Friction **2**(1): 1–26 (2014)
DOI 10.1007/s40544-013-0033-z ISSN 2223-7690

REVIEW ARTICLE

The nature of friction: A critical assessment

Stephen HSU[1,*], **Charles YING**[2], **Fei ZHAO**[1]

[1] *George Washington University, Washington DC, USA*

[2] *National Science Foundation, Arlington, VA, USA*

Received: 03 September 2013 / Revised: 30 October 2013 / Accepted: 19 November 2013

Abstract: Friction is an essential part of human experience. We need traction to walk, stand, work, and drive. At the same time, we need energy to overcome the resistance to motion, hence, too much friction costs excess energy to perform work, introducing inefficiencies. In the 21st century, we are facing the dual challenges of energy shortage and global warming from burning fossil fuels. Therefore, the ability to control friction has become a top priority in our world today. Yet our understanding of the fundamental nature of friction is still lacking.

Friction has always been a subject of curiosity. Intensive study of the origin of friction began in the 16th century, after the pioneering work by Leonardo da Vinci. Yet progress in understanding the nature of friction has been slow, hampered by the lack of instrument to measure friction precisely. Ingenious experiments performed by Amontons, Coulomb, and others have yielded important insights to build the foundation of our understanding. Beginning in the late 1800s and early 1900s, the advent of steam engines, locomotives, followed by the automobiles airplanes, and space exploration demands a clear understanding of friction and the ability to control it for the machinery to last. Significant progress on how to apply and control friction in engineering friction was made through trial and error. At the beginning of the 21st century, a new dimension of nanoscale friction came into the picture in conjunction with the arrival of nanotechnology. Our understanding of atomic and molecular friction has been expanding rapidly. However, integration of the new found knowledge of nanofriction into engineering practices has been elusive. Why? What is the scaling relationship between atomic friction and macro-friction? Is it possible to predict friction at the macro-level from nanoscale results? Why nanofriction values often do not agree with the macrofriction values given the same materials pair? Could it be there is a length scale dependent characteristic friction value?

In engineering practice, progress since the 1980s has been slow. Most of the effort has been focused on lubrication research such as elastohydrodynamic theories and solid lubricants. Friction mechanisms and failures have received relative little attention while nanofriction received much of the attention.

Today, energy efficiency and renewable energy generation demand our immediate attention while we seek reduction in carbon emission. The ability to control friction becomes an essential step in seeking sustainable technologies. Friction, after all, is an indicator of energy efficiency. If we can reduce the unnecessary parasitic energy losses and increase our current energy efficiency, it will give us time to develop alternative energy sources. This paper examines our current understanding of friction, filling some voids with experimental data, and attempts to integrate the various pieces to identify the gaps of our knowledge, hopefully to spark new avenues of investigations into this important area.

Keywords: friction; scaling laws; molecular friction; length scale dominant friction values

* Corresponding author: Stephen HSU.
E-mail: stevehsu@gwu.edu

1 Introduction

Friction is a concept long understood by early civilizations when humans began to use tools, build monuments, and design shoes to control slippery paths. Yet its origin and nature have eluded understanding. On one hand, gravitation produces mass and weight and when the mass has to move across the gravitational fields, horizontal forces need to be exerted to overcome the gravitational forces. In tribology, friction is intrinsically coupled to the wear control and reliability aspects of all moving parts under various operating conditions and environments. Thus, the complexity of friction increases with the rapid increase in designs and machineries. Our understanding of the fundamental nature of friction requires a critical review.

To discuss the nature of friction, we need to have some common definitions: friction can be defined as the resistance to motion of a mass; frictional force is the force necessary to overcome the resistance to motion; static frictional force is the force necessary to overcome the resistance to motion from rest; kinetic frictional force is the force necessary to maintain motion of the sliding mass. The coefficient of friction is a normalized scalar to enable the comparison of the relative magnitudes of frictional force across a spectrum of materials and operating conditions.

To understand how we arrive at our current understanding of friction, it is necessary to trace back to the work by early pioneers on friction such as Leonardo da Vinci (1452–1519), Guillaume Amontons (1663–1705), and Charles-Augustin de Coulomb (1736–1806). They conducted simple yet elegant experiments to define friction that have framed much of our understanding on the subject. They experimented with blocks of wood/materials sliding on incline planes to observe the relationship between frictional force, velocity, and load. Based on their observations, they established some simple rules that become known as the laws of friction.

2 The laws of friction (for macroscale dry friction)

(1) Frictional force is directly proportional to the applied load, i.e., to the total force acting normal to the sliding surfaces (Amontons' First law, 1699).

(2) Frictional force for a constant load is independent of the apparent area of contact (Amontons' Second Law, 1699).

(3) The kinetic frictional force is independent of the sliding velocity (Coulomb's Law, 1785) (Actually, Coulomb states: at very low speeds, frictional force increases with speed; at medium speeds (1 in/s to a few ft/s), frictional force is nearly independent of speed; at high speeds, frictional force decreases with speed).

(4) Frictional force depends on the nature of the material in the contact (Coulomb's Law).

These laws over the years have been validated by many engineering applications. In the 19th century, the high performance automobile engines, space explorations, automation, etc., require more precise understanding of friction and the ways to control it. At the time, friction has been recognized as an instantaneous energy dissipation process where the energy is transformed into heat, work, and materials deformation processes. With better instrumentation, these laws were reexamined in detail, such as why friction is not related to the apparent contact pressure but on machine loading. Bowden and Tabor and others conducted carefully designed experiments in an attempt to trace the basic assumptions that lead to the laws.

2.1 Tabor's explanation of the friction laws

Coulomb's attribution of friction to interlocking surface asperities influences Tabor's thinking especially surface profilometers have just been developed and begin to be used at that time, so surface roughness can be quantified easily. Based on the measured surface roughness data, Greenwood and Williamson [1], and Whitehouse and Archard [2] proposed contact models to describe the processes of how rough surfaces can come into contact, including the asperity distributions, skewness, and waviness of surfaces. They classified machined surfaces based on their roughness distributions (Gaussian, exponential, stochastic, etc.). Bowden, in an effort to explain the origin of the Amontons' Laws, conducted experiments to measure real area of contact by using an electrical resistance method [3]. The data suggested at any one time, the real area of contact was extremely small. Bowden

suggested for steel on steel (polished surfaces), the real area of contact may be as small as 1/10,000 of the apparent area of contact (bear in mind that these surfaces are contaminated and covered by oxides). He further found that the real area of contact is directly proportional to the applied load, and almost independent of the sizes of the surfaces, confirming Amontons observations.

