the same velocity as the spring and the smooth sliding kinetic friction is probed. At small values of $k$, the system manages to lock into a potential energy minimum, similar to what happens in the Prandtl-Tomlinson model. The surface then undergoes plugging or “stick-slip” motion as a whole. In that regime, the measured friction approaches the value for static friction. Commensurability affects the measured values for $F_k$ in both regimes sensitively. From reference 25.
Figure 7: Projections of atoms from the bottom (solid circles) and top (open circles) surfaces into the plane of the walls. (A through C) The two walls have the same structure and lattice constant, but the top wall has been rotated by 0°, 11.6°, and 90°, respectively. (D) The walls are aligned, but the lattice constant of the top wall has been reduced by 12/13. The atoms can only achieve perfect registry in the commensurate case (A). From reference 14.

identical crystalline surfaces. These difficulties arise from the fact that only a limited number of geometries conform to the periodic boundary conditions in the lateral direction. Each geometry needs to be analyzed separately to ensure that the contact remains incommensurate and there are few general suggestions that one can offer. For surfaces with hexagonal symmetry, such as (111) surfaces of face-centered cubic crystals, it is often convenient to rotate the top wall by 90°. This rotation does not map the hexagonal lattice onto itself. The number of unit cells in the x and y directions should be chosen so that only marginal strain is needed to form an interface with a square geometry. The top view of some incommensurate structures between hexagonal surfaces is shown in Figure 7. In most cases, the measured friction between incommensurate walls is relatively insensitive to how incommensurability is achieved, as long as the roughness of the two opposing walls remains constant. \(^{15}\)

As a final point, we note that typical surfaces are usually not crystalline but instead are covered by amorphous layers. These layers are much rougher at the atomic scale than the model crystalline surfaces that one would typically use for computational convenience or for fundamental research. The additional roughness at the microscopic level due to disorder increases the friction between surfaces considerably, even when they are separated by a boundary lubricant. \(^{15}\) However, no systematic studies have been performed to explore the effect of roughness on boundary-lubricated
systems and only a few attempts have been made to investigate dissipation mechanisms in the amorphous layers under sliding conditions from an atomistic point of view.

3 Computational Aspects

A typical model system used in tribological simulations is shown in Figure 8. In this system, two walls are separated by a fluid and shear is applied by pulling the top wall with an external device while the bottom wall is held fixed. In atomic-level simulations, the two walls correspond to atomically-discrete surfaces and the fluid is composed of atoms or molecules, which represent lubricants or contaminants.

In most atomic-level tribological simulations, the behavior of such systems is explored with molecular dynamics (MD). However, as noted above, certain aspects of MD simulations performed under the non-equilibrium conditions that are inherent to tribology require more attention than in typical MD simulations of bulk systems that are at equilibrium. For instance, shear and load must be introduced in a meaningful way. Additionally, a great deal of heat is generated during tribological simulations, which must be removed from the system with thermostats. This requires
an understanding of how thermostats affect the behavior of the system. In this section, we cover these, and other, key points related to tribological simulations.

We start by discussing various means of incorporating surface roughness into the model systems in order to perform more realistic simulations. Means of subjecting the system to shear and load are discussed in Section 3.2. Thermostats are discussed in Section 3.3. In Section 3.4, we consider cases in which one can neglect the walls and treat the system as a bulk fluid. We finish with a discussion of different computational methodologies that are used in tribological simulations.

### 3.1 Surface Roughness

The natural starting point for a tribological simulation, or any other simulation, is to define the system that is to be simulated. Working within the context of the generic model shown in Figure 8, it is apparent that this starts by specifying the walls, which form the contact. In simulations and experiments aimed at exploring the fundamental aspects of friction, it has become common practice to employ crystalline surfaces. Setting up such surfaces is rather straightforward and will not be elaborated upon here. Instead, we focus on situations in which one is interested in studying realistic engineering contacts. As mentioned in Section 2.2, surfaces in such contacts are rough on many length scales. Recently, there has been an intensified interest in modeling more realistic surface profiles; however, this research has focused so far on contact mechanics rather than on sliding motion between fractal surfaces. In this section, we examine ways of characterizing and generating rough surfaces.

The roughness of a surface can be characterized by averaging its height difference auto-correlation function over one or several statistically identical samples. The height difference auto-correlation function, $C_2(\Delta r)$, is given by

$$C_2(\Delta r) = \langle [h(r) - h(r + \Delta r)]^2 \rangle$$ \hspace{1cm} [8]

where $h(r)$ is the height of a sample’s surface at the position $r = (x, y)$. Thus, $C_2(\Delta r)$ states the
variation in height we would expect to encounter if we move a distance $\Delta r$ away from our current position. For many real surfaces, $C_2(\Delta r)$ exhibits power law behavior according to Eq. [9]

$$C_2(\Delta r) \propto \Delta r^{2H}$$  \hspace{1cm} [9]$$

where $H$ is called the Hurst roughness exponent. $H = 1/2$ corresponds to a random walk in height as we move laterally over the surface. Surfaces satisfying Eq. [9] are called self-similar. A profile of a self-similar surface is shown in Figure 9 along with a flat elastic object pressed onto the rough substrate.

Various means of constructing self-similar surfaces are known. Some of them do not allow one to produce different realizations of surface profiles, for example by making use of the Weierstrass function. These methods should be avoided in the present context because it would be difficult to make statistically meaningful statements without averaging over a set of statistically independent simulations. An appropriate method through which to construct self-similar surfaces is to use a representation of the height profile $h(x)$ via its Fourier transform $\hat{h}(q)$.

In reciprocal space, self-similar surfaces described by Eqs. [8] and [9] are typically characterized by the spectrum $\tilde{S}(q)$ defined as

$$\tilde{S}(q) = \langle \hat{h}(q)\hat{h}^*(q) \rangle$$  \hspace{1cm} [10]$$
with

\[
\langle \tilde{h}(q) \rangle = 0 \\
\langle \tilde{h}(q)\tilde{h}^*(q') \rangle \propto q^{-2H-d}\delta(q - q')
\]  \[11\]

where \( d \) is the number of independent coordinates on which the height depends, i.e., \( d = 1 \) if \( h = h(x) \) and \( d = 2 \) if \( h = h(x, y) \), and \( \tilde{S}(q) \) is the Fourier transform of the height autocorrelation function \( S(\Delta x) = \langle h(x)h(x + \Delta x) \rangle \). The height-difference autocorrelation function \( C_2(\Delta x) \) and \( S(\Delta x) \) are related through \( 2S(\Delta x) = C_2(0) - C_2(\Delta x) \).

The full characterization of the stochastic properties of a surface requires consideration of higher-order correlations of the height function. However, it can be difficult construct surfaces in this manner without experimental input. As an approximation, it may be reasonable to neglect the higher-order terms.

One means of generating height profiles is to draw (Gaussian) random numbers for the real and complex parts of \( \tilde{h}(q) \) with a mean zero and defined variance, and then divide the random number by a term proportional to \( q^{H+d/2} \) such that Eq. [11] is satisfied. Furthermore, \( \tilde{h}(-q) \) must be chosen to be the complex conjugate to ensure that \( h(x) \) is a real-valued function.

Alternatively, one may simply write \( h(x) \) as a sum over terms \( h(q) \cos(qx + \varphi_q) \). In this case, one needs to draw one (Gaussian) random number with the proper second moment of \( h(q) \) with zero mean and one random number for each phase \( \varphi_q \), which is uniformly distributed between 0 and \( \pi \) and filter the absolute value of \( h(x) \) in the same way as described in the previous paragraph. Other methods exist with which to generate self-similar surfaces, such as the midpoint technique, described in reference 24.

### 3.2 Imposing Load and Shear

To explore tribological phenomena, it is necessary to subject the system to external shear and load. When working within the context of a model such as that shown in Figure 8, this is achieved by
pulling the top wall with an external driving device while holding the bottom wall fixed. It is
only natural to subdivide such a system into a slider (the top wall), a substrate (the bottom wall)
and the remaining system. In general, it is desirable to keep the interface as unperturbed by the
external forces as possible, and hence it is important to only explicitly apply any external forces
and constraints to the outermost layers of the substrate and the slider. In this section, we discuss
how this can be achieved in a manner that allows one to accurately simulate real sliding contacts.
It is noted that some systems do not require the explicit consideration of confining walls, and can
be treated as bulk fluids; we defer the discussion of such systems to Section 3.4.

It will be useful to establish some general nomenclature before proceeding. We have already
defined the terms slider and substrate within the context of the model in Figure 8. To properly
impose shear and load, it will prove convenient to subdivide the system even further. Specifically,
only the outermost layer of the slider will be coupled to an external driving device. We will refer
to this layer as the top layer (tl). Similarly, the substrate will be constrained by fixing the center of
mass of the bottom layer of atoms in the substrate. This layer will be called the bottom layer (bl).
All other atoms will be referred to as the embedded system, regardless of whether they are in the
slider, substrate or fluid.

There are three commonly used modes under which top layers are driven:

1. Predefined trajectory, e.g., \( X = X(t) \)

2. Predefined force, e.g., \( F = F(t) \)

3. Pulling with a spring, e.g., \( F_x = -k[X - X_0(t)] \), where \( F_x \) is the force acting on the top
   layer in the \( x \) direction, \( k \) reflects the (effective) stiffness of the driving device, and \( X_0(t) \)
   denotes the position of the driving device as a function of time.

The typical choices for the predefined trajectories or forces are constant velocity, including zero
velocity, constant separation, constant forces, and/or oscillatory velocities and forces. It is certainly
possible to apply different driving modes along different directions, e.g., applying a constant force
or load perpendicular to the interface and using a predefined velocity parallel to a direction that has
no component normal to the interface. Simulating tribometer experiments would best be achieved using constant velocity or constant force modes, whereas it is probably best to employ oscillatory motion in a lateral direction to mimic measured by a rheometer. In general, one should avoid using a scheme where the surfaces are held at a constant separation during sliding because such conditions rarely occur in experiments and such simulations may lead to erroneous results.

Note that pulling a point particle over a periodic potential in driving mode (3) resembles the Prandtl-Tomlinson model discussed in Section 2.1. As in that model, the calculated (kinetic) friction force can depend sensitively on the stiffness of the driving spring (see also Figure 6). Weak springs tend to produce higher friction than do stronger springs. The fact that the calculated friction is not only a function of the interface, but also depends upon how the interface is driven is important to keep in mind when comparing simulations to experiments.

It is often beneficial to define a coordinate $R_{tl}$ that describes the center of mass of the top layer. There are three common ways to set up the top layer. (i) The positions of top layer atoms $r_n$ are confined to (lattice) sites $r_{n,0}$, which are connected rigidly to the top layer. (ii) The top layer atoms are coupled elastically to sites $r_{n,0}$ fixed relative to the top layer, e.g., with springs of stiffness $k$. (iii) An effective potential, such as a Steele potential, $V_S$, is applied between embedded atoms and the top layer. There are specific advantages and disadvantages associated with each method. Approach (i) may be the one that is most easily coded, (ii) allows one to thermostat the outermost layer in an effective manner, while (iii) is probably cheapest in terms of CPU time.

