Rubbing and Scrubbing

Though simply expressed, the laws of friction encapsulate a host of microscopic and nanoscopic phenomena whose elucidation has become one of the most fascinating pursuits in applied physics.

Georg Hähner and Nicholas Spencer

The "rubbing and scrubbing department" was how David Tabor's friction, lubrication and wear laboratory was described by certain uncharitable colleagues at the Cavendish Laboratory in Cambridge, England, some 40 years ago. The tables have turned. Tribology, as Tabor named his discipline (from the Greek *tribos*, meaning "rubbing"), has become respectable—even positively modish—in physics departments worldwide. And Tabor, having become the revered elder statesman of this flourishing field, is often accorded a place in reference 1 of even the most hardcore tribo-physics papers.[1]

Although Tabor brought physics to tribology in the 1950s, the origins of the field lie in the engineering sciences and stretch back more than 5000 years to the neolithic period. Duncan Dowson, in his fascinating history of the subject,[2] describes an early use of bearings in door sockets in Assyrian villages before 4000 BC. Dowson's treatise also shows an ancient Egyptian tomb painting of the first recorded tribologist pouring a liquid (oil, water, milk?—the archaeologists are uncertain) in front of a large statue as it is being dragged over wooden planks by teams of slaves. This image has subsequently become a staple of tribology lectures and overview articles, and the temptation to include it here as figure 1 has been too hard to resist.[2,3]

Laws of friction

Although Aristotle had already identified the existence of friction, it was not until Leonardo da Vinci turned his extraordinary mind to tribology at the end of the 15th century that the subject was treated in a truly scientific manner.

As well as being the first person to formulate laws of friction, Leonardo introduced what has come to be the standard high school friction experiment: sliding objects on an inclined plane. He also grappled with the nature of wear, the effects of lubricants and the design of bearings. Possibly the most significant of Leonardo's tribological findings were his two observations that frictional force is (1) independent of the apparent contact area and (2) dependent on the normal force exerted on the sliding body. Leonardo extended the second observation to the definition of what he termed a coefficient of friction—that is, the ratio of the frictional force to the normal load N. Known later as the laws of friction, these empirical observations of Leonardo's are obeyed under a remarkably large range of circumstances.

Interestingly, much of Leonardo's writing on friction did not come to light until the 1960s, which is why the individual more often associated with the laws of friction is Guillaume Amontons, who independently studied both lubricated and unlubricated friction at the end of the 17th century.

Amontons's observations on friction, as presented to the Royal Academy of Sciences in Paris on 19 December 1699, were as follows

▷ That the resistance caused by rubbing increases or diminishes only in proportion to greater or lesser pressure (load) and not according to the greater or lesser extent of the surfaces.

▷ That the resistance caused by rubbing is more or less the same for iron, lead, copper and wood in any combination if the surfaces are coated with pork fat.

▷ That this resistance is more or less equal to one third of the pressure.[2]

The first of these observations became what are now known as Amontons's laws—that is

1. The force of friction is directly proportional to the applied load.
2. The force of friction is independent of the apparent area of contact.

These two, extraordinarily simple empirical laws hold true under a remarkably large set of sliding conditions, both lubricated and unlubricated.

A further law has been attributed to Charles Augustin de Coulomb,[4] although he was acutely aware of its limitations—that is

3. Dynamic friction is independent of the sliding velocity.

Coulomb's treatment of the difference between the static and dynamic coefficients of friction was perhaps more useful. (Static friction is always higher than dynamic friction, simply for energetic reasons: When the two bodies come to rest, the system falls into a potential well as stronger, time-dependent forces come into play. The system must be brought out of this well if further motion is to occur.)

Coulomb also made the valuable suggestion that friction could be made up of two terms: one term that varies with load (Amontons's first law) and a second, usually smaller, term due to adhesion. Coulomb's model has become quite relevant in recent times, since, on the nanometer scale, the adhesion term begins to predominate.