Bowden's results provided an explanation to the Amontons' First Law, at the same time, it suggested that the Amontons' Laws are approximations of the macro-contact systems, which is based primarily on the fact that the real area of contacts are so small for "rough surfaces", and that for practical purposes, friction varies with load and not contact pressure. For dry sliding systems and "normal" engineering rough surfaces, this may be true, but what about lubricated systems? We have examined this issue using surface profilometer traces during a lubricated sliding system to map the real areas of contact as a function of time. Results suggested that the real areas of contact ranged from 20% to 70% of the apparent area of contact [4]. This raised the question whether Amontons' First law actually applies to well-lubricated contacts or to highly polished bearings operating under lubricated conditions.

Amontons' Second Law assumes that the real areas of contacts are so small, given the normal roughness that statistically, the large apparent contact areas and the roughness are approximately the same, yielding approximately the same amount of contacting asperities that changes in apparent area of contact do not affect the real area of contact much (given the fact that roughness in the 17th century cannot be measured precisely or characterized quantitatively). In modern engineering surfaces such as precision bearings, magnetic hard disks, and MEMS devices, not only the surface roughness is tightly controlled at the nanometer scale but the surface roughness has directionality to control contact area and adhesion. With multiscale surface designs, sequential multiscale contact comes into play [5]. In these instances, contact pressure becomes a more relevant parameter when the contact area is carefully controlled, contrarily to the Second Law.

Coulomb's Third Law that the kinetic frictional force is independent of velocity ignores the interfacial temperature build-up as speed increases (energy balance $\Delta E = \mu L$+ heat + deformation). As the interfacial temperatures changes, so will the frictional forces.

While these laws have served us well over the past three centuries since they were proposed, we need to recognized their assumptions and limitations before we can move forward.

2.2 Modern concept of friction

The current concept of friction is that friction is an energy dissipative process. Frictional energy from resistance to sliding dissipates into friction, heat, and materials deformation and fracture. Friction can be classified into two categories: one is pure interfacial friction (no asperity penetration into opposing sliding surface); the other is the global resistance to motion, i.e., total resistance to motion as encountered in most engineering cases including plowing, adhesion, deformation, heat, microcracks, and delamination, etc.

The pure interfacial friction with non-adhesive surfaces (no intrinsic bonding when surfaces come into contact) has been extensively studied by many authors [6–12]. They focus on the nature of nanoscale friction on atomically flat surfaces. Insights on the atomic, molecular scale contacts and the fluid behavior of molecules under confined conditions were gained. The instruments used are surface force apparatus, atomic force microscope (AFM), and nanoindenters and have conducted theoretical studies using molecular dynamics (pair-wise potentials), force fields, and other simulation techniques to elucidate the nature of friction at the atomic and molecular levels. Translation of these new insights into macroscale phenomena, however, has been elusive. As a result, the nature of engineering friction and its prediction remained unresolved.

For engineering macrofriction, if we do an energy balance of the friction process (an energy dissipation event) including all the energy dissipation processes possible, we get:

Energy in = energy expended

= energy expended to overcome frictional resistance + energy expended to produce heat + energy expended for materials deformation and strain + energy expended to displaced the volume of material from

surface to the side by asperity scratching + energy expended to produce strain and micro-cracks + energy expended to produce gross fracture

This describes a much more complex process of friction occurring at the engineering level. This would imply that friction will change under high sliding speeds under dry sliding conditions. Since higher velocity tends to generate more heat which may change the materials properties and influence the heat dissipation process.

2.3 Tabor's formulation of friction equations

Tabor in his effort to understand the frictional processes, found Coulomb's interlocking asperities model failed to account for the higher friction values observed in his experimental measurements [13]. He postulated that there may be an adhesion term involved. Based on his study of adhesion between clean surfaces in vacuum, he found clean metal surfaces formed strong adhesive bonds [13], and he suggested that during sliding of asperities, cold welding or junctions might form, hence additional friction resistance might come from rupturing these bonds. He also recognized that plowing was another factor to account for the total frictional resistances in sliding systems [14]. Based on these considerations, he formulated various terms to be considered [15]. Here we presented an integrated form of the terms he proposed:

$$F = \mu L + \tau A_r + A_s P' \quad (1)$$

where F = friction force; μL = load term; τA_r = adhesion term, τ is the interfacial shear strength and A_r is the real area of contact; $A_s P'$ = plowing term, A_s is the projected area of the plowing path, and P' is the plastic flow stress of the softer material.

Following Tabor's formulation of friction, Buckley also studied the fundamental aspects of adhesion between two clean metal surfaces in ultrahigh vacuum for space exploration applications [16]. He found that adhesion depended on the specific crystalline phases in contact among pure transition metals. Adhesion was very sensitive to trace amount of impurities and contaminants in the material. For transition metals, the adhesion tends to stem from the transition metals' D-orbital electrons interactions with the oxygen, suggesting the important role that oxides may play in the adhesion process.

Tabor's formulation is based on theoretical consideration but the terms of A_r, τ, and A_s are difficult to determine inside the contact and almost impossible on the asperity level. P', the plastic flow stress of the softer material, is also difficult to determine for self-mated material and often depends on whether it is on the moving or stationary surface (the stationary surface usually has higher temperatures than the rotating surface, hence softer). The material transformation during sliding (higher temperatures often induced martensitic transformation of some metals, increasing the hardness) also introduces additional complications.