The force on the top wall, $F_{tl}$, is evaluated differently for each of the driving modes mentioned above.

$$F_{tl} = F_{ext} + \begin{cases} 
\sum_{n \in tl} f_n & \text{(i)} \\
- \sum_{n \in tl} k (r_n - r_{n,0}) & \text{(ii)} \\
- \sum_{n \in em} \nabla_n V_S (r_n) & \text{(iii)} 
\end{cases}$$

[12]

where $f_n$ in line (i) denotes the force on atom $n$ and $\nabla_n V_S$ in line iii is the gradient of the surface potential with respect to an embedded atom’s position. $F_{tl}$ will be used to calculate the acceleration of the top layer, resulting in a displacement $\Delta R_{tl}$. This displacement needs to be added to the sites
\( r_{n,0} \) contained in the top layer in cases (i) and (ii).

It is possible to set the mass of the top layer, \( M_{tl} \), arbitrarily. For example, one may increase \( M_{tl} \) beyond the total mass of the atoms in the top layer to incorporate some mass of the top wall that is not explicitly included in the simulation. However, one should be aware of two effects that arise from altering \( M_{tl} \). First, altering \( M_{tl} \) affects the dynamics of the system. Specifically, if \( M_{tl} \) is increased, a time scale gap will arise between the fast atomic motion of the embedded system and the slow collective motion of the top layer. However, having the top wall move on shorter time scales than in real systems may help to overcome the time scale gap between simulations and experiments. Second, one should be aware that the measured friction may depend on the mass of the top wall when it is pulled with a spring. Large masses favor smooth sliding over stick-slip motion, and hence reduce the calculated friction.\(^{35,36}\)

### 3.3 Imposing Constant Temperature

The external driving force imposed on solids leads to the dissipation of energy or as heat. In experiments, this heat diffuses away from the interface into the bulk and eventually into the experimental apparatus. However, this does not occur in simulations due to limited system size, and hence artificial means of controlling temperature must be employed. Ideally, this is done by applying thermostats to only the outermost layers of the system. In some cases, such as the simulation of bulk fluids, there are no confining walls, and in other cases it is necessary to keep the confining walls in a rigid configuration. Under such circumstances, which are described in Section 3.4, thermostats must be applied directly on the sheared system.

Numerous ways of applying thermostats in MD simulations of systems that are in equilibrium exist, each with its specific advantages and disadvantages. When the system is far from equilibrium, such as during tribological simulations, stochastic thermostats have proven particularly beneficial. Langevin thermostats are the prototypical stochastic thermostats,\(^{37}\) and dissipative particle dynamics (DPD) is a modern variation thereof.\(^{38}\) Although the use of such thermostats can be motivated from first principles through linear response theory, i.e., there are rigorous schemes for the
derivation of the damping terms and the fluctuation terms contained in stochastic thermostats,\(^{39}\) we will not provide these arguments here. Instead, we will focus on their implementation and properties.

### 3.3.1 Langevin Thermostat

In the Langevin description, one assumes that the degrees of freedom within the system that are not explicitly considered in the simulation, exert, on average, a damping force that is linear in velocity, \(\gamma_i \dot{r}_i\), along with additional random forces \(\Gamma_i(t)\). This leads to the following equation of motion for particle number \(i\):

\[
m_i \ddot{r}_i + \gamma_i (\dot{r}_i - \langle v_i \rangle) = -\nabla_i V + \Gamma_i(t) \quad [13]
\]

where the damping coefficient \(\gamma_i\) and the \(\alpha\) component of the random forces \(\Gamma_{i\alpha}(t)\) acting on particle \(i\) should obey Eq. \([14]\)

\[
\langle \Gamma_{i\alpha}(t) \rangle = 0 \\
\langle \Gamma_{i\alpha}(t) \Gamma_{j\beta}(t') \rangle = \sqrt{2\gamma_i k_B T} \delta(t - t') \delta_{ij} \delta_{\alpha\beta} \\
\rightarrow \sqrt{2\gamma_i k_B T} \frac{1}{\Delta t} \delta_{t,t'} \delta_{ij} \delta_{\alpha\beta} \quad [14]
\]

in order to satisfy the fluctuation-dissipation theorem. In Eq. \([13]\), \(V\) denotes the total potential energy and \(\langle v_i \rangle\) the expected drift velocity,\(^{40}\) e.g., \(\langle v_i \rangle = 0\) in the bottom layer and \(\langle v_i \rangle = v_{tl}\) if atom \(i\) belongs to the top layer. The last line in Eq. \([14]\) refers to the discrete time description used in molecular dynamics in which \(\Delta t\) is the time step. When using predictor-corrector methods (velocity Verlet is a second-order Gear predictor corrector method), it is necessary to keep in mind that random terms cannot be predicted. Therefore, one should only apply the predictor-corrector schemes to the deterministic parts of the equation of motion. In cases where very high damping is applied, time steps can be kept large when employing efficient integration schemes.\(^{41}\) In general, however, one should ensure that the coupling between the system and the thermostat is sufficiently weak in order to avoid externally imposed overdamping.
It should also be emphasized that there is no need to choose the random forces from a Gaussian distribution unless one is interested in short-time dynamics. It is much faster to generate uniformly distributed random numbers for the $\Gamma_i(t)$’s on an interval $[-\sqrt{3}\sigma, \sqrt{3}\sigma]$, where $\sigma$ is the standard deviation of the Gaussian distribution. Moreover, having a strict upper bound on the $\Gamma_i(t)$’s eliminates potentially bad surprises when using higher-order predictor corrector schemes, and thus allows one to use a large time step while producing accurate thermal averages and trajectories.

It is also certainly possible to apply damping, and hence the thermostat, in a direction-dependent manner. For example, the damping terms can be suppressed parallel to the sliding direction. This is particularly important, for instance, when the system has a small viscosity or when the shear rates are high. Otherwise, one is likely to create artificial dynamics. Using the correct velocity profile $\langle v \rangle$ prior to the simulation can also reduce the problem of perturbing the dynamics in an undesirably strong fashion. However, anticipating certain velocity profiles will always suppress other modes, e.g., assuming laminar flow in a thermostat is likely to artificially bias the system towards laminar flow\textsuperscript{42} and may create additional artifacts.\textsuperscript{43–45}

### 3.3.2 Effects of Damping on Calculated Friction

Making assumptions regarding the dissipation of heat can also influence solid friction, although typically it is less of an issue. This can be explored most easily within the Prandtl-Tomlinson model; however, the lessons to be learned apply to a large degree to more general circumstances. In the original formulation of this model, see Eq. [1], damping takes place relative to the substrate. However, one may also assume that the conversion of energy into heat takes place within the top solid.\textsuperscript{46} Thus a generalized Prandtl-Tomlinson model would be

\begin{equation}
\dot{m} \ddot{x} + \gamma_{\text{sub}} \dot{x} + \gamma_{\text{top}} (\dot{x} - v_0) = -\nabla V - k(x - v_0 t) + \Gamma_{\text{sub}}(t) + \Gamma_{\text{top}}(t),
\end{equation}

where the indices “sub” and “top” denote the thermal coupling to substrate and top solid, respectively.
Figure 10: Friction velocity relationship $F_k(v_0)$ in the Prandtl Tomlinson model at (a) zero and (b) finite thermal energy, i.e., $k_B T = 0.1 V_0$. Different damping with respect to substrate $\gamma_{\text{substrate}}$ and top solid $\gamma_{\text{tip}}$ for different realizations of damping. The arrow in (b) points to the zero-velocity limit in the athermal case.

To investigate the effect of the thermostat on the frictional forces, it is instructive to study slightly underdamped or slightly overdamped motion. In the following, we will set $m = 1$, $a = 1$, $V_0 = 1$, $k = 0.5 V''_{\text{max}}$. Damping, $\gamma$, and temperature, $T$, will be varied, but we will first consider the athermal case where $T = 0$. With this choice of parameters, the maximum curvature of the potential, $V''_{\text{max}}$, will be greater than $k$ so that instabilities will occur during sliding and lead to finite kinetic friction at small $v_0$ in the absence of thermal fluctuations.

Figure 10(a) shows the friction-velocity dependence for the following choices of thermostats (i) $\gamma_{\text{tip}} = 1$, $\gamma_{\text{substrate}} = 0$, (ii) $\gamma_{\text{tip}} = 1/4$, $\gamma_{\text{substrate}} = 0$, (iii) $\gamma_{\text{tip}} = 1$, $\gamma_{\text{substrate}} = 0$, (iv) $\gamma_{\text{tip}} = 4$, $\gamma_{\text{substrate}} = 0$. We see that the kinetic friction is rather insensitive to the precise choice of the thermostat for small values of $v_0$, at least when the temperature is sufficiently small. This is because friction is dominated by fast “pops” in that regime. This conclusion becomes invalid only if $\gamma$ is sufficiently small such that $a/v_0$, the time it takes the driving stage to move by one lattice constant, is not long enough to transfer most of the “heat” produced during the last instability into the thermostat. At high velocities, the sliding velocity $v_0$ is no longer negligible when compared to the peak velocity during the instability. This renders the friction-velocity dependence susceptible to the choice of the thermostats. Damping with respect to the substrate leads to strictly monotonically increasing friction forces, while damping with respect to the top wall can result in non-monotonic
friction-velocity relationships. So far, we have considered the zero temperature case. Once finite thermal fluctuations are allowed, there is a qualitative friction-velocity relationship, which can be shown by choosing thermal energies as small as $k_B T = 0.1 V_0$, see Figure 10(b). Jumps can now be thermally activated and the friction force decreases with decreasing velocity. Yet again, at small $v_0$ there is little effect of the thermostat on the measured friction forces. Even changing $\gamma$ by as much as a factor of 16 results in an almost undetectable effect at small $v_0$. This behavior originates from the fact that the system can get very close to thermal equilibrium at every position of the top wall for very small sliding velocities. As such, linear response theory applies and it follows that friction and velocity are linearly related at sufficiently small values of $v_0$. This linearity is generally valid unless the energy barriers to sliding are infinitely high and explains the linear dependence of friction upon the velocity at small $v_0$ and finite $T$, as shown in Figure 5.

3.3.3 Dissipative-Particle-Dynamics Thermostat

A disadvantage of Langevin thermostats is that they require a (local) reference system. Dissipative particle dynamics (DPD) overcomes this problem by assuming that damping and random forces act on the center-of-mass system of a pair of atoms. The DPD equations of motion read

$$m \ddot{\mathbf{r}}_i = -\nabla_i V - \sum_j \gamma_{ij}(\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_j) + \Gamma_{ij}(t)$$

where $\Gamma_{ij}(t) = -\Gamma_{ji}(t)$. The usual relations for fluctuation and dissipation apply

$$\langle \Gamma_{ij,\alpha}(t) \rangle = 0$$

$$\langle \Gamma_{ij,\alpha}(t) \Gamma_{kl,\beta}(t') \rangle = 2k_B T \gamma_{ij} (\delta_{ik}\delta_{j\ell} + \delta_{il}\delta_{jk}) \delta_{\alpha\beta} \frac{\delta_{t,t'}}{\Delta t}$$

Note that $\gamma_{ij}$ can be chosen to be distance dependent. A common approach is to assume that $\gamma_{ij}$ is a constant for a distance smaller than a cut-off radius $r_{\text{cut,DPD}}$ and to set $\gamma_{ij} = 0$ otherwise. Since calculating random numbers may be a task of relatively significant computational effort in force-
field based MD simulations, it may be sensible to make $r_{\text{cut,DPD}}$ smaller than the cut-off radius for the interaction between the particles, or to have the thermostat act only every few time steps.