Friction mechanisms and adhesion

Although the laws of friction are empirical, considerable effort has been devoted to understanding their underlying

Georg Hähner *and* Nicholas Spencer *both work in the Swiss Federal Institute of Technology's department of materials and its laboratory for surface science and technology in Zurich, Switzerland.*

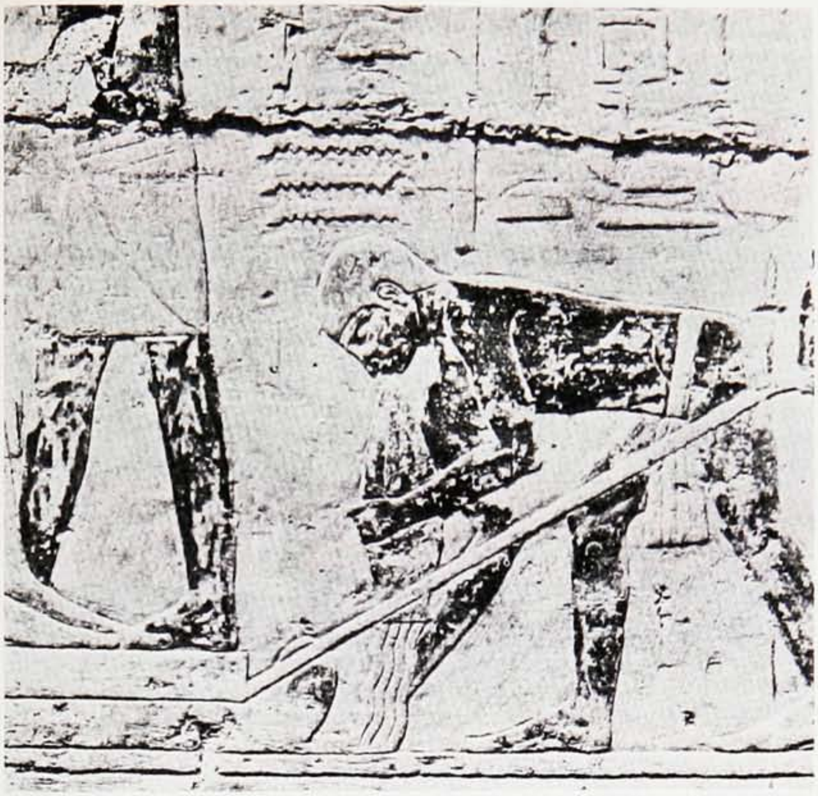


FIGURE 1. THE FIRST RECORDED TRIBOLOGIST pouring a liquid lubricant in front of the sledge used to transport the statue of Ti (about 2400 BC). (From refs. 2 and 3.)

mechanisms.

Nowadays, most friction theories assume that the shear strength (the force per unit area that resists sliding) is constant, from which it follows that frictional force is proportional to the true area of contact. This notion is quite consistent with Amontons's second law, for the sum of all contact points—established by microscopic surface irregularities—determines the true area of contact and, hence, the observed frictional force. Whereas the friction coefficient as defined by Amontons varies only slightly for different materials (in fact, it was believed for some time to be completely independent of the material and always approximately 0.3), the shear stress can vary over several orders of magnitude for various interfaces. (See the table on page 25.) This behavior is a consequence of the dependence of friction on both shear strength (low for indium, high for steel) and real contact area (high for indium, low for steel).

It was Tabor's group that supplied experimental evidence for a linear relation between load and true contact area. The researchers measured the electrical conduction (which was assumed to represent the real contact area) across a junction of two metal surfaces in contact as a function of the applied load.[1]

The field of contact mechanics deals with this issue theoretically. In the 1960s, Jim Greenwood (University of Cambridge) and coworkers showed that, under loading conditions that produce wholly plastic deformation, the contact area is proportional to the load and thus to the frictional force for a single spherical contact or for an array of similar spheres that all have the same height. In the case of elastic contacts, Heinrich Hertz's pioneering theory predicted a nonlinear (two-thirds power) relation between load and contact area and, hence, frictional force. Consequently, the friction coefficient would not be independent of load.

The issue of plastic versus elastic deformation becomes nicely resolved, however, when the height distribution of surface asperities (roughnesses) is taken into account. Greenwood showed that in the case of an exponential asperity height distribution, the dependence of total contact area on load is linear—regardless of whether the contact is elastic or plastic. This finding is fully consistent with Amontons's second law. Moreover, for the case of a Gaussian distribution of asperity height, which, for engineering surfaces, is often the case, the load dependence of contact area is also very nearly linear.

For some systems, shear stress increases with load, indicating that additional mechanisms contribute to the total frictional force. Deviations from the simple friction laws have been observed, for example, in metal–metal contact, where cold welding can occur and lead to a significant adhesive force. The linear relation between load and frictional force, however, is still often applicable, but with an offset—the adhesion contribution that was described by Coulomb.

Deviations from Amontons's second law have frequently been observed for polymeric or, more generally, viscoelastic systems. One of the materials that belongs in this category is rubber.