The interfacial shear strength, τ, is difficult to define and measure. Briscoe and Tabor [17] designed and constructed an apparatus of the shear strength of polymeric interfaces or boundary lubricating films. We have also developed instrument to measure film rupture strength [18] of thin films. However, these instruments cannot measure solid–solid contact shear strength. Under dry sliding conditions, from oxide covered surface to nascent metal surfaces, the sheat strength values could chang over several orders of magnitudes. We need instantaneous real area of contact data, frequency, locations of asperity-asperity contact, and associated asperity material properties to use in the equation. Following this conclusion, we built an apparatus to observe asperity–asperity collision process using a high speed video camera [19], but no evidence was found to suggest adhesion and junction growth taking place in dry contacts. Instead, we found strong evidence of particle detachment at the starting edge position and scratch across the surface. This can be explained by the fact that the surfaces are covered by oxide layer and the shear processes are insufficient to produce clean "nascent" surfaces similar to ultra-high vacuum conditions that Tabor and Buckley used in their observation. Other researchers also failed to confirm the existence of junction growth during normal friction and wear conditions.

While the formulation of these terms is theoretically reasonable, the use of the equation proved to be difficult. Therefore, we propose to modify Tabor's formulation to the following form:

$$F = \mu_0 L_0 + \tau A + \sum \varepsilon V \quad (2)$$

where
F = friction force;
$\mu_0 L_0$ = load term limited to pure interfacial friction without plowing;
τA = adhesion term, τ is the interfacial shear strength and A is the real area of contact;
εV = plowing term, ε is the force necessary to displace unit volume, and V is the total volume displaced (or volume times the average hardness of materials in front of the tip).

If we ignore the adhesion term (if we define adhesive surfaces of having chemical bonds when coming into contact, most engineering surfaces will not have adhesive term), then the equation simplifies to:

$$F = \mu_0 L_0 + \sum \varepsilon V \tag{3}$$

This would allow practical determination of pure interfacial friction (as discussed in the later sections of this paper) and measurement of the sum of displaced volume from asperities scratch marks inside the wear scar.

3 Nanofriction and the scaling laws

The beginning of the national nanotechnology initiatives around the world has enabled the establishment of facilities and instruments to image, measure, and manipulate materials at nanoscale and made them widely available. The availability of this capability stimulates innovations and creativity. It also provides opportunities to study atomic, molecular events and how they relate to practice. The magnetic hard disk technology took advantage of this capability to study nanofriction, stiction, and wear failures.

The initial barriers to nanoscale investigations are instrument limitations and our understanding of the potential artifacts introduced by the hardware and software. As time progresses, many issues have been resolved. In nanofriction measurement, many problems still persist since AFM and nanoindenters (the major instrumentation for studying nanofriction) are primarily designed for imaging and nanoindentation. Probe tip-surface interactions under sliding conditions create distortions of the cantilevers (twist and bend) and tip damage in AFM, while the nanoindenters are designed to indent with rigidity with relatively large diamond tip (100 nm tip diameter), and sliding the tip across the surface to measure friction often produces unintended scratching and plowing, which significantly change the magnitude of the measured friction force.

Goddard at Caltech first proposed a continuum of events from atoms to molecules, materials of various length scales (Fig. 1). Being a chemist, that all phenomena can be traced across the length scale range to explain the origin of the event is logical and reasonable. Drexler [20] also proposed scaling laws linking events at various length scales to their atomic or molecular origin. The question is will friction follows some sort of scaling laws?

From Eq. (3), at least in macrofriction, we will need to define the pure interfacial friction ($\mu_0 L_0$), which depends on surface forces between the tip and various material surfaces.

3.1 Classification of friction: Interfacial term and the plowing term

Accurate measurement of the frictional forces in devices is central to successful design of reliable and durable microsystems and devices. The common way to measure nanofriction is to use a sharp tip sliding on a sample surface [21–30]. There are two nanoscale behaviors in friction measurement: one is the tip radius and the other is the penetration depth of the tip across the surface. To quantify the contributions to friction at nanoscale, one would need to know precisely the tip shape, tip size, tip penetration depth, and forces in the normal and lateral directions. These parameters, however, are difficult to characterize and there are no standard measurement techniques to follow, so many literature reports, at least in the early

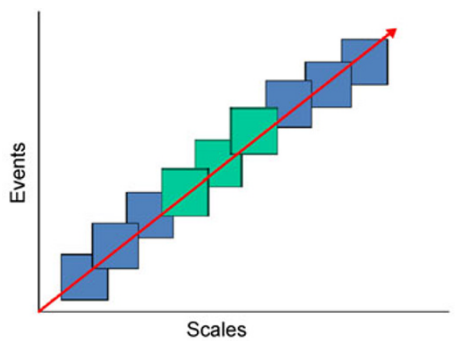

Fig. 1 A continuum model of scales and events, after Bill Goddard (Gordon Research Conference on Tribology, 1990 and his subsequent talks).

stages, did not report these parameters they used in conducting the nanofriction experiments but just assumed the manufacturers' nominal values.

3.2 AFM and tip shape

Cantilever spring constant and tip shape can significantly influence the measured nanofriction. This point can be illustrated in a typical AFM set up, as shown in Fig. 2, for nanofriction measurement. Cantilevers of various stiffness levels are available. Depending on the hardness of the surface and intended friction measurement, a proper level of cantilever spring constant needs to be chosen. Too soft a cantilever inevitably will introduce bending and twisting, complicating the translation of the voltage signals from the quadrant photodiode detector to force. Too stiff a cantilever will lose sensitivity and high resolution imaging capability. For nanomechanical property measurements, reasonably accurate determination of real area of contact is necessary to apply the contact mechanics formalism. Figure 2 also shows a typical silicon nitride tip at different magnifications. If the cantilever is twisted under bending and sliding forces, different parts of the tip will touch the sample, introducing uncertainty on the actual contact area, and significant errors on the reported friction value.

3.3 Interfacial friction and plowing study

Careful control of experimental conditions is crucial for quantitative friction studies. Literature shows large variations in friction data when the same material system was measured by AFM, nanoindenters, and microtribological instruments [25, 31–36] as shown in Table 1.