Among the advantages of DPD over Langevin dynamics are conservation of momentum and the ability to properly describe hydrodynamic interactions with longer wavelengths,\textsuperscript{47,48} which ensures that “macroscopic” properties are less effected with DPD than with Langevin dynamics. To illustrate this point, it is instructive to study the effect that DPD and Langevin thermostats have on a one-dimensional, linear harmonic chain with nearest neighbor coupling, which is the simplest model to study long wave length vibrations.

The Lagrange function, $L$, of harmonic chain without thermostats is given by

$$L = \sum_{i=1}^{N} \frac{m}{2} \dot{x}_i^2 - \frac{k}{2} (x_i - x_{i-1} - a)^2$$ \[19\]

where $a$ is the lattice constant, and $k$ the stiffness of the springs. Periodic boundary conditions are employed after a distance $Na$. The equations of motion at zero temperature with damping are

$$m\ddot{x}_i + \gamma \dot{x}_i = -k(2x_i - x_{i+1} - x_{i-1}) \quad \text{(Langevin)}$$ \[20\]

$$m\ddot{x}_i + \gamma(2\dot{x}_i - \dot{x}_{i+1} - \dot{x}_{i-1}) = -k(2x_i - x_{i+1} - x_{i-1}) \quad \text{(DPD)}$$ \[21\]

As usual, it is possible to diagonalize these equations of motion by transforming them into reciprocal space. The equations of motion of the Fourier transforms $\tilde{x}(q, \omega)$ then read

$$-m\omega^2 \tilde{x} + im\omega \gamma \tilde{x} + 4 \sin^2(qa/2)k \tilde{x} = 0 \quad \text{(Langevin)}$$ \[22\]

$$-m\omega^2 \tilde{x} + 4 \sin^2(qa/2)im\omega \gamma \tilde{x} + 4 \sin^2(qa/2)k \tilde{x} = 0 \quad \text{(DPD)}$$ \[23\]

Thus, while Langevin and DPD damping do not alter the eigenfrequencies of the chain, i.e., their $q$ dependence is $\omega_0(q) = \sqrt{4 \sin^2(qa/2)/m}$, the quality factor $Q$, defined as the ratio of the eigen-
frequency to the damping, does differ between the two methods, as shown in Eq. [24]

\[
Q(q) = \frac{\omega_0(q)}{\gamma} \quad \text{(Langevin)}
\]

\[
Q(q) = \frac{\omega_0(q)}{\gamma} \frac{1}{4 \sin^2(qa/2)} \quad \text{(DPD)}
\]  \[24\]

In the long wavelength limit where \( q \to 0 \), Langevin dynamics will always be overdamped, while DPD dynamics will be underdamped, provided that the system is not intrinsically overdamped, as would be the case in the vicinity of continuous phase transition. Although this example was based on calculations of a linear harmonic chain, the results suggest that DPD has little effect on dynamical quantities that couple to long wavelengths. The bulk viscosity of a system is an example of a property where the measured value in sheared fluids depends on the precise choice of \( \gamma \) only in a negligible manner as long as \( \gamma \) is reasonably small.\(^{48}\)

In some cases, it may yet be beneficial to work with Langevin thermostats. The reason is that (elastic) long wavelength modes equilibrate notoriously slowly. Consider the case where an elastic solid is pressed onto a fractal surface. Because the DPD thermostat barely damps long-range oscillations, we must expect a lot of bumping before the center-of-mass of the top wall finally comes to rest. Conversely, Langevin dynamics can lead to faster convergence because the thermostat couples more strongly to long-wavelength oscillations, as shown in Figure 11. When the Langevin thermostat is used, the system quickly reaches its mechanical position, while the DPD-thermostatted system is strongly underdamped, despite the fact that the damping coefficient was 10 times larger for DPD than for Langevin. In general, one needs to keep in mind that equilibrating quickly and producing realistic dynamics (or calculating thermal expectation values) are often mutually exclusive in simulations and it may be necessary to consider carefully which aspect is more important for a given question of interest.
Figure 11: Time dependence of the normal position $Z_{tl}$ of an elastic solid, which is pressed against a self-affine substrate similar to the one shown in figure 9. Two different damping/thermostatting schemes are employed, Langevin (broken lines) and DPD (full lines). Although the damping coefficient is 10 times greater in DPD than in Langevin, DPD-based dynamics are too strongly underdamped to relax efficiently to the right position.

### 3.4 Bulk Systems

In some cases, friction between two surfaces is dominated by the bulk viscosity of the fluid embedded between them. In these cases, it is often suitable to model the bulk sheared fluid and neglect the presence of confining walls. In this section, we describe computational approaches for shearing bulk systems and identify the conditions under which it is appropriate to treat the system in this manner. We start in Section 3.4.1 with a discussion of the conditions under which one may neglect confining walls. This is followed in Section 3.4.2 with a discussion of how to impose shear on bulk systems. We close in Section 3.4.3 by exploring ways in which the system can be constrained to accurately reproduce certain phenomena.

#### 3.4.1 Theoretical Background

When surfaces are sufficiently far from one another and shear rates are low, one can usually assume that the solid and first layer of fluid move at the same velocity. This is called a stick condition. For the calculation of shear forces in this scenario, it is possible to ignore the walls altogether by simply applying shear directly to the fluid. As the distance, $D$, between the two walls of the contact decreases, slip may occur, in which case the wall and nearby fluid no longer move at the same
velocity. The concept of slip is alluded to in Figure 8, where the slip length, \( \Lambda \), is introduced. The calculation of \( \Lambda \) from atomistic simulations is a subtle issue,\textsuperscript{50} which we will not touch upon here. When the fluid is confined even further, the concept of slip length might break down altogether and the measured friction becomes a true system function that cannot be subdivided into smaller, independent entities. The discussion is summarized in the following equation

\[
F/A = \begin{cases} 
\eta v/D & \text{hydrodynamic regime} \\
\eta v/(D + \Lambda) & \text{moderate confinement} \\
? & \text{strong confinement}
\end{cases} \tag{25}
\]

where \( F/A \) is the force per surface area that is required to slide two solids separated by a distance, \( D \), at a velocity, \( v \), and \( \eta \) denotes the (linear-response) viscosity of the fluid between the walls.

Figure 12 illustrates the effect of \( D \) on the measured friction of a system similar to that shown in the inset of Figure 6. At large separations, the behavior is reminiscent of hydrodynamic lubrication, i.e., the damping coefficient, \( \gamma_{\text{rheo}} = F/Av \), is approximately inversely proportional to \( D \), and \( \gamma_{\text{rheo}} \) is relatively independent of the orientation of the surfaces. As \( D \) is decreased, the damping force is no longer inversely proportional to \( D \), and begins to exhibit a dependence upon the orientation of the walls. In fact, the damping force for commensurate surfaces increases by several orders of magnitude by going from 4 layers of lubricant atoms to 3, as indicated by the vertical arrow in Figure 6. The large values for the effective damping can be understood from the discussion of lubricated commensurate surfaces in Section 2.4. Incommensurate walls do not show such dramatic effects because a fluid lubricant does not lock the surface on very long time scales. However, the system with incommensurate surfaces still deviates from hydrodynamic behavior as \( D \) decreases.

### 3.4.2 Boundary Conditions

In the preceding section, it was demonstrated that the walls of the system can be neglected when the system is in the hydrodynamic regime, i.e., when \( D \) is sufficiently large. In this situation, it is
often desirable to treat the system as a bulk fluid and apply shear directly without any boundary effects. In this section, we describe two different methods with which to shear bulk systems.

The first method we will discuss was proposed by Lees and Edwards, and is outlined qualitatively in Figure 13. In this technique, periodic boundary conditions are employed in all three spatial directions; however, while the center-of-mass of the central simulation cell remains fixed in space, many of its periodic images are moved parallel to the shear direction. Thus, even when a particle is fixed with respect to the central image, the distance to its periodically-repeated images will change with time if the vector connecting the two images contains a component parallel to the shear gradient direction.

To be specific, let $R_{ij}$ denote the position in the periodically repeated cell which is the $i$th image to the right and the $j$th image on top of the central cell. (A potentially third dimension remains unaffected and will therefore not be mentioned in the following.) The position in real
Figure 13: Visualizations of Lees-Edwards periodic boundary conditions. At time zero, $t = 0$, regular periodic boundary conditions are employed. As time moves on, the periodic images of the central simulation cell move relative to the central cell in the shear direction as shown in the middle and the right graph. Circle and square show points in space that are fixed with respect to the (central) simulation cell. It is important to distribute the effect of shear homogeneously through the simulation cell. Otherwise, velocities will be discontinuous in shear direction whenever a particle crosses the simulation cell’s boundary across the shear gradient direction. In this graph, $x$ corresponds to the shear direction and $y$ to the shear gradient direction.
space of the vector $R_{ij} = (X, Y)_{ij}$ would be

$$
\begin{pmatrix}
X \\
Y
\end{pmatrix}_{ij} =
\begin{pmatrix}
X \\
Y
\end{pmatrix}_{00} +
\begin{bmatrix}
1 & \dot{\epsilon} t \\
0 & 1
\end{bmatrix}
\begin{pmatrix}
iL_x \\
jL_y
\end{pmatrix}
$$

[26]

with $\dot{\epsilon}$ being the shear rate, and $L_x$ and $L_y$ being the lengths of the simulation cell in $x$ and $y$ directions, respectively. Conventional periodic boundary conditions can be reproduced by setting $\dot{\epsilon}$ to 0. When using the Lees-Edwards periodic boundary condition, thermostating is most naturally performed with DPD thermostats because no reference system needs to be defined.

When integrating the equations of motion, it is important to not impose the shear only at the boundaries because this would break translational invariance. Instead, we need to correct the position in the shear direction at each MD step of size $\Delta t$. This is done, for instance, in the following fashion:

$$X_{n+1} = X_n + \Delta X_n + \dot{\epsilon} \Delta t Y_n$$

[27]

where $\Delta X_n$ is the change in the $x$ coordinate if no external shear was applied. This way, the effect of shear is more homogeneously distributed over the system.

An alternative to Lees-Edwards boundary conditions is the formalism put forth by Parrinello and Rahman for the simulation of solids under constant stress.$^{52,53}$ They described the positions of particles by reduced, dimensionless coordinates $r_\alpha$, where the $r_\alpha$ can take the value $0 \leq r_\alpha < 1$ in the central image. Periodic images of a given particle are generated by adding or subtracting integers from the individual components of $r$.