Automobile tires should possess low rolling losses but high sliding friction. The contact between such materials and more rigid countersurfaces (that is, between the rubber and the road) is often predominantly elastic, and the frictional properties differ fundamentally from those of many other pairings. The coefficient of friction varies considerably as a function of normal load, temperature and sliding speed, so no single value can adequately describe the material. For materials like rubber, the two main mechanisms for the dissipation of energy are deformation and adhesion.

The observation that friction may occur even if no load is applied indicates that friction is due to the shearing of adhesive junctions, which, in the case of metal–metal contact, is generally a wear process. However, wear and friction are often independent of each other, and there are numerous examples of high wear–low friction systems (chalk on a blackboard, a pencil on paper), as well as low wear–high friction systems (brakes). Nanotribological experiments have revealed that high friction does not necessarily involve wear at all.

The relation between adhesion and friction is the subject of continuing debate. Although it has been suggested that frictional force is correlated with adhesive force (by analogy to its dependence on an externally applied load), Jacob Israelachvili (University of California, Santa Barbara) and his group have recently shown experimentally that, for some systems, friction is correlated with the irreversible component of adhesion—that is, the adhesion hysteresis—rather than with adhesion itself.

To explain this observation, Israelachvili's group has proposed a thermodynamic model.[5] For nonadhering surfaces, the frictional force is described in terms of the work required to confine molecules between opposing surfaces. Irreversible compression work is then responsible for the dissipation of energy during sliding. In this model, then, the whole concept of a contact area becomes moot.

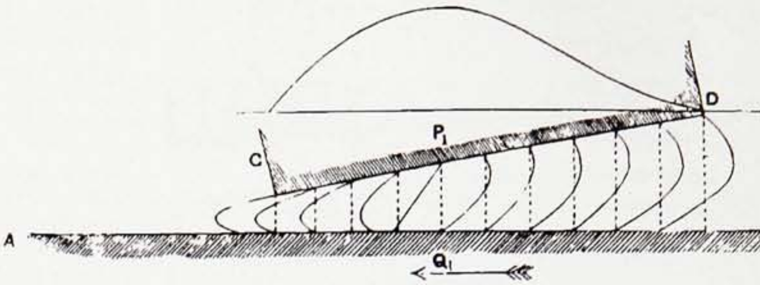


FIGURE 2. OSBORNE REYNOLDS'S picture of the action of lubricants on nonparallel plane surfaces in relative motion. The fluid-film wedge develops load-carrying pressures. (From ref. 7.)

Lubrication

For industrial applications, the principal tribological concerns are to reduce the friction coefficient—and, hence, the dissipated energy—and to avoid wear.

Wear occurs as a natural consequence when two surfaces in relative motion interact with each other. Lubricants or lubricant films between the opposing surfaces can ensure that shearing occurs inside the liquid—that is, between liquid–liquid junctions, which determine the resistance against sliding. If surfaces are covered by a thin lubricant film, liquid junctions will grow under contact until the critical shear stress of the film is reached, at which point gross sliding will occur. The introduction of a fluid film between components in relative motion solves a vast number of tribological problems.

Although lubricants had been used since the earliest times (recall figure 1), it was not until the mid-19th century that the development of railroads, the increasing use of lubricated machinery and the discovery of mineral oil combined to bring lubrication to the forefront of tribological investigation.

In those early days of lubricant studies, some of the most important experiments were carried out by Beauchamp Tower,[6] who discovered, interestingly, that a hole made in the top of a lubricated journal bearing (cylinder within cylinder) became the source of an oil fountain once the bearing went into motion. At that time, oil was commonly fed into journal bearings from the top, which, based on Tower's finding, was exactly the wrong position! Tower went on to determine the pressure distribution around a journal bearing, both axially and longitudinally.

Tower's work helped lay a firm experimental foundation for the work of Osborne Reynolds, who, in 1886, published his famous fluid film lubrication equation that links lubricant pressure, relative velocities of the moving surfaces, film thickness and lubricant viscosity.[7] The Reynolds equation, which underlies our modern understanding of fluid film lubrication, could quantitatively account for Tower's results.

Crucial to Reynolds's explanation was the formation of a physical wedge, in which the film thickness decreases in the direction of motion. (See figure 2.) The wedge generates load-carrying pressures within the film.

At the beginning of the 20th century, Richard Stribeck reported his carefully performed experiments with sliding bearings.[8] Tower had already noticed, several decades earlier, that the tangential force of friction on the journal bearing went through a minimum and then increased with velocity, at which point it was fairly independent of load. Extending Tower's work, Stribeck showed the systematic dependence of the friction coefficient for different loads on the sliding velocity.