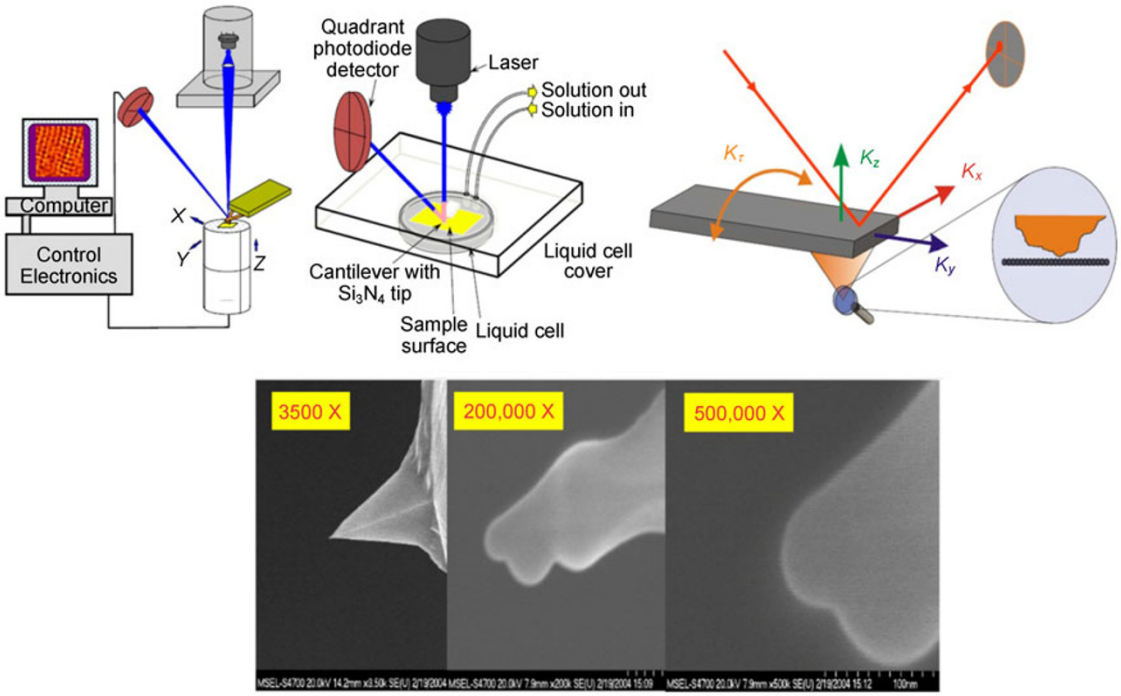

Fig. 2 Schematic diagram of an atomic force microscope and the tip shape. The image were obtained using field emission SEM on a silicon nitride tip at different magnifications.

Table 1 Experimental conditions and results of friction measurements on silicon [31].

Instrument	Tip material	Tip radius (nm)	Load (nN)	Friction force (nN)	Apparent contact pressure (GPa)	COF
Atomic force microscope	Silicon	150	5–700	0.15–21	0.3–1.54	0.03
Scanning force microscope	Diamond	100	1×10^4–7×10^4	250–1.8×10^4	9.4–18	0.25
Scanning force microscope	Diamond	1.6×10^4	2×10^4–1×10^6	220–1.1×10^5	0.19–1.51	0.11
Pin on disk	Diamond	1.2×10^6	1.7×10^7–2.7×10^8	1.4×10^6–2.2×10^7	0.21–0.54	0.08

From Table 1, we can see the coefficient of friction (COF) ranges from 0.03 to 0.11 for the same materials system measured by different instruments. To understand whether these difference stems from the scaling laws or measurement artifacts or instrument differences, we decided to use a triboindenter to study friction at nano-, micro-scales. This apparatus has rigid and stiff frame with potentially interchangeable capacitance force sensors for force measurement with minimum error. We also developed a simple way to characterize the tip shape and size. We used a diamond tip sliding on silicon wafer or fused silica at various loads to examine the friction variations in order to provide insights to this problem. In our study, lateral resistance forces as a function of load and penetration depth were measured.

Friction measurements were conducted using a triboindenter apparatus, shown in Fig. 3. Three 3-plate capacitive transducers are used for motion and force measurements in normal and lateral directions. In each of the 3 transducers, the two outer plates were fixed in space while the center plate was attached to metal springs and therefore moveable. The position of the center plate was determined through an AC capacitance measurement, while the force on the center place was actuated electrostatically. A probe was attached to the center plate of the middle capacitive transducer, which provided force and displacement control in the normal or z direction. The middle transducer was in turn connected to the center plates of the two other transducers located at its sides, which controlled motion along the y direction. By

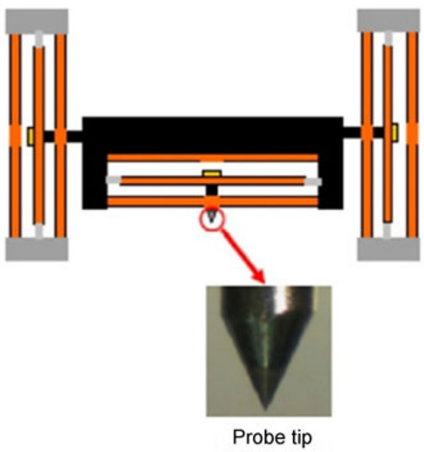

Fig. 3 Schematic illustration of capacitive sensor.

using multiple transducers of the same design, the instrument achieved independent motion control and force sensing in the normal and lateral directions. The instrument design also provided the rigidity that is desirable for quantitative friction measurements. In each friction measurement, the instrument recorded 4 parameters: force and displacement in the y and z directions, versus time. The normal displacement, or penetration depth, varied from less than 1 nm to more than 100 nm, depending on load and tip size. The travel distance in the y direction during each friction measurement was 8.0 μm ± 0.1 μm, and the travel speed was 0.2 μm/s. Careful control of experimental conditions is crucial for quantitative friction measurement. All measurements were performed with the instrument purged with nitrogen to minimize meniscus effect which could otherwise condense near the tip-substrate contact area. After friction measurements, the substrate surfaces were examined by an AFM (Bruker MultiMode).