The real coordinates of $R$ are obtained by multiplying $r$ with the matrix $h$ that contains the vector spanning the simulation cell. In the present 2-D example, this would read

$$R_\alpha = \sum_\beta h_{\alpha\beta} r_\beta$$

[28]

$$h = \begin{bmatrix}
L_x & 0 \\
0 & L_y
\end{bmatrix} \begin{bmatrix}
1 & \dot{\epsilon} t \\
0 & 1
\end{bmatrix}$$

[29]
In this approach, the potential energy, $V$, is a function of the reduced coordinates and the $h$-matrix. For the kinetic energy, one would only be interested in the motion of the particle relative to the distorted geometry so that a suitable Lagrange function, $L_0$, for the system would read

$$L_0 = \sum_i \frac{1}{2} m_i \left( \sum_{\beta} h_{\alpha \beta} \dot{r}_{i\beta} \right) - V(h, \{r\})$$ \[30\]

in which the $h$ matrix may be time dependent. From this Lagrangian, it is straightforward to derive the Newtonian equations for the reduced coordinates, which can then be solved according to any number of integration schemes. One advantage of the scheme outlined in Eqs. [28] to [30] is that it is relatively easy to allow for fluctuations of the size of the central cell. This is described further below.

### 3.4.3 Geometric and Topological Constraints

An important aspect of the methods described in the preceding section is that $L_x$ and $L_y$ can be time dependent. As we will show in this section, this flexibility allows the simulation cell to fluctuate independently along different spatial dimensions during the simulation. This capability is useful in simulations of systems such as self-assembled monolayers under shear. However, care must be taken when allowing for this additional flexibility because for some systems, e.g., simple fluids under shear, there is no particular reason why $L_x$ and $L_y$ should be chosen to be independent from one another. In this section, we use the specific example of simulating a diblock copolymer to explore aspects of allowing $L_x$ and $L_y$ to fluctuate independently.

The most simple diblock copolymers are linear chains, in which one part of the chain consists of one type of monomer, say polystyrene (PS), and the other one of another type, say polybutadiene (PB), as illustrated in Figure 14. PS and PB usually phase separate at low temperatures; however, due to their chemical connectivity, block copolymers cannot unmix on a macroscopic scale. They can only phase separate on a microscopic scale, the size of which is determined by the length of the polymers.
Figure 14: Schematic representation of the microphase separation of block copolymers. The left graph shows atomic-scale details of the phase separation at intermediate temperatures, the right graph a lamellar phase formed by block copolymers at low temperatures. The block copolymers have solid-like properties normal to the lamellae, due to a well-defined periodicity. In the other two directions, the system is isotropic and has fluid-like characteristics. From reference 55.

When lamellar structures are formed, it is necessary to ensure that the dimensions of the simulation cell are commensurate with the intrinsic periodicity of the lamellae. This prevents unintentionally subjecting the system to artificial pressure as a result of the geometric constraints. Subjecting the system to a predetermined pressure, or stress, in a controlled manner can be achieved by allowing the system to fluctuate parallel to “solid directions,” which are introduced in Figure 14. For these directions, it would be appropriate to employ the usual techniques related to constant stress simulations.\textsuperscript{52,53}

Let us consider the three-dimensional case and work within the Parrinello-Rahman framework. A rather general three-dimensional $h$ matrix of Eq. [31] will be considered.

\[
h = \begin{bmatrix}
L_{xx} & 0 & L_{xz} \\
0 & L_{yy} & L_{yz} \\
L_{xz} & L_{yz} & L_{zz}
\end{bmatrix}
\begin{bmatrix}
1 & \dot{\epsilon} t & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]  

[31]

It is now possible to treat the variables $L_{ij}$ as generalized coordinates and allow them to change during the MD simulation. For this purpose, it is necessary to define a kinetic energy $T_{\text{cell}}$ associated with the fluctuation of the cell geometry as a bilinear function of generalized velocities $\dot{L}_{ij}$,

\[
T_{\text{cell}} = \sum_{\alpha\beta\gamma\delta} \frac{1}{2} M_{\alpha\beta\gamma\delta} \dot{L}_{\alpha\beta} \dot{L}_{\gamma\delta}
\]

[32]
where $M_{\alpha\beta\gamma\delta}$ must be a positive definite matrix with the unit of mass. Although the optimum choice for the $M$-matrix is a matter of debate, a reasonable approach is to treat the various $\dot{L}_{\alpha\beta}$ as independent, uncoupled variables and assign the same mass $M_{\text{cell}}$ to all of them, which simplifies Eq. [32] to Eq. [33]

$$T_{\text{cell}} = \sum_{\alpha, \beta \leq \alpha} \frac{1}{2} M_{\text{cell}} \dot{L}_{\alpha\beta}^2$$  \[33\]

It is often sensible to choose $M_{\text{cell}}$ such that the simulation cell adjusts to the external pressure and the thermostat on microscopic time scales.

The Lagrangian, $L$, for Lees-Edwards boundary conditions combined with Parrinello-Rahman fluctuations for the cell geometry now reads

$$L = L_0 + T_{\text{cell}} - p \det h$$  \[34\]

where $p$ is an isotropic pressure and $\det h$ the volume of the simulation cell. The Newtonian equations of motion for the generalized coordinates $L_{\alpha\beta}$ and $r_i$ follow from the Lagrange formalism. Furthermore, it is possible to couple fluctuating cell geometries not only to constant isotropic pressure but to non-isotropic stresses as well. The description of these approaches is beyond the scope of the present tutorial, but can be found in the literature.\textsuperscript{52,53}

When studying systems with mixed “fluid” and “solid” directions, it is important to keep in mind that each solid direction should be allowed to breathe and fluid directions need to be scaled isotropically or constrained to a constant value. Allowing two fluid directions to fluctuate independently from one another allows the simulation cell to become flat like a pancake, which we certainly would like to avoid. As an example, consider Figure 15, in which a lamellar block copolymer phase is sheared. The convention would be to have the shear direction parallel to $x$ and the shear gradient direction parallel to $y$. There is no reason for the simulation cell to distort such that $L_{xz} = L_{yz} = 0$ would not be satisfied on average, so one may fix the values of $L_{xz}$ and $L_{yz}$ from the beginning. As a result, there is one solid direction plus two fluid directions. We can also constrain $L_{xx}$ to a constant value, because the shear direction will always be fluid and there
is another fluid direction that can fluctuate. This means that we should allow the simulation cell to fluctuate independently in only the directions of the shear and the shear gradient. Yet during the reorientation process, i.e., during the intermediate stage shown in Figure 15, simulation cells do have the tendency to flatten out, because periodicity, and hence solid like behavior, is lost for a brief moment in time.

It is interesting to note that the lamellar structure in Figure 15 does not find the true equilibrium state during reorientation, but instead adopts a metastable state. This behavior occurs because the periodic boundary conditions impose a topological constraint and prevent the system from simply reorienting. It is conceivable that similar metastable states are also obtained experimentally; however, the nature of the constraints leading to these states differs in both cases. One means of overcoming this topological constraint is to impose a higher temperature, $T_{ti}$, at the boundaries of the simulation cell (e.g., at $0 \leq r_y \leq 0.2$ and $0 \leq r_z \leq 0.2$) and to keep the temperature low in the remainder of the system. This would melt the lamellar structure at the boundary and allow the remaining lamellae to reorient freely with respect to the shear flow.

Figure 15: A lamellar block copolymer phase is reoriented through external shear. The initial phase has the director of the lamellae parallel to the shear gradient direction. The most stable state would be to orient the director parallel to the shear and shear gradient direction. However, the reorientation process gets stuck before true equilibrium is reached. The stuck orientation is relatively stable, because the lamellae have to be broken up before they can further align with respect to the shear flow. From reference 56.
3.5 Computational Models

In the preceding sections, we have discussed how to set up a tribological simulation by properly defining the system, applying thermostats and imposing shear. However, in order to perform the simulation one must describe the interactions between the particles that comprise the system. Numerous techniques exist for this purpose ranging from continuum-based models, which neglect the atoms altogether, to quantum chemical methods, which explicitly consider the system down to the level of the electrons. Generally, the choice of model is dictated by the nature and goals of the simulation that will be performed. In cases where one needs to describe chemical reactions, it is necessary to use first-principles methods or, in some special cases, reactive force-fields. However, if one wishes to study large systems or examine longer time scales, it is necessary to use more approximate methods.

In this section, we give a brief overview of theoretical methods used to perform tribological simulations. We restrict the discussion to methods that are based on an atomic-level description of the system. We begin by discussing generic models, such as the Prandtl-Tomlinson model. In Section 3.5.2, we explore the use of force-fields in MD simulations. In Section 3.5.3, we discuss the use of quantum chemical methods in tribological simulations. Finally, we briefly discuss multiscale methods which incorporate multiple levels of theory into a single calculation.

3.5.1 Generic Models

The Prandtl-Tomlinson (PT) model introduced in Section 2.1 is the most commonly used generic model for tribological simulations. It has tremendous didactic value because it clearly demonstrates the role of instabilities in energy dissipation and friction. Moreover, it is often used to describe the dynamics of systems such as an atomic force microscope tip dragged over a periodic substrate quite accurately. Even though the PT model is a significant approximation to realistic systems, it is worthwhile to run a few simulations with this model to explore its rich behavior. In particular, interesting dynamics occur when \( k \) and \( \gamma \) are so small that the surface atom is not only bistable but multistable and the motion is underdamped. In such a situation, the atoms do
not necessarily become arrested in the next available, mechanically stable site after depinning and interesting non-linear dynamics can occur, such as non-monotonic friction-velocity dependence.

Another frequently-used model system is the Frenkel-Kontorova (FK) model, in which a linear harmonic chain is embedded in an external potential. For a review, we direct the interested reader to reference 62. The potential energy in the FK model reads

$$V = \sum_i \left\{ \frac{1}{2} k (x_{i+1} - b - x_i)^2 - V_0 \cos(2\pi x_i / a) \right\}$$  \[35\]

where $a$ and $b$ are the lattice constant of substrate and slider respectively, and $k$ is the strength of the spring connecting two adjacent atoms in the slider. As in the PT model, finite friction is found when atoms can find more than one mechanically metastable position and become unstable during sliding. Experience indicates that it is not possible to reproduce the results of many tribological experiments with the FK model, despite the increased complexity with respect to that of Prandtl and Tomlinson. In particular, when parametrized realistically and generalized to higher dimensions, it is found that most incommensurate interfaces between crystals should be superlubric within the approximations of the FK model.\(^7\) In other cases, when instabilities do occur, the FK model can only describe the early time behavior of flat sliding interfaces.\(^63\) As such, one must be cautious when using the FK model.