Now known as Stribeck curves (figure 3), plots of frictional force against sliding velocity start at low velocities with a nearly load-independent frictional force. With increasing velocity, the frictional force drops steeply to a minimum—the Stribeck friction minimum—and then increases slowly. The minimum is significant. It separates the favorable wearless region (at higher velocities) from the region where strong wear may occur and in which the load is not completely compensated for by the hydrodynamic pressure of the lubricant film.

On the high-velocity side of the minimum, hydrodynamic or elastohydrodynamic friction operates. The latter mechanism involves not only the pressure induced in the liquid by the relative motion of the sliding partners, but also the elastic deformation of the partners themselves. It is the regime most often encountered in nonconforming or highly loaded systems. At the low-velocity end of the curve, the system is said to be running under boundary lubrication conditions—the situation found in either very slow moving machines, such as watches, or when loads are extremely high, as in the machining of metals. In this regime, adsorbed molecules from the lubricant can play a major role in keeping the sliding surfaces separated and in reducing wear.

Reynolds himself noticed that wear can occur as a result of metal contact before the movement-induced fluid film is able to separate the sliding surfaces. As mechanical systems became increasingly complex over the half century following Reynolds's observation, the issue of wear induced either by these startup effects or by extreme contact pressures assumed ever greater importance.

In an effort to improve the performance of lubricating

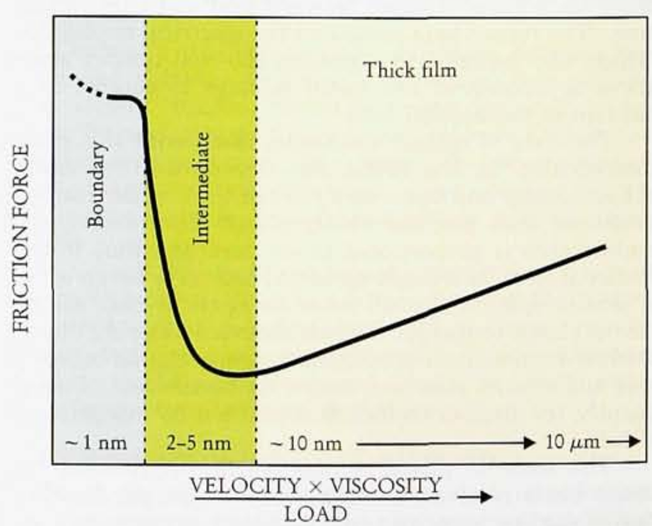


FIGURE 3. STRIBECK CURVE showing the frictional force as a function of sliding velocity, bulk viscosity of lubricant and applied load. Also shown is the thickness of the lubricant film in three frictional regimes. (From J. Israelachvili, *Molecular Adhesion and Tribology*, University of Lausanne, 1994.)

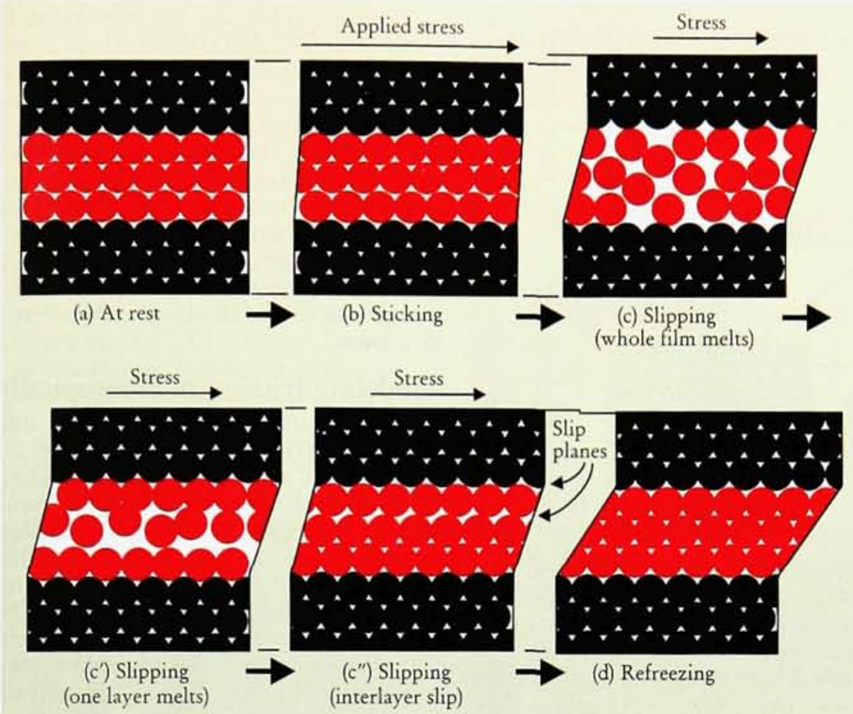


FIGURE 4. MOLECULAR REARRANGEMENTS occurring in a molecularly thin film of simple chain molecules between two solid surfaces during shear. (From ref. 13.)