The materials used in this study consisted of well-characterized, non-adhering diamond tips and flat substrates of fused silica (CVI Laser PW-0508-UV) and Si (100) (n type, phosphorus doped, 10^2 Ω·cm). The diamond tips are conical in shape with spherical tip diameters of 0.5 μm, 1.2 μm, and 4 μm. The substrate and tips were cleaned using ethanol, then dried with nitrogen. Experiments were performed under nitrogen purged environment to minimize humidity influence. Potential wear of the diamond tips was checked with friction tests at low loads before and after the high load tests, and the results were repeatable within 10%, suggesting negligible wear.

The shape, size, and orientation of the tips are crucial parameters for quantitative study of friction. Accurate measurements of these geometric parameters were achieved using a digital "replica" method, as described below.

3.4 Tip characterization method

A tip was pressed onto a CaF_2 crystal surface at several locations under various loads to achieve several penetration depths. The impressions were subsequently imaged using AFM. A replica of the tip was obtained by digital inversion of the AFM image of the impression

and by correction of small elastic recovery of the CaF_2 surface. The replica method allows one to examine the tip geometry in detail, including symmetry or asymmetry of the tip and smoothness of the tip surface. Moreover, the replica method has other advantages including high accuracy of tip orientation and straightforward experimental procedures.

Figure 4(a) shows a replica of a tip that has nearly spherical shape with radius $R = 1.2$ μm. The tip surface is smooth, actually with roughness no more than 3 nm or 0.25% of the tip radius. As a result, this tip can be treated as a single asperity. Another spherical tip with a radius of 0.5 μm also has a smooth surface. Figure 4(b) shows the geometry of the third tip used in the experiment. This tip shape is not symmetrical and it resembles an ellipsoid. The long axis in the surface plane, having a radius of 6.4 μm, is oriented at an angle of 45° from the y axis, the direction of the tip sliding. The radius of the short axis in the surface plane is 2.8 μm. Three tip sizes were used in order to probe the scaling issue.

3.5 Results and discussions

A series of experiments was conducted to measure the interfacial friction of diamond on Silicon and silica under controlled environment. Three diamond tips with nominal diameters of 0.5 μm, 1.2 μm, and 4.0 μm were used but in the data analysis, and they are corrected by the tip replica characterization results.

The friction forces as a function of load between the diamond tips and fused silica and silicon substrate were plotted in Fig. 5. The data for the two spherical tips with radii $R = 0.5$ μm and 1.2 μm can be clearly divided into two regions. In the first region (low-load region), the friction force shows a linear dependence on the load. While in the second region (high-load region), the friction force deviates from the linear relationship and becomes higher with increasing load. Nanoindentation and topographic studies of the substrate surfaces after friction measurements reveal the nature of this friction transition: from elastic deformation to plastic plowing with increasing load.

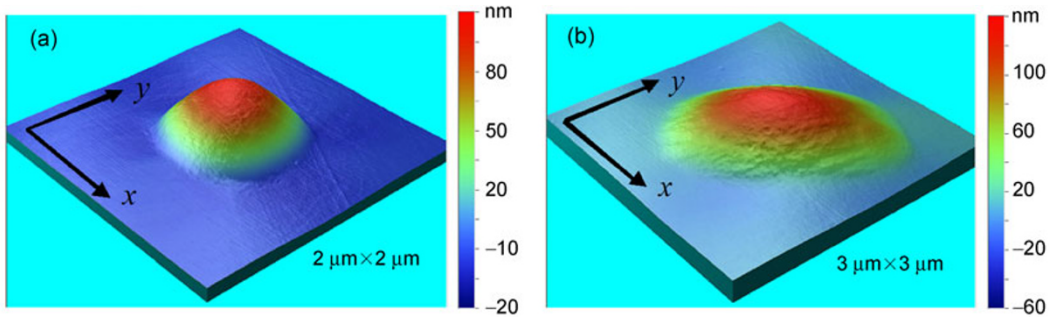

Fig. 4 Geometry of diamond tips used during friction measurements. (a) A spherical tip with radius $R = 1.2$ μm. (b) A tip with a nominal value of radius of 4 μm but actually having an ellipsoidal shape.

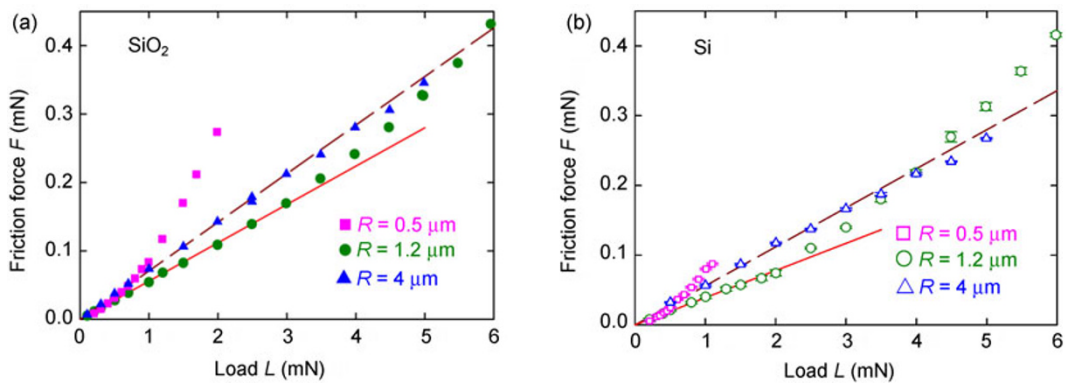

Fig. 5 Friction force as a function of load using three diamond tips with different radii R. (a) uses fused silica substrate and (b) uses silicon substrate.

As seen in Fig. 6, when a low load of 2.0 mN was applied, the loading and unloading curves for both silicon and fused silica substrates are well overlapped and free of any pop out. Moreover, the wear grooves are too shallow (~0.4 nm, near the noise level of the measurement) to be discerned, in contrast to nanometers deep grooves at higher loads where plastic deformation occurs, as shown in Fig. 7.