### 3.5.2 Force-field Methods

The vast majority of atomic-scale tribological simulations use force-fields (FFs) to describe the interactions between atoms. A huge amount of literature exists regarding the development and use of FFs, and we will not attempt to cover this vast topic here. Instead, we will point out aspects of FFs and their use which are relevant to tribological simulations. The reader interested in a more general discussion of FFs is directed to the chapters by Bowen and Allinger,\(^58\) and Dinur and Hagler\(^59\) in volume 2 of this series and by Landis et al. in volume 6.\(^60\)

FFs are parameterized functions that relate the structure of a system to its potential energy,
which is taken as a sum of simpler energy terms arising from, for instance, bond distances and angles, as well as non-bonded terms, such as electrostatics and van der Waals interactions. In most cases, the parameters used in the FF are chosen to reproduce experimental results obtained under specific conditions. It is then assumed that the FF will be accurate for similar systems under similar conditions. As discussed above, systems are exposed to large pressures and shear during tribological simulations; however, FFs are typically not parameterized with data obtained under such conditions and there is no guarantee that FFs developed for simulations at ambient conditions are suitable for use under more extreme situations. As a result, one must take particular care to ensure that the FF used in a tribological simulation provides an accurate representation of the system over the pressure range encountered in that simulation. A natural starting point is to test the ability of existing FFs to reproduce experimental data obtained under the conditions of interest; however, in some cases it may be necessary to generate new sets of parameters.

FFs that are parameterized for high pressure conditions can still lead to behavior that differs from that observed in experiments. For instance, it is common practice to treat the interatomic interactions with Lennard-Jones (LJ) potentials. Although this is convenient from a computational standpoint, it is known that LJ potentials do not reproduce experimentally-observed behavior such as ‘necking’, where a material attempts to minimize surface area will break under large tensile stresses. Many other examples exist where particular types of FFs are unable to reproduce properties of materials and, once again, we emphasize that one should ensure that the FF used in the simulation is sufficiently accurate.

The calculation of electrostatic interactions in FF-based tribological simulations also require care. The typical model used in tribological simulations consists of two surfaces separated by a fluid, with the whole system subject to periodic boundary conditions (PBCs). If we define the system such that the surfaces extend in the $x - y$ plane, it seems only natural to apply (PBCs) in these two dimensions. However, care must be taken when treating the third dimension, $z$, which lies normal to the surfaces. Specifically, one must ensure that the length of the simulation cell in the $z$ direction is large enough to leave a large vacuum space between periodic images, which
will minimize electrostatic interactions between these images. This is particularly important when performing simulations with standard packages that use regular routines for the Ewald summation. In some cases, however, it may be possible to employ fast-multipole methods that are specific to 2-D systems.

We end this section by mentioning reactive FFs,\textsuperscript{61,64,65} which are capable of describing changes in chemical bonding and atomic hybridization. The development of reactive FFs has emerged as a topic of significant interest. However, as one can imagine, this task is exceedingly difficult and these FFs only apply to very specific systems and reactions. Nonetheless, reactive FFs have been used in many areas of simulation as a cost-effective alternative to first-principles calculations. In tribology, reactive FFs have been used to study tribochemical reactions in hydrocarbons.\textsuperscript{66} The simulations of these reactions are discussed in greater detail in Section 4.5.

3.5.3 Quantum Chemical Methods

Although most atomic-level tribological simulations are performed with force-fields, circumstances exist when such approaches are insufficient. This is particularly true when studying tribochemistry - chemical reactions that are induced by shear and load. Because chemical reactions involve changes in bonding, it is necessary to employ first-principles methods that explicitly consider the electronic structure of the system.\textsuperscript{67} These methods include those typically used in quantum chemistry (QC), such as molecular-orbital methods, e.g., Hartree-Fock, and density functional theory (DFT), as well as approximate QC techniques, such as tight-binding DFT and semi-empirical molecular orbital methods. In this section, we focus on the use of QC methods to study tribological phenomena.

The greatest limitation of QC methods is computational expense. This restricts system sizes to a few hundred atoms at most, and hence it is not possible to examine highly elaborate systems with walls that are several atomic layers thick separated by several lubricant atoms or molecules. Furthermore, the expense of first-principles calculations imposes significant limitations on the time scales that can be examined in MD simulations, which may lead to shear rates that are orders of
magnitude greater than those encountered in experiments. One should be aware of these inherent differences between first-principles simulations and experiments when interpreting calculated results.

In general, two types of first-principles tribological simulations have been reported in the literature. The first type of simulation is based on static energy calculations in which the dynamics of the system are not considered.\textsuperscript{68,69} A typical simulation would be set up as follows. An initial structure is generated where the two surfaces are aligned in the desired configuration and fixed at some constant separation distance, $D$. The structure is relaxed and the energy is recorded. Then the top layer is moved by some small distance relative to the bottom surface, while holding $D$ constant. The new structure is relaxed and the energy is recorded. This procedure is then repeated until the desired sliding distance is covered. In this approach, one generates an energy profile along the given trajectory. The (static) friction force can be estimated as the maximum slope, $\partial E/\partial x$, of the energy along this trajectory between a potential energy minimum and the following maximum. To determine the kinetic friction force one must identify instabilities and evaluate $F_k$ as

$$F_k = -\left( \frac{E_{\text{diss}}}{\Delta \alpha} \right)$$

where $E_{\text{diss}}$ is the total energy lost to instabilities over the total sliding distance, $\Delta \alpha$.\textsuperscript{70} It is important to note that the number of instabilities that occur over a given sliding distance increases with the size of the system, and hence for large systems one must use a very small step size to avoid missing instabilities.

The calculated energetics also allow one to estimate the normal load, thereby providing access to the friction coefficient once the friction force is known. This method, albeit crude, has been shown to yield good results when compared to experiments in cases such as graphite sheets sliding past one another.\textsuperscript{68} However, one should realize that approximations in the determination of the normal load, the assumption that friction only depends on energy barriers and the lack of a consideration of dynamical aspects of system, may lead to significant deviations from experimental results.
results for many other systems.

The second approach used in first-principles tribological simulations focuses on the behavior of the sheared fluid. That is, the walls are not considered and the system is treated as bulk fluid, as discussed in Section 3.4.2. These simulations are typically performed using *ab initio* molecular dynamics (AIMD) with DFT and plane-wave basis sets. A general tribological AIMD simulation would be run as follows. A system representing the fluid would be placed in a simulation cell repeated periodically in all three directions. Shear or load is applied to the system using schemes such as that of Parrinello and Rahman, which was discussed in Section 3.4.2. In this approach one defines a (potentially time-dependent) reference stress tensor, \( \sigma_{\text{ref}} \), and alters the nuclear and cell dynamics, such that the internal stress tensor, \( \sigma_{\text{sys}} \), is equal to \( \sigma_{\text{ref}} \). When \( \sigma_{\text{sys}} = \sigma_{\text{ref}} \), the internal and external forces on the cell vectors balance, and the system is subject to the desired shear or load.

When performing variable-cell AIMD simulations with plane-wave basis sets, problems arise from the fact that the basis set is not complete with respect to the cell vectors.\(^7\) This can introduce fictitious forces (Pulay forces) into \( \sigma_{\text{sys}} \) and lead to artificial dynamics. To overcome this problem, one must ensure that \( \sigma_{\text{sys}} \) is well-converged with respect to the basis set size. In general, it is found that one needs to employ a plane-wave kinetic energy cutoff, \( E_{\text{cut}} \), that is approximately 1.3 - 1.5 times greater than that used in a comparable constant volume simulation. This is only a general guideline, as opposed to a strict principle, and one should test the convergence of \( \sigma_{\text{sys}} \) with respect to \( E_{\text{cut}} \) before performing any serious first-principles tribological simulations.

A related issue arises from the fact the quality of a plane-wave basis set is dependent upon the volume of the cell. In most variable cell AIMD simulations performed with plane-wave basis sets, the number of plane waves is determined for the initial cell volume and held fixed throughout the simulation. That is, as the cell changes shape and size, the number of plane waves remains constant, and hence the quality of the basis set, which is more closely related to \( E_{\text{cut}} \), is not constant throughout the simulation. This can lead to problems when calculating *relative* energies, which govern chemical reactions. Essentially, a relative energy is the difference between the absolute
energies of two systems and if the systems are treated at the same level of theory, any errors in the absolute energies of the individual species due to quality of the basis set will roughly cancel. On the other hand, if one system is treated with a higher quality basis set, these errors will no longer cancel, but instead will become incorporated into the relative energy. These errors can become significant if the cell undergoes a substantial change in volume, which may occur during a tribological simulation. Two popular approaches exist for minimizing the errors due to changes in basis set quality during variable cell AIMD simulations. The first approach is to set $E_{\text{cut}}$, which determines the number of plane waves considered in the calculation, high enough to ensure that the absolute energies are sufficiently converged for all structures within the range of volumes that one would anticipate encountering in the simulation. The second approach involves altering the number of plane waves that are explicitly considered during the simulation such that $E_{\text{cut}}$ remains constant with changes in volume.\textsuperscript{72} We note that regardless of which approach one uses it is necessary to ensure that the basis set is large enough to yield an accurate $\sigma_{\text{sys}}$. An advisable strategy involves performing the AIMD simulation with a basis that is large enough to yield converged results for the pressure and exploring the energetics of any observed processes in greater detail through a second set of calculations with a basis set that yield converged absolute energies. Such a strategy has been used in static quantum chemical calculations of phase transitions in CO$_2$\textsuperscript{73} and could also be used to study tribochemical reactions.

### 3.5.4 Multiscale approaches

In many simulations, it is desirable to simulate as many layers of the confining walls as possible in order to closely reproduce experimental situations. However, from a computational point of view one would like to simulate as few degrees of freedom as possible. Unless conditions are special, all processes far away from the interface can be described quite accurately within elastic theory or other methods that allow for a description of plastic deformations, such as finite elements. The advantage of these continuum-theory based methods is that it is possible to increasingly coarse-grain the system as one moves away from the interface, thereby reducing the computational effort.
Figure 16: Representation of a finite-element mesh for the simulation between a fractal, elastic object and a flat substrate. From reference 24.

New methodological developments even allow one to couple atomistic simulations to continuum-theory descriptions.\textsuperscript{74,75} It is beyond the scope of this chapter to provide a detailed description of all of the available methods; however, the coordinate discretization scheme shown in Figure 16 alludes to how one must proceed when incorporating different mesh sizes into a simulation. The reader interested in a more detailed discussion of multi-scale simulation methods is directed to the chapter by Cagin in this volume of this series.

While quasi-static processes can be modeled quite well with continuum-mechanics based models employing varying mesh size, this is not the case for dynamic processes. Whenever there is a region where the coarse-grain level is changed, one risks introducing artificial dynamics. In particular, the transmission of sound waves and energy density is suppressed whenever the mesh size changes, and hence it is not possible to have the proper momentum and energy transfer across the boundary when employing a Hamiltonian-based description. It is important, however, to realize that the computational effort required to simulate a three-dimensional system of linear dimension
only scales with $L^2$, or sometimes $L^2 \ln L$, using coarse-grained models, as opposed to the $L^3$ scaling for brute-force methods. It is often advisable to sacrifice realistic dynamics rather than system size.