oils, the use of small concentrations of additives became more common. The introduction of the hypoid gear (the gear that connects the drive shaft to the rear axle from above) into automobiles in the 1920s would have been impossible without the development of additives (usually fatty acids) that adsorb on the sliding surfaces, where they prevent destructive metal–metal contact at low speeds or high loads.

A more sophisticated class of wear-reduction extreme-pressure additives was introduced in the second half of this century. These additives, which include chlorinated hydrocarbons, as well as compounds containing metal, phosphorus or sulfur, react with surfaces only under extreme conditions, when they form products of low shear strength in a controlled corrosion process.

Virtually all friction and wear additives have been empirically developed, and not until the advent of surface science approaches in the last 30 years has the molecular basis of these additives begun to be understood. Andrew Gellman[9] (Carnegie Mellon University) examined the effect of submonolayer lubricant (ethanol) coverages on the friction measured between sulfided nickel single crystals in ultrahigh vacuum. Interestingly, the friction coefficient decreases monotonically with increasing coverage until one monolayer is reached, at which point it remains constant.

Wilfred Tysoe (University of Wisconsin—Milwaukee) investigated the now widely outlawed chlorinated hydrocarbons, which are highly effective extreme-pressure additives for steel applications. Tysoe showed that the effectiveness of several compounds in this class ceases dramatically at the melting point of iron chloride (940 K), which suggests that the formation of $FeCl_2$ is what protects the surface under extreme conditions.

The advent of railroads was clearly an important driving force in the development of lubrication technology. In early Russian trains, it was common practice to use lard for lubricating journal bearings, instead of the mineral oil more commonly used in the West. Before Tower's journal-bearing investigations and the subsequent design changes to reduce oil loss, the more viscous lard was an attractive option. An unfortunate consequence of the use of lard, however, was its theft by hungry peasants, who, presumably, consumed it between slices of bread. The Russian railroad authorities' solution to this problem was to adulterate the lard with soot. Not only did the soot render the lard inedible, thereby eliminating theft, but, thanks to its graphite component, the soot also significantly and unexpectedly improved the lubrication.

Here, again, is a case of a lubricant that works by presenting a layer of low shear strength within the contact region—in this case the weakly interacting graphite planes, which readily slide over each other.

The soot–lard mixture was an early example of the use of a solid lubricant. Graphite and other solid lubricants (such as molybdenum disulfide and boron nitride) are extensively used nowadays, both in combination with oils and alone when the use of liquid lubricants presents problems—in spacecraft, for example.

Tribology today

Gerd Binnig and Heinrich Rohrer's invention of the scanning tunneling microscope in the early 1980s and the subsequent development of the atomic and lateral force microscopes[10,11] unwittingly launched tribology into a new era of fundamental investigation.

Lateral force microscopy (LFM) offered the possibility, for the first time, of monitoring the forces acting on a single asperity during sliding. (See box 1 on page 26 and box 2 on page 27.) No bigger than a few nanonewtons and with ranges measured in nanometers, these forces lie at the very foundation of frictional behavior. What is clear from LFM measurements carried out over the last decade is that, in the nanoworld, Coulomb's two-term adhesion model is more useful than Amontons's simple one-term model, since adhesion predominates at low loads.

Although the fundamental causes of friction are still being debated, it is clear that both mechanical and chemical effects are involved. LFM has boosted the involvement

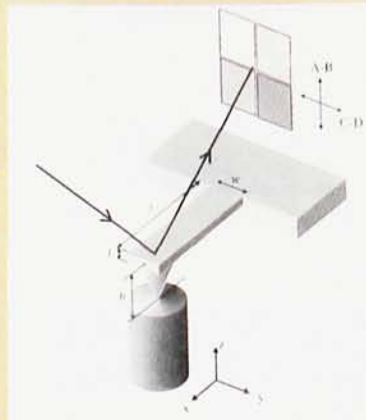
Values for friction coefficient and shear strength between steel slider and four materials

Material	Friction coefficient (μ)*	Shear strength calculated from friction measurements (g/mm)**
Indium	2	325
Lead	1.2	1600
Copper	0.8	28 000
Steel ball	0.8	140 000

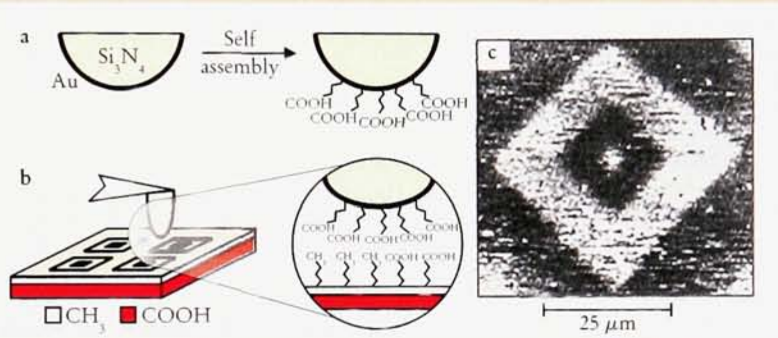
*From ref. 17
**From ref. 1

Box I. Friction-based Chemical Imaging Using Lateral Force Microscopy

Lateral Force Microscopy. The sample, supported on a piezodriven stage, is rastered underneath a sharp microfabricated tip, which is positioned at the end of a cantilever. A laser beam is reflected off the back side of the cantilever onto two sets of photodiodes such that the height of the contacted region and the frictional force on the tip can be simultaneously measured by monitoring the vertical displacement (A–B) and the torsion of the cantilever (C–D), respectively. These quantities, when displayed as a function of x and y position, yield topographical and frictional maps of the surface. (From reference 11, Overney and Meyer paper.)



Tip modification. Atomic force microscopy can be used to discern the chemistry of a surface. Here, a self-assembled monolayer (SAM) of carboxyl (COOH) groups covers the tip of an atomic force microscope (**a**), which is then dragged across a surface on which a regular pattern of two different SAMs—terminated by COOH and methyl (CH_3) groups—has been lithographed (**b**). When modified in this way, the tip encounters different degrees of interaction at the COOH- and CH_3-terminated regions and, therefore, records different frictional coefficients (**c**). (From C. D. Frisbie *et al.*, *Science*, volume 265, page 2071, 1994.)



of chemistry by using frictional images to provide chemical data of high spatial resolution, often on systems not readily accessible by other surface-imaging techniques. (See boxes.) Over the last few years, numerous examples of tribological chemical imaging have appeared in the literature.

A particularly interesting observation, which was obtained on several atomically smooth, clean surfaces studied by LFM, concerns stick–slip phenomena—the rapid stop–start movements, whose familiar manifestations in the macroworld include the squealing of brakes, the creaking of doors and the bowing of violin strings. (For more on stick–slip motion, see PHYSICS TODAY, September 1997, page 17.) It appears that stick–slip occurs not only in the macroworld, but also on an atomic scale, as the LFM tip moves from one potential well to the next across a surface. Since, during the stick phase, the two interacting bodies are at rest, stick–slip motion can be viewed as a continuous sequence of static friction events.

Using data obtained from the surface forces apparatus,[12] which brings atomically flat mica surfaces into contact with subnanometer precision by means of interferometric methods, Israelachvili has suggested an interesting mechanism to explain stick–slip in lubricated systems. In Israelachvili's model (see figure 4), molecular order–disorder transitions in the thin fluid film are proposed as the source of stick–slip motion. Highly branched molecules, which are too entangled to form ordered surface structures, are found to display smoother sliding behavior when they lubricate surfaces as molecularly thin films.[13]

Although this model might explain stick–slip motion under certain lubricated conditions, it is clear that many different mechanisms could be responsible. In engineering practice, stick–slip phenomena are observed only in situations where the Stribeck curve slopes downward.

Modeling friction microscopically

Significant effort has also been put into modeling frictional phenomena, although a full microscopic understanding of the interfacial processes that occur when two bodies are brought together is still lacking. One reason for this deficiency is that continuum mechanics loses its applicability as the scale of the bodies and the separation between them becomes very small. In addition, the mechanical properties of materials strongly depend on the size of the sample, and, since the junctions between contacting bodies can be small, their mechanical properties may be significantly different from those of the bulk material.

Nanotribological experiments have led to numerous simulations and the results, in turn, have led to further experiments, since, in both experiment and simulation, the number of particles is low enough in some cases to enable the two approaches to be compared.

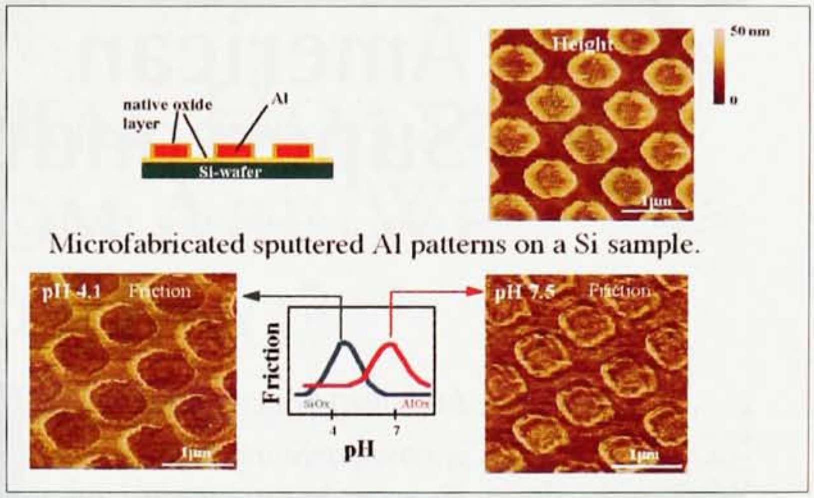
Uzi Landman (Georgia Institute of Technology) and his coworkers have extensively modeled single asperity contacts using a molecular dynamics approach. They have found that molecular ordering and layering processes occur in lubricants that are confined and sheared at high velocities by topographically nonuniform solid surfaces. They have also observed a correlation with oscillatory patterns in the frictional force, as well as the dynamic formation of elastic–plastic states of the lubricant due to extreme confinement between sliding asperities.

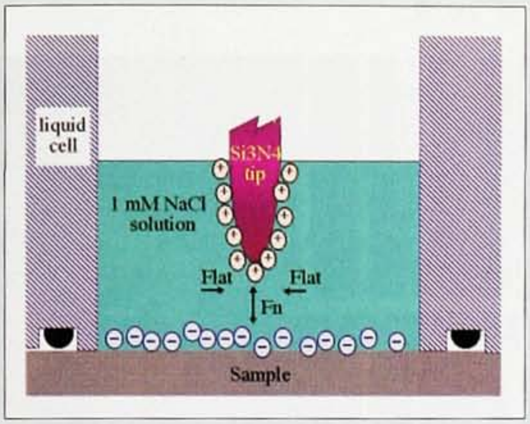
One of the questions at the very center of tribology is how energy is dissipated during frictional processes. For wearless friction, phonons and electrical effects[14] are both good candidates for carrying away energy, but their respective roles are still debated. Recent results seem to suggest that both mechanisms may contribute to energy dissipation, and, depending on the system, one or the other may predominate.

Elucidating the role of phonons was one of the aims of a nanotribology experiment conducted by Jacqueline Krim (now at North Carolina State University) and coworkers, who, with a quartz crystal microbalance (QCM), measured the forces on a monolayer of krypton as it was pushed along a gold surface.[14] They observed that a liquid film shows higher friction than a solid one. Usually, liquid films are used for lubrication, in which case shearing takes place between liquid interfaces. In the QCM experiment, the solid–liquid interface determines the friction, which is what accounts for the unexpected result. Mark Rob-

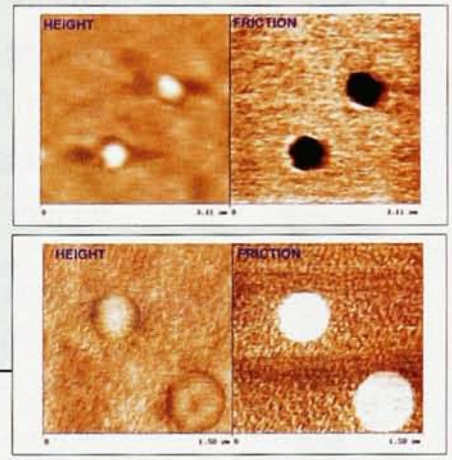
Box II. Friction-based Chemical Imaging Using Lateral Force Microscopy

PH sensitivity: When they are in contact with an electrolyte, oxidic surfaces (including that of the tip) become charged as a function of the electrolyte's pH value. And the pH value at which the charge changes from + to − depends on the chemical nature of the oxide. Since charge interaction between the tip and the sample contributes to the friction measured by lateral force microscopy, a frictional image of a multicomponent oxide surface will change as the pH of the electrolyte is varied, depending on the specific oxides involved. (From G. Hähner *et al.*, *Tribology Letters*, volume 3, page 359, 1997.)





Force imaging of polymers.▶ Height (left) and friction (right) images of a spin-cast polystyrene/poly(methyl methacrylate) polymer blend (PS/PMMA 1:10), obtained with gold-coated (top) and silica-coated (bottom) tips under perfluorodecalin. The inversion in frictional contrast is due to differences in interactions between the polar (silicon dioxide) and nonpolar (gold) tips, and the polar (PMMA) and nonpolar (PS) polymers. The role of the perfluorodecalin is to enhance the London component of the van der Waals force, due to its low refractive index. (From Feldman *et al.*, Langmuir, volume 14, page 373, 1998.)



bins (Johns Hopkins University) and coworkers have confirmed the result by means of computer simulations. In the case of krypton on gold, phonons appear to be almost solely responsible for the dissipation of energy.

Only very recently, Krim's group again used a QCM to investigate the temperature dependence of shear stress for a lead substrate 150 nm thick and covered with a nitrogen layer 1–2 molecules thick. When the lead substrate was cooled below the superconducting transition temperature, the friction between it and the solid nitrogen dropped by half. This seems to be the very first experimental evidence that conduction electrons can contribute to friction. In this case, electrical effects seem to play the dominant role in dissipation.[15]

Tribology tomorrow

Tribology has become a respectable research area not only for engineers but also for chemists and theoretical and experimental physicists. With the combined power of molecular dynamics, scanning probe microscopes and surface science, the prospects for achieving a much better understanding of the fundamentals of friction, lubrication and wear are very rosy indeed.

References

1. F. P. Bowden, D. Tabor, *The Friction and Lubrication of Solids*, Clarendon Press, Oxford, England (1985).
2. D. Dowson, *History of Tribology*, Longman, London (1979).
3. G. Steindorff, *Das Grab des Ti*, Hinrichs, Leipzig (1913).
4. C. A. Coulomb, Mém. Math. Phys. **10**, 161 (1785).
5. B. Bhushan, ed., *Micro/Nanotribology and its Applications*, Kluwer, Dordrecht, The Netherlands (1997). A useful collection of articles on micro and nanotribology.
6. B. Tower, Proc. Inst. Mech. Engr. (November, 1883), p. 632. B. Tower, Proc. Instn. Mech. Engr. (January, 1885), p. 58.
7. O. Reynolds, Philos. Trans. R. Soc. **177**, 157 (1886).
8. R. Stribeck, Z. Ver. dt. Ing. **46**, 1341 (1902); **46**, 1432 (1902); **46**, 1463 (1902).
9. C. McFadden, C. Soto, N. D. Spencer, Tribology International **30**, 881 (1997).
10. G. Binnig, C. F. Quate, C. Gerber, Phys. Rev. Lett. **56**, 930 (1986).
11. C. M. Mate, R. Erlandsson, G. M. McClelland, S. Chiang, Phys. Rev. Lett. **59**, 1942 (1987). R. Overney, E. Meyer, MRS Bull. **18**, 26 (1993).
12. D. Leckband, Nature **376**, 617 (1995). G. Luengo, F.-J. Schmitt, R. Hill, J. Israelachvili, Macromolecules **30**, 2482 (1997).
13. I. L. Singer, H. M. Pollock, eds. *Fundamentals of Friction*, Kluwer, Dordrecht, The Netherlands (1992).
14. B. N. J. Persson, E. Tosatti, eds. *Physics of Sliding Friction*, Kluwer, Dordrecht, The Netherlands (1996).
15. A. Dayo, W. Alnasrallah, J. Krim, Phys. Rev. Lett. **80**, 1690 (1998).

Further reading

There are a number of excellent general textbooks that cover tribology. Particularly recommended are G. W. Stachowiak and A. W. Batchelor, *Engineering Tribology* (Elsevier, Amsterdam 1993) and I. M. Hutchings, *Tribology—Friction and Wear of Engineering Materials* (CRC Press, Boca Raton 1992). For a more mathematical treatment, we recommend J. A. Williams, *Engineering Tribology* (Oxford University Press, Oxford 1994). ■