The elastic region shows a linear relationship between friction force F and load L

$$F = \mu_0 L, \tag{4}$$

where μ_0 is COF and the subscript denotes the elastic region. A common value of $\mu_0 = 0.055 \pm 0.002$ is observed for the two spherical tips on silica substrates (Fig. 6(a)). The plastic region exhibits friction force higher than $\mu_0 L$, load-dependent COF value, and permanent groove created on the substrate surface. The onset load values of the plastic region are 0.5 mN and 4 mN for $R = 0.5$ μm and 1.2 μm, respectively. The friction data obtained using the ellipsoidal tip with a nominal radius value of 4 μm are all in the elastic region. But the COF value is higher, 0.072 ± 0.003.

The friction data for the silicon substrate (Fig. 6(b)), are qualitatively similar. The COF value in the elastic region is 0.039 ± 0.002 for the two spherical tips. A higher value of 0.056 ± 0.003 is seen for the ellipsoidal tip. The onset load values of the plastic region are 0.4 mN and 2.5 mN for $R = 0.5$ μm and 1.2 μm, respectively.

The data presented in Fig. 5 clearly shows the different onset load values of the plastic region for different tips. The onset of plastic yield in indentation is described by a yield stress or yield pressure, so the different onset load values for different tips can be explained in terms of contact pressure, which could be calculated from the load and contact area A_c. Direct measurement of real area of contact is a difficult task. This is especially true for friction in the elastic region because the materials are recovered after friction measurements. In this study, we computed the contact area for the two spherical tips from the experimentally measured tip radius (given by the 3D replica image) and tip penetration depth, and crosschecked it with a direct experimental measurement.

Consider a spherical tip and a flat surface (see Fig. 8). A simple geometry consideration provides a relationship among the contact radius $a = w_c / 2$, tip radius R, and contact depth h_c,

$$a^2 = 2Rh_c - h_c^2 \tag{5}$$

In the limit of $h_c \ll R$, Eq. (5) becomes

$$a^2 = 2Rh_c \tag{6}$$

From Hertzian contact mechanism, the contact depth h_c in the elastic regime should be one half of the tip penetration depth h [27, 37],

$$h_c = h / 2. \tag{7}$$

Therefore, the contact width w_c for a spherical tip is given by

$$w_c = 2a = 2\sqrt{2Rh_c} = 2\sqrt{Rh}, \tag{8}$$

So the contact area can be expressed as

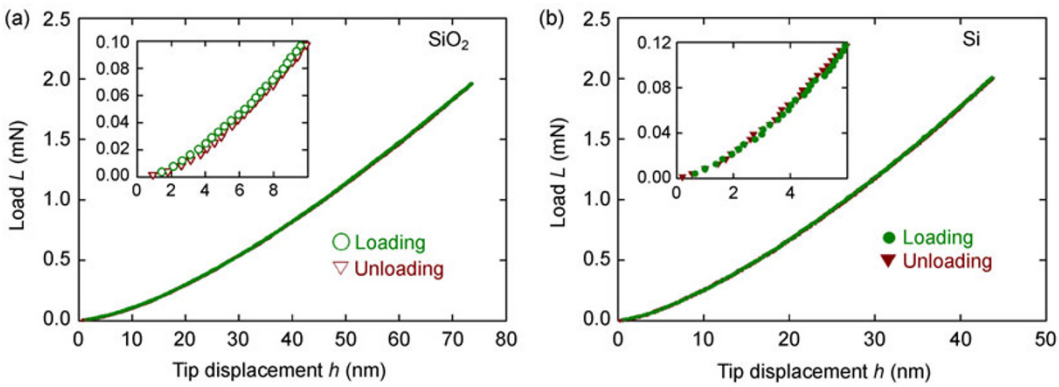

Fig. 6 Load versus displacement data measured during a single indent on silica and silicon using the tip with radius $R = 1.2$ μm.

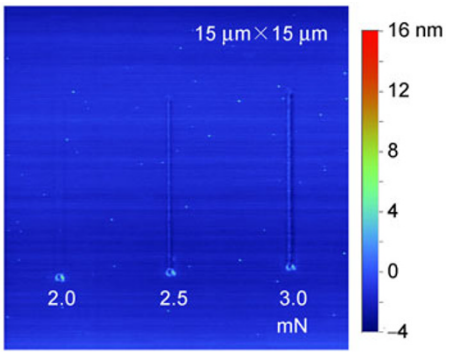

Fig. 7 AFM image of scratch marks on Si (100) at various loads using tip with radius R = 1.2 μm.

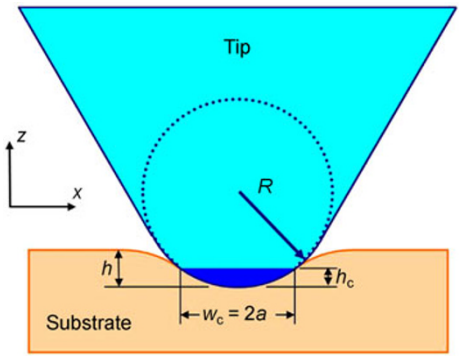

Fig. 8 Geometry of an ideal tip for adhesion and friction measurements.

$$A_c = \pi a^2 = \pi R h \tag{9}$$

Equation (9) relates the contact area to experimentally measured quantities: R and h. The tip radius R could be accurately measured by the replica method described above. The latter was measured in the sliding experiments by the TriboIndenter.

The major assumption of its derivation is Eq. (8), which was derived from continuum theory [38]. To check validity of this result, we studied AFM images of the silicon surface after friction measurements and found strong supporting evidence. An AFM image shows that the friction under 2 mN had created slight perturbation of the silicon surface with a width of 0.45 μm ± 0.05 μm. The theoretical value of the contact width based on Eq. (9) is $2a = 2\sqrt{Rh}$, assuming $h_c = h/2$. The penetration depth was measured to be h = 45.1 μm ± 0.6 nm under 2 mN load and with the 1.2 μm diamond tip, which gives a contact width of 0.47 μm ± 0.02 μm, in excellent agreement with the value measured from the AFM image. The agreement between the computed and experimentally measured contact width lays a solid foundation to compute contact area using Eq. (9) from the measured tip radius and penetration depth.

Figure 9 re-plots the COF data shown in Fig. 5, as a function of tip contact pressure. In contrast to different dependences for the two spherical tips shown in Fig. 5, the COF data for these two tip sizes now merge into a single curve in Fig. 9. The plastic region starts at contact pressures of ~ 9 GPa and ~ 12 GPa for fused silica and silicon, respectively. These values are independent of tip size and consistent with the hardness values of these materials.

Figure 9 clearly shows that the high contact pressure at the tip apex induces plastic plowing which introduces additional resistance to the tip motion, i.e., higher COF. An understanding and quantification of the plastic plowing is important because it allows us not only to avoid plastic plowing, if possible, but also to deduct the plastic plowing contribution to friction in cases that it is unavoidable.

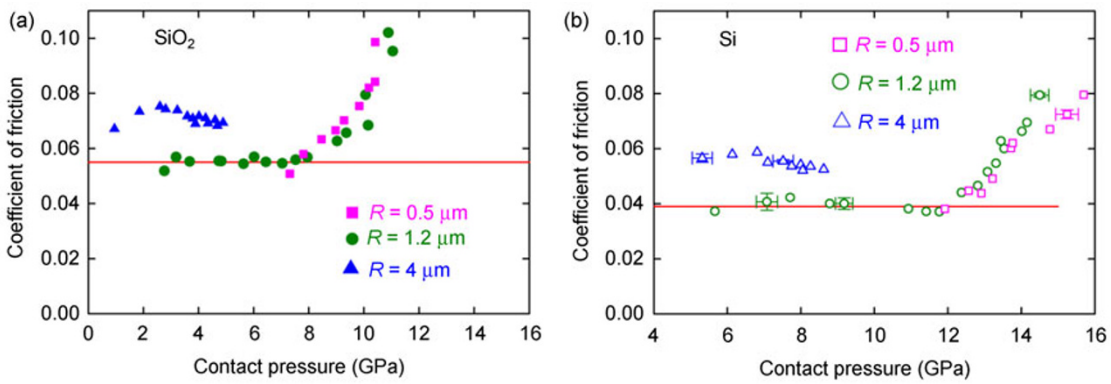

Fig. 9 Coefficient of friction as a function of contact pressure.

3.6 Plastic plowing mechanism

Plastic plowing by a sharp moving tip is a phenomenon of plastic flow of materials. It creates a continuous groove on the surface of substrate materials, with pile up of the materials along the sides and at the end of the tip travel path. It is a quite complex dynamic process in terms of materials response to stress. As a tip penetrates into a substrate and starts to slide horizontally, substrate material is dug up by the tip and accumulated in front of the tip. As the tip slides further, the material accumulated in front of the tip flows sideway around the tip, forming side pile up [39], as seen in Fig. 10. Eventually, a steady state is reached where the substrate material is continuously being dug up, first accumulated in front of the tip and then pushed to the sides of the tip. Figure 11 shows an AFM 3D image of CaF_2 surface after a friction measurement using a spherical tip of micrometer radius. The image clearly shows a groove, pile up at two sides of the groove, and pile up near at the end of the groove, where the tip was lifted up. The side pile up is nearly absent at the very beginning of the tip travel path, builds up gradually during the first ~1.5 μm of the path, and then reaches a constant shape and size.

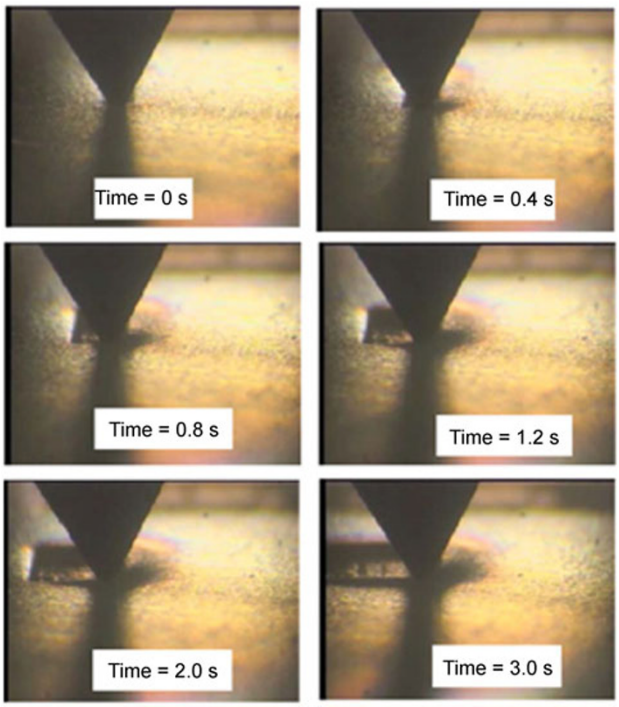

Fig. 10 Video frames of sliding process of a tip penetrating into UHMWPE [39].

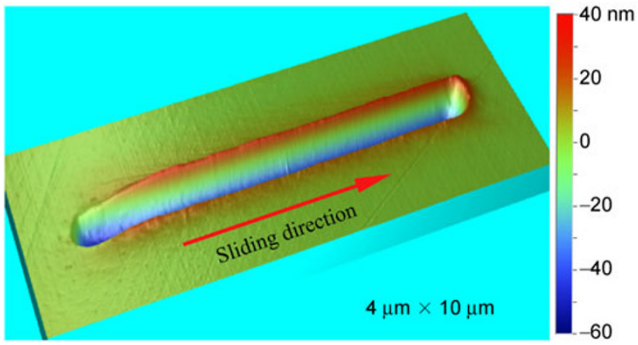

Fig. 11 AFM image of a groove created on CaF_2.

Now we are attempting to quantify the plastic plowing contribution to friction. When a tip moves at the substrate surface in the x direction by a distance d, creating a continuous groove with projected cross section area of $S(x)$, both this area and friction force are functions of tip sliding distance x. The work done to overcome the resistance due to plastic plowing is given by

$$\int_0^d F_{\text{plowing}}(x)\mathrm{d}x$$

and the displaced volume of materials is

$$V = \int_0^d S(x)\mathrm{d}x$$

An energy conservation consideration leads to

$$\int_0^d F_{\text{plowing}}(x)\mathrm{d}x = \varepsilon V = \varepsilon \int_0^d S(x)\mathrm{d}x \tag{10}$$

where ε is the energy per unit volume expected in plowing. In a special case that both the projected cross section area and friction force do not vary with x,

$$F_{\text{plowing}} = \varepsilon V / d = \varepsilon S \tag{11}$$

To quantify the plowing effect, one needs to compute volume of materials displaced by a tip. In general, there are two experimental approaches to obtain the displaced volume: (i) AFM imaging of the substrate surface after friction measurements and (ii) computation from the tip geometry and tip penetration depth. The latter requires the precise measurements of the tip size and shape in the direction of sliding with the tip penetration depth data throughout the plowing process. Depending on materials, corrections to volume displaced for elastic recovery may be

necessary to avoid over-estimation of the displaced volume. Part of the material displaced may be elastically deformed but recovered after the tip passes over. Since the present study concentrates on the plastic deformation, the volume measured from the AFM image should be accurate.

In computing the volume from the AFM data, one could use either the volume of the groove or the volume of the pile up material around the groove. The volume of the groove corrected for the surface plane baseline was used.

As discussed above, the data in Fig. 5 can be clearly divided into elastic and plastic regions. So the quantification of plastic plowing should only compute the additional work done by the enhanced friction force, which is defined as the magnitude of friction force deviated from the linear friction relation, and compare it with the volume of groove left on substrate surface. In a simple case that neither the friction force nor groove profile varies as the tip slide, one could simply compare the enhanced friction force with cross-section area of groove (Eq. (11)).

Figure 12 plots the friction force versus cross-section area of groove on fused silica. A linear relationship can be clearly observed. Fitting the data to a straight line yields a proportionality constant of ~ 11.4×10^{-6} mN/nm^2 (or ~11.4 GPa). The same value of the proportionality constant appears to be applicable to both tips with radii of 0.5 μm and 1.2 μm. The value of 11.4 GPa is greater than but close to the hardness of fused silica.

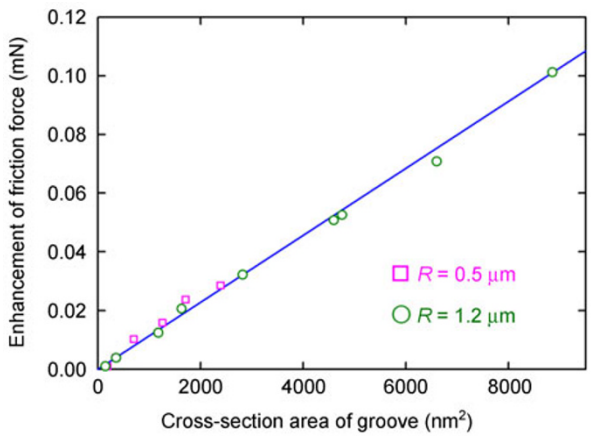

Fig. 12 Friction force versus cross-section area of groove on silica.

One possible explanation is that the enhanced friction force reflects the additional lateral resistance experienced by the tip because it needs to push the substrate material in front of the tip. The literature value of cohesive energy is 2.5×10^{10} J/m^3. Comparison between these two data (11.4 GPa = 11.4 J/m^3) suggests that approximately 45% of the bonds were broken during the plastic plowing process. This fraction appears reasonable. Our experimental data presented here show that the total friction force can be expressed as:

global frictional resistance = interfacial friction + plowing friction

The plowing contribution to friction could be estimated from the material hardness and displaced volume. The upper limit of the plowing contribution can be obtained by equating parameter ε in Eq. (11) with the cohesive energy.

$$\mu = \mu_0 + F_p / L = \mu_0 + \varepsilon V / dL = \mu_0 + \varepsilon S/L \qquad (12)$$

where, F_p is the plowing contribution to friction force. Numerically, this contribution to friction force is equivalent to the energy necessary for the tip to displace the materials inside the groove. The proportionality constant ε is a materials specific parameter that is mechanistically associated to bonding breaking and is close to hardness. In addition, the principal component stresses also cause dislocation concentration, micro-cracks to form, grain alignment, etc. This produces a wide variety of friction values even though the same materials pair and experiments performed by the same person.

3.7 Friction scaling laws

Table 1 shows for the same materials system, using different instruments to measure friction at different length scales produces wide range of friction levels [31]. Do these values suggest an intrinsic and implied scaling law for friction?

The results in the table appear to show a scaling effect, i.e., different levels of friction at different scales (nm, μm, and mm). However, an un-intended plowing by the sharp tip was observed at $\mu = 0.25$ [31]. Based on our results, does the observed scaling effect come from measurement artifacts?

Our results in Figs. 5 and 9 show different COF

values for different tip sizes even in the elastic region. In order to explain this phenomenon, let us analyze the relationship between friction force and contact area (Fig. 13). Calculation of the contact area has been described in detail above. As seen, in both cases of silica and silicon substrates, the friction force does not exhibit a linear relationship with the contact area. In addition, the friction forces at a given contact area are different for different tip sizes.

The observed dependence of friction force with contact area can be understood on the basis of Hertz theory of contact mechanics and a linear friction force-load relationship. Combining Eqs. (4) and (9) with Hertz theory of contact mechanics,

$$h = \left(\frac{3L}{4E_r}\right)^{2/3} \frac{1}{R^{1/3}} \tag{13}$$

where E_r is the reduced modulus of the system, one obtains,

$$F = \frac{4\mu_0 E_r}{3R}\left(\frac{A_c}{\pi}\right)^{3/2} \tag{14}$$

The lines in Fig. 13, drawn based on Eq. [14] without free fitting parameters, are consistent with the experimental data, which provides strong support for the non-adhering nature of the materials studied here. Equation (14) contains three parameters: μ_0, E_r and R. The value for μ_0 is obtained from the friction data in elastic region presented in Fig. 5. The reduced modulus E_r is obtained through analysis of indentation data.

Friction is a phenomenon with relative motion between two contact surfaces in the lateral direction