An alternative approach involves integrating out the elastic degrees of freedom located above the top layer in the simulation. The elimination of the degrees of freedom can be done within the context of Kubo theory, or more precisely the Zwanzig formalism, leading to effective (potentially time-dependent) interactions between the atoms in the top layer. These effective interactions include those mediated by the degrees of freedom that have been integrated out. For periodic solids, a description in reciprocal space decouples different wave vectors, $q$, at least as far as the static properties are concerned. This in turn implies that the computational effort also remains in the order of $L^2 \ln L$, provided that use is made of the fast Fourier transform for the transformation between real and reciprocal space. The description is exact for purely harmonic solids, so that one can mimic the static contact mechanics between a purely elastic lattice and a substrate with one single layer only.

There is even the possibility of including dynamical effects with time-dependent friction terms (plus random forces at finite temperatures). However, it may not be advisable to take advantage of this possibility, as the simulation would become increasingly slow with increasing number of time steps. Moreover, the simulation will slow down considerably in higher dimensions due to the non-orthogonality of the dynamical coupling in reciprocal space.

To be specific regarding the formalism, let $\tilde{u}_{qia}(t)$ denote the $\alpha$ component of the displacement field associated with wave vector $q$ and eigenmode $i$ at time $t$. In the absence of external forces, which can simply be added to the equation, the equation of motion for the coordinates that are not thermostatted explicitly, $\tilde{u}_{qia}$, would read:

$$
M \ddot{\tilde{u}}_{qia}(t) = -G_{qia} \dot{u}_{qia}(t) + \int_{-\infty}^{t} dt' \sum_{q',j} \sum_{\beta=1}^{3} \gamma_{q'ia}(t-t') \dot{u}_{q'j\beta}(t') + \Gamma_{q'ia}(t') 
$$

[37]

where the $G_{qia}$ are the (static) Green’s functions, or effective spring constants, associated with
eigenmode $i$ and wavelength $q$. The knowledge of these functions enables us to work out the static contact mechanics. The time-dependent damping coefficients $\gamma^{q j \beta}_{q i \alpha} (t - t')$ in equation (37) reflect the dynamical coupling between various eigenmodes. There is no reason why this coupling should be diagonal in any of its indices, and thus including the terms related to dynamics increases memory requirements and slows down the speed of the calculation tremendously, i.e., beyond the expense of approximating a semi-infinite solid by a discrete, elastic lattice of size $L^3$. Included in Eq. [37] are random forces $\Gamma^{q j \beta}_{q i \alpha} (t')$, which must be used at finite temperature to counterbalance the time-dependent damping term. The random and damping terms have to be chosen such that they satisfy the fluctuation-dissipation theorem.\(^{39}\)

4 Selected Case Studies

The last few years have seen an explosion in the number of atomic-level tribological simulations aimed not just at understanding fundamental aspects of friction, but also at determining the frictional properties of systems used in real-world applications. In this section, we will discuss selected studies in an effort to demonstrate how the principles discussed earlier in this chapter are used in practice. Unfortunately, it is not possible to consider all of the important, high-caliber research within the space of this chapter. Instead, we focus on a few key areas which encompass both fundamental and applied research in computational tribology.

4.1 Instabilities, Hysteresis and Energy Dissipation

In Section 2, we used a relatively simple model composed of a slider and substrate to demonstrate how mechanical instabilities lead to energy dissipation and friction. However, realistic contacts are much more complex. For instance, real contacts can rarely be described as one-dimensional and almost always contain some molecules that act as impurities. Understanding the frictional aspects of these systems will require a consideration of the role instabilities play in systems that are more complex than the PT model. In this section, we discuss studies that investigate instabilities in more
realistic systems.

The role of instabilities involving confined impurity atoms has been investigated by Müser using a model in which two 1-D or 2-D surfaces were separated by a very low concentration of confined atoms and slid past one another.\textsuperscript{25} The motion of the confined atoms was simulated with Langevin dynamics where the interactions between these atoms were neglected and the atom-wall interactions were described by

\begin{align*}
V_t &= V_t^{(0)} \cos((x - v_0 t)/b_t) + V_t^{(1)} \cos(2(x - v_0 t)/b_t) + \ldots \\
V_b &= V_b^{(0)} \cos(x/b_b) + V_b^{(1)} \cos(2x/b_b) + \ldots
\end{align*}

where the subscripts $t$ and $b$ denote the top and bottom walls respectively, $2\pi b_t$ and $2\pi b_b$ are the periods of the top and bottom walls, and $v_0$ is the sliding velocity. Only the first higher order harmonic was considered and defined as $V_1 = V_{t,b}^{(1)}$. In that study, the ‘impurity’ limit was considered in which the concentration of impurity atoms is so low that interactions between these atoms can be neglected.

Simulations with commensurate, 1-D surfaces showed that the behavior of the system was sensitive to $V_1$, as summarized in Figure 17. In the case where $V_1 < 0$, the atom becomes unstable at some point in time and slides into the next minimum. However, the curve representing the mechanically stable position, $x_{\text{ms}}$, is continuous and the atom can always remain close to the equilibrium position, which minimizes energy dissipation. When $V_1 = 0$, the motion of $x_{\text{ms}}$ is discontinuous; however, the instability leads to a transition between symmetrically equivalent minima. Consequently, there will be no net change in energy and dissipation will be minimal. On the other hand, when $V_1 > 0$, the motion of $x_{\text{ms}}$ is discontinuous and the transitions, or ‘pops’ as they were termed by the author of the study, occur between inequivalent minima. This leads to energy dissipation and friction. Thus, different types of instabilities were observed in commensurate systems depending upon the interference between $V_t$ and $V_b$, which was altered with $V_1$. Instabilities that occur when $V_1 \leq 0$ were termed \textit{continuous instabilities} and lead to vanishing friction as the sliding velocity,
$0 \leq \pi \leq \Delta x_{\text{wall}} \leq [\text{a.u.}] \leq 0 \leq \pi \leq 2\pi \leq 3\pi \leq x_n \leq [\text{a.u.}]$

Figure 17: Mechanical equilibrium position for adsorbed atoms between two commensurate solid surfaces as a function of relative displacement $\Delta x_{\text{wall}}$ between the walls. The gray lines indicate the motion of the adsorbed atom if the walls are in slow relative sliding motion. (a) $V_1 < 0$, (b) $V_1 = 0$ and $V_1 > 0$. From reference 25.

$v_0$, goes to zero. When $V_1 > 0$, first-order instabilities occur, which lead to energy dissipation and finite friction as $v_0$ approaches zero.

Simulations of incommensurate surfaces showed a similar dependence on $V_1$, with first-order instabilities occurring if $V_1 < V^*_1$, where $V^*_1$ is some positive, critical value that depends on the degree of mismatch between the lattice constants of the top and bottom surfaces. This leads to non-vanishing $F_k$ as $v_0$ goes to zero. In the case where $V_1 < V^*$, the atoms are dragged with the wall that exerts the maximum lateral force. This in turn, leads to friction that scales linearly with the sliding velocity. As a result, the friction force will go to zero with $v_0$.

The effect of dimensionality was also considered in that study. It was found that systems with commensurate 2-D walls yield results that are similar to the 1-D case because the interference between $V_i$ and $V_b$ persists. This is no longer true for the incommensurate case, where the adsorbed atoms can circumnavigate the points of maximum lateral force, which permits first-order instabilities regardless of the nature of the higher order harmonics in the wall-atom potential. Thus, one would expect friction to remain finite as $v_0$ goes to zero for incommensurate systems, while this may not be the case for commensurate surfaces. These differences were demonstrated by considering the change in the friction force as a system proceeds from stick-slip behavior to smooth sliding motion. In the stick-slip regime, the friction force is dominated by static friction, while
kinetic friction plays the major role in the smooth sliding regime. The simulations showed that an abrupt change in the friction force was observed when commensurate systems underwent a transition from stick-slip to smooth sliding motion, which is consistent with vanishing kinetic friction as the sliding velocity goes to zero. For incommensurate surfaces, the friction force varied smoothly as the system passed between these two regimes. These results are summarized in Figure 6.

That study was extended by considering higher concentrations of confined atoms or molecules, 1-D, 2-D, and 3-D surfaces, and the internal structure of the confined molecules. In what follows, we will focus on the effect of the concentration of confined particles. The relevant results of the study are summarized in Figure 18, and show that the dependence of the friction coefficient upon the concentration of lubricant molecules differs for systems with commensurate and incommensurate walls. For incommensurate systems, it was found that the magnitude of the friction coefficient, \( \mu \), was relatively insensitive to the concentration of confined molecules for coverages between 0.25 and 2.5 monolayers. Furthermore, \( \mu \) was not highly dependent upon \( v_0 \) and remained finite as the sliding velocity went to zero. This behavior is due to the ability of first-order instabilities to occur within incommensurate systems under all conditions. Meanwhile, first-order instabilities do not necessarily occur within systems with commensurate walls, and hence such systems exhibit a significant dependence on surface coverage. The results indicate that for sub-monolayer coverages, commensurate systems exhibit \( \mu \) that vanishes as \( v_0 \) goes to zero because the confined molecules move coherently, which prevents first-order instabilities. Above one monolayer coverage, the confined molecules do not move in this manner and first-order instabilities can occur, which leads to non-vanishing friction as \( v_0 \) goes to zero. The increase in first-order instabilities with coverage for commensurate systems is apparent from the data in Figure 18, where, in some cases, \( \mu \) of the commensurate system is higher than that of systems with incommensurate walls.

First-order instabilities may not only involve the translational motion of atoms confined within contacts, but may also involve chemical reactions within the confined fluid itself. This has been demonstrated recently in first-principles studies of (zinc) phosphates, which are found in protective films formed in automobile engines. Here, we focus on simulations of systems containing
phosphate molecules in which pressure-induced chemical reactions lead to hysteresis and energy dissipation. The reactions involving zinc phosphates are discussed in Section 4.5 along with other tribocatalytic reactions.

 Systems composed of phosphate molecules were exposed to isotropic pressures that were increased from 2.5 GPa to 32.5 GPa and then returned to the initial pressure. The resultant equation of state, see Figure 19, exhibited a hysteresis loop between 18 and 26 GPa. Such hystereses lead to energy dissipation and friction. Analysis of the dynamics of the system indicated that this hysteresis was the result of a reversible chemical reaction. Essentially, when the system was compressed to 26 GPa, a P-O bond formed abruptly between two neighboring phosphate molecules, as indicated by the rapid decrease in the relevant P-O distance shown in Figure 20. This rapid transition between two inequivalent states is consistent with a first-order instability, as defined above. When the system was decompressed, the P-O bond persisted until the pressure dropped below 18 GPa at which point P-O bond dissociated rapidly, as indicated by the data in Figure 20. Once again, the abruptness with which this transition occurred is consistent with the formation of instabilities discussed earlier in this chapter.

Integrating over the hysteresis loop between the ‘compression’ and ‘decompression’ curves in

Figure 18: Coverage dependence of the kinetic friction coefficient, $\mu_k$, of a system containing 0.25 - 2.5 monolayers of a simple fluid. Commensurate systems (c) are denoted with open symbols and incommensurate systems (ic) are designated with closed symbols. From reference 82.
Figure 19: Volume, $V$, of phosphates as a function of pressure, $p$. An initial compression cycle and a subsequent compression/decompression cycle is shown. A hysteresis occurs in the pressure range $18 \text{ GPa} < p < 26 \text{ GPa}$, indicated by points A and B. From reference 19.

Figure 20: Bond lengths, $d_{PO}$, between the phosphorus and oxygen atoms as a function of pressure, $p$, during the compression of phosphates. From reference 19.
Figure 19 yields the amount of energy dissipated due to the reversible bond formation/dissociation process. Unfortunately, it is not possible to determine the contribution of these transitions to the friction of phosphate films because such a calculation would require knowledge of the number of similar instabilities that occur per sliding distance, which is certainly beyond the limits of first-principles calculations. Nonetheless, the results do indicate that pressure- and shear-induced chemical reactions can contribute to the friction of materials.

4.2 The Role of Atomic-Scale Roughness

Recently, interest has arisen in simulating realistic contacts. A key question involves the degree to which the atomistic nature of the surfaces in these contacts affects calculated values of friction. In this section, we will discuss studies that shed light on the importance of including atoms and atomic-scale roughness in tribological simulations. We start with a recent study in which continuum models (no atoms) were compared with atomistic models representing similar contact geometries. We then examine simulations demonstrating that even sub-monolayer atomic roughness can significantly influence the calculated friction forces.

The importance of treating surfaces atomistically in tribological simulations can be determined by comparing results from simulations that employ an atomistic description of the system with those based on continuum models. Such a comparison was reported recently by Luan and Robbins in an investigation of the breakdown of continuum models at small length scales. In their study, calculations were performed using a system composed of a cylinder pressed against a flat, elastic substrate. Four distinct atomistic models of the cylinder were considered: (a) a bent crystal with lattice spacing matching that of the substrate to render the system commensurate, (b) a bent crystal with lattice spacing differing from that of the substrate to render the system incommensurate, (c) a stepped crystal, and (d) an amorphous structure. Despite clear differences at the atomic level, all of these systems can be treated with the same continuum level model in which the atomic details are neglected. These models were used to evaluate various quantities related to contact mechanics. We will focus on the results pertaining to static friction.
It was found that the static friction force was affected by the atomistic nature of the system. The commensurate and stepped cylinders yielded results that were in good agreement with the predictions from the continuum model. This agreement was attributed to the fact that in these systems, the atoms on the cylinder can lock into registry with those on the substrate, which leads to a large friction force that is proportional to the apparent area of contact. On the other hand, the results showed that the amorphous and incommensurate systems gave values for the static friction force that were approximately an order of magnitude lower than those calculated with the continuum model. This was attributed to the fact that the atoms on the cylinders do not lock into registry with those on the substrate, which decreases the force necessary to initiate sliding. Overall, these results indicate that an atomistic description of the surfaces can have a tremendous influence on the calculated friction. This is not surprising given the discussion in earlier sections of this chapter.

It is also important to note that the results for the amorphous and incommensurate systems, which are typical of surfaces in real contacts, differ most from the continuum models. This indicates that including atomic-level details of the contact may be necessary to simulate realistic systems.

The comparison of continuum and atomistic models by Luan and Robbins demonstrates that the atomic details of this contact can have a significant influence on the calculated friction. However, those calculations did not explore atomically-rough surfaces, which are most likely found in real engineering contacts. The effect of roughness has been investigated recently by Qi et al. in a study of the friction at the interface between two Ni(100) surfaces. Two models were considered in that work. In the first model, both surfaces were atomically flat, i.e., the rms roughness was 0.0 Å. In the second model, 25 percent of the surface atoms were removed to give an rms roughness of 0.8 Å. The surfaces were either perfectly aligned or misoriented through rotation.

The results showed that roughening the surfaces increased the friction coefficients by up to an order of magnitude, as can be seen in Figure 21. In the case of atomically flat surfaces, misoriented by 45°, the static friction was very low (µs = 0.21), as anticipated by arguments from analytical theories of friction. Roughening these surfaces increased µs to 2.06. The increase in the calculated friction with roughening was less dramatic for the aligned surfaces, where the atomically flat
surfaces form a commensurate contact which, as discussed in Section 2.4, results in large static friction forces regardless of surface roughness.

The microscopic origin of the increase in friction with roughening can be understood from the microstructures shown in Figure 22. These structures clearly show that plastic deformation occurs during sliding, except for the atomically flat, incommensurate case. This difference accounts for the low friction coefficient that was calculated for the incommensurate system. In the three cases where plastic deformation occurs, the atoms are no longer elastically coupled to their lattice sites and can interlock the surface in a manner akin to that shown in Figure 4. This interlocking increases the force required to move the surface. It is interesting to point out that upon bringing the two atomically flat commensurate surfaces together, the identity of each surface disappears and one ends up shearing a crystal, which exhibits finite resistance to shear. In such a scenario, no intervening layer is needed to interlock the surfaces.

The calculated results were compared with experimental data regarding the influence of the surface orientation on the friction of Ni(100)/(100) interfaces. Interestingly, the calculated results obtained with the atomically flat surfaces differed significantly from the experimental findings. The agreement between theory and experiment was vastly improved when the surface roughness
was taken into account in the simulations. This demonstrates that the mere presence of atoms in the surfaces of the contact may not be sufficient to obtain realistic results, but instead it may be necessary to include surface roughness at the atomic level.

4.3 Superlubricity

Another area of current interest is superlubricity, which was introduced in Section 2.1. In that section, an argument in favor of superlubricity was given based on the translational invariance of solids. Basically, if plastic deformation and wear are negligible, one would anticipate having the same free energy at the start and end of a sliding process. Since there is no net change in free energy, there will be no friction. An additional, microscopic argument for superlubricity is as follows. When a slider is moved relative to a substrate, in a statistical sense, there will be as many surface irregularities pushing the slider to the right as there are pushing it to the left. This will lead to a cancellation of lateral forces, see Figure 23 and ultra-low friction. Once again, this mechanism is only operative when energy dissipation is minimal.

In recent years, the search for superlubric materials has become a subject of practical importance, which may be most evident in the development of miniaturized systems, where nanoscopic surfaces slide past one another. In many cases friction and wear are the main impediments to
Figure 23: Cancellation of lateral forces between two surfaces. The atoms in the top layer, represented by circles, experience forces that are dependent upon the position of the atom with respect to the periodic substrate. The arrows on the atoms indicate the magnitude and direction of these forces. For contacts lacking commensurability that are contain a sufficiently large number of surface atoms or irregularities, these forces will cancel in a statistical sense.

Miniaturizing such devices even further. Identifying or designing superlubric materials may allow these technological barriers to be overcome. A thorough understanding of the theoretical aspects of superlubricity will aid in achieving these goals.

Based on the discussion in earlier sections of this chapter one may expect atomically flat incommensurate to be superlubric. Indeed the first suggestion that ultra-low friction may be possible was based on simulations of copper surfaces.\(^6,7\) Furthermore, the simulations of Ni(100)/(100) interfaces discussed in the previous section showed very low friction when the surfaces were atomically flat and misoriented (see the data for the atomically flat system between 30° and 60° in Figure 21). In general, however, it is reasonable to assume that bare metals are not good candidates for superlubric materials because they are vulnerable to a variety of energy dissipation mechanisms such as dislocation formation, plastic deformation and wear.

Layered materials such as graphite and MoS\(_2\) have been the focus of much attention in terms of superlubricity. These systems are expected to exhibit low friction because the coupling between atoms within a given sheet is much stronger than those between layers. As a result, the sheets will remain relatively rigid under sliding conditions, which in turn will minimize energy dissipation and lead to ultra-low friction. A major problem associated with layered materials is that they tend to rub away, leading to debris formation, which increases friction. Indeed, in some of the first experimental studies to demonstrate superlubricity it was found that MoS\(_2\) sheets exhibited ultra-low friction prior to the onset of wear.\(^87\)

Carbon nanotubes have also received a great deal of attention within the context of low-friction materials. These structures are consistent with rolled up graphite sheets and can exist as either soli-
tary tubes (single-walled nanotubes) or several concentric tubes of increasing size (multi-walled nanotubes). In multi-walled nanotubes, the interactions between atoms within the tubes are much stronger than those between the atoms in two neighboring tubes, which opens up the possibility of ultra-low friction. As such, it has been suggest that multi-walled nanotubes could be used as molecular-level bearings and springs, as well as gigahertz nano-oscillators. In the remainder of this section we will discuss recent theoretical studies of friction in double-walled carbon nanotubes.

Tangney et al. studied the friction between an inner and an outer carbon nanotube. Realistic potentials were used for the interactions within each nanotube and Lennard Jones potentials were employed to model the dispersive interactions between nanotubes. The intra-tube interaction potentials were varied and for some purposes even increased by a factor of 10 beyond realistic parameterizations, thus artificially favoring the onset of instabilities and friction. Two geometries were studied, one in which inner and outer tubes were commensurate and one in which they were incommensurate.

In the simulations, the inner nanotube was initially displaced relative to the outer tube by a distance, \(x\). In order to reduce surface area, the inner nanotube was pulled into the outer tube and a potential energy minimum was found to exist when the inner tube is completely embedded. However, due to the accumulation of kinetic energy, the inner tube will move past this minimum and extend beyond the other end of the outer tube. In the absence of frictional forces, one would anticipate that this process would be repeated indefinitely, with the relative displacements of the two nanotubes oscillating between \(+x\) and \(-x\). In the presence of frictional forces, the maximum displacement will decrease with time. This can be seen from the data shown in Figure 24, where the relative displacement of the nanotubes and frictional forces are shown as functions of time.

Interestingly, the displacement curves shown in Figure 24 are virtually identical for the commensurate and incommensurate systems. This seems counterintuitive as one would expect low energy dissipation for the incommensurate system and high friction for the commensurate system. To understand this apparent discrepancy, let us return briefly to the system of two idealized egg
Figure 24: Top: Friction force between two nanotubes as a function of time. Bottom: Displacement of the nanotubes as a function of time. Grey and black lines indicate incommensurate and commensurate geometries, respectively. From reference 92.

cartons mentioned in Section 2.4. Aligning the two egg cartons explains static friction because the cartons will be locked. However, once motion is initiated, there is no reason why the system should exhibit kinetic friction because whenever the top egg carton slides downwards, kinetic energy is produced that will help it to climb up the next potential energy maximum. In order to encounter kinetic friction, an additional ‘microscopic’ dissipation mechanism must be operative. The same scenario applies to atomic systems, and is exactly demonstrated by the data for the commensurate system shown in Figure 24.

The differences between the incommensurate and commensurate systems are evident from the relative forces between the nanotubes, which are also plotted in Figure 24. In both cases, the forces change sign as the system passes through the potential energy minimum, which is found when the inner tube is completely embedded in the outer tube. However, for the portions of the simulation between these transitions, the magnitude of the force for the incommensurate system is relatively constant, while that for the commensurate system oscillates rapidly with what was described as a ‘butterfly’ pattern. This behavior is directly related to commensurability. For the commensurate tubes, the oscillations are due to aligned atoms on the two tubes passing over one another. The
number of such interactions depends on the degree to which the two tubes overlap. Thus, when the inner tube is completely embedded, i.e., when the force changes sign, the magnitude of the force will be largest. As the inner tube moves out of either end of the outer tube, the number of inter-tube interactions will decrease, leading to the butterfly pattern as the inner tube moves back and forth. No such systematic alignment of atoms on the two tubes exists for the incommensurate system and the oscillations in the force are of a much smaller amplitude. It is important to note that despite the large oscillations, the average value of the force for the commensurate system is the same as that for the incommensurate system, which leads to the same behavior regarding the relative displacement of the tubes over time.

The study by Tagney et al. indicates that energy dissipation in double-wall nanotubes does not depend on commensurability. The absence of significant dissipation on long time scales, indicates that no instabilities occur. Thus, the simulated, low-dimensional rubbing system exhibits superlubricity, not only for incommensurate but even for commensurate surfaces. Finite static and zero kinetic friction forces have also been observed experimentally, albeit for a different system. Although the nanotube simulations were based on reasonably realistic potentials, it needs to be emphasized that real carbon nanotubes contain a lot of chemical defects. Additional simulations of double-wall nanotubes have demonstrated that defects in the walls of the nanotubes as well as the chemical details of the ends of the nanotubes can lead to deviations from superlubric behavior and induce non-viscous types of friction forces, which are also observed in experiments.

### 4.4 Self Assembled Monolayers

Self assembled monolayers (SAMs) are highly-ordered organic thin films which have been proposed as potential low friction, protective coatings for devices such as MEMS/NEMS and computer hard drives. The influence of the chemical details of the SAM, such as chain length, chemical nature of the terminal group, etc., on the observed friction have been investigated both experimentally and theoretically. As anticipated, it has been found that these chemical details can have a significant effect on the frictional properties of SAMs. For instance, certain types of end-groups
promote adhesion between two SAMs, thereby increasing friction. Recent simulations have shown that the frictional aspects of SAMs are not only influenced by the chemical details of the film, but can also be affected by the degree of order within the SAM.

In their study, Park and coworkers\textsuperscript{100} investigated the frictional properties of fluorine-terminated alkanethiol SAMs grafted to gold surfaces. The frictional properties of the system were investigated by sliding two SAMs past one another at velocities in the stick-slip regime under various external loads. The simulations yield the shear stress, $\sigma_s$, and the kinetic friction coefficient, $\mu_k$, can be estimated from the slope of a plot of $\sigma_s$ versus load, using the relationships contained in Eqs. [4] and [5].

Two distinct types of fluorine-terminated SAMs emerged from the simulations. One exhibited low shear strength, while the other showed a much higher resistance to shear. Interestingly, it was found that the two films had significantly different friction coefficients. Considering external pressures of above 400 MPa, it was found that the low-shear strength SAM had $\mu_k \approx 0.12$, while the friction coefficient of high-shear strength SAM was approximately zero. The authors of that study investigated the origins of these differences.

The differences in the frictional aspects of the two SAMs were found to arise from differences in the structures of the films. Specifically, the high-shear strength SAMs exhibited a high degree of order, such that each of terminal groups of the SAM on one surface was perfectly aligned with a terminal group of the SAM on the opposing surface. Thus, during sliding all of the chains of one SAM will lock with all of those on the opposite surface, leading to a stick condition and high friction. The low-shear strength same, on the other hand, exhibited less order and only a fraction of the terminal groups would lock, thereby leading to lower friction. Thus, the differences in the order of the SAMs accounts for the differences in the shear strengths of the films. The authors also attributed the lower kinetic friction coefficient for the high-shear strength SAM to the increased order in the film, as higher order had previously been shown to lead to lower values for $\mu_k$.\textsuperscript{101} Overall, this work demonstrates the structure and order of films can have a significant influence on friction, even when the surfaces are of identical chemical composition.
4.5 Tribochemistry

An emerging sub-discipline of tribological simulation involves the study of tribochemical reactions - that is, reactions which are activated by pressure and shear. These reactions alter the structure of lubricants and films that are used to protect surfaces from wear. Understanding the effects of these reactions on the intended behavior of these films is of utmost importance. However, simulation studies of tribochemical reactions have been impeded by the difficulty in accurately describing changes in chemical bonding. In a limited number of cases, this can be achieved with the use of reactive force-fields (FFs), as noted in Section 3.5.2, while in other cases, one must resort to expensive quantum chemical calculations. In this section, we will describe two studies where such methods were used to examine tribochemical reactions.

Chateauneuf et al. investigated the reactions that occur when SAMs containing diacetylene moieties are compressed and sheared. The authors were interested in exploring the origin of experimental results indicating that polymerization within such films can significantly influence the observed friction. To investigate this behavior, model SAMs composed of long alkyne chains grafted onto a diamond surface were compressed and sheared with a model for an AFM tip composed of amorphous carbon. The authors employed a reactive FF that was capable of modeling changes in hybridization of carbon atoms, and therefore able to account for the formation of chemical bonds leading to polymerization within the film.

The results demonstrated that both compression and shear can induce the formation of C-C bonds between sp-hybridized carbons atoms, which leads to polymerization within the SAM. Interestingly, it was found that the location of these reactive sites within the film could influence the calculated friction. For instance, if the diacetylene components in the chains were close to the tip/film interface, reactions between the film and tip could occur, which led to wear and high friction. On the other hand, if the diacetylene moieties were far from the tip, the reactions did not lead to wear and had little effect on the average calculated friction. These observations demonstrate that a proper treatment of the chemical reactivity of the system may be necessary in some cases to calculate friction accurately.
Reactive FFs can only be applied to a few specific cases for which they have been developed, such as the hydrocarbon systems discussed in the first part of this section. For other systems, describing tribochemical reactions requires the use of quantum chemical methods. In recent studies, such methods have been applied to investigate the behavior of zinc phosphates (ZPs) in response to high pressures. ZPs form the basis of anti-wear films derived from zinc dialkylthiophosphates (ZDDPs), which are additives that have been incorporated into virtually all motor oils for the last seven decades. It has recently become apparent that these additives must be replaced because they are harmful to the environment and do not inhibit wear effectively on modern engines composed of aluminum. Despite extensive experimental research,\textsuperscript{102} the mechanisms through which the ZP films form and function have remained unknown, thus hindering efforts to develop new additives.

One key experimental observation regarding the ZP films is that the films found on the tops of asperities are stiffer and exhibit chemical spectra indicative of longer phosphate chain lengths than films found in the valleys between asperities. This indicates that differences in the conditions at the two distinct locations alter the elastic and chemical properties of the films. One of the key differences between the tops of asperities and the valleys is the pressure experienced by the zinc phosphates. Since the highest pressures, and greatest potential for wear, are achieved at the tops of the asperities, determining the response of ZPs to these pressures may aid in developing a clear picture of how the anti-wear films work.

In a recent study, Mosey, Müser and Woo used \textit{ab initio} molecular dynamics simulations to investigate the behavior of ZPs under high pressure conditions encountered during asperity collisions in running engines.\textsuperscript{83} Systems composed of zinc phosphate molecules, which are formed when ZDDP decomposes thermally in the engine, were subjected to pressures that were increased from a low value to the theoretical yield strength of iron or aluminum and then decreased to the initial ambient values of pressure. Several successive compression cycles were performed. Relevant structures observed during these simulations are shown in Figure 25.

The simulations demonstrated that compressing the system to a minimum pressure of 5 GPa results in the \textit{irreversible} formation of chemical cross-links between the phosphate groups with
Zn acting as a cross-linking agent. This irreversible change is evident through a comparison of structures (a) and (c) in Figure 25, which show the structure of the system before and after compression. In the first case, the system is composed of disconnected ZP molecules, while in the latter, extended bonding is present throughout the system. It is important to note that the presence of chain-like structures in (c) is a consequence of the limited size of the systems that were considered, and in larger systems cross-linking would lead to structures where networking extends in all three spatial dimensions on longer length scales. A comparison with simulations of pure phosphates, see Section 4.1, showed that this irreversible behavior does not occur without zinc.

The results indicated that cross-link formation increased the bulk modulus of the system. As noted above, cross-linking was a pressure-induced effect that was facilitated by a change in the coordination at zinc when the pressure reached 5 GPa. The observation that stiffening of the film is a pressure-induced phenomenon is consistent with the differences in the measured elastic properties of films found on the tops of the asperities and those found in the valleys between asperities as mentioned above. Basically, in real systems, pressures high enough to form stiff cross-linked films are achieved on top of the asperities, but are not encountered between the asperities.

It was also found that cross-linking, which occurs at pressures accessible on aluminum, causes the films to become harder than aluminum. Thus, on aluminum surfaces one could expect the films to act as abrasives that will induce wear, as has been observed in sliding experiments. The authors of the study suggested that the inability of ZDDP additives to protect aluminum surfaces from wear may be due to the pressure-induced stiffening of the film.
An additional pressure-induced reaction was observed during the simulations at a pressure of 16 GPa, which is accessible on iron. This reversible process involved an increase in the coordination number of Zn through Zn-O bond formation and further increased the degree of cross-linking in the system. It was found that this increase in cross-linking increased the stiffness of the film to a value approaching that of iron. It was suggested that this may contribute to the anti-wear properties of the film. Essentially, stiffening allows the film to accommodate and redistribute applied loads, thereby protecting an underlying iron surface. This behavior clearly requires the presence of cations that can undergo changes in coordination, which may explain why the anti-wear capabilities of the ZP films are reduced significantly if some of the zinc atoms are replaced with calcium (which does not exhibit variable coordination).

Overall, this work highlights how quantum chemical methods can be used to study tribochemical reactions within chemically-complex lubricant systems. The results shed light on processes that are responsible for the conversion of loosely-connected ZP molecules derived from anti-wear additives into stiff, highly connected anti-wear films, which is consistent with experiments. Additionally, the results explain why these films inhibit wear of hard surfaces such as iron, yet do not protect soft surface such as aluminum. The simulations also explained a large number of other experimental observations pertaining to ZDDP anti-wear films and additives. Perhaps most importantly, the simulations demonstrate the importance of cross-linking within the films, which may aid in the development of new anti-wear additives.

5 Concluding Remarks

Interest in tribological simulations has developed because of the need to understand the fundamental details of friction and wear. In this chapter, we have provided a detailed overview of several aspects related to such simulations. Basic theories of friction have been covered, and potential pitfalls associated with tribological simulations were described. Particular emphasis was placed on designing simulations that are representative of realistic systems, so that meaningful results can be
calculated. Finally, several recently-reported tribological simulation studies were discussed. These studies span topics ranging from fundamental issues, such as the nature of instabilities during sliding, to practical issues, such as the function of engine anti-wear films. These studies represent the current status of simulations in exploring both fundamental and applied areas of tribology.

Acknowledgments

Financial support from the Natural Sciences and Engineering Research Council of Canada and SHARCnet (Ontario) is gratefully acknowledged.

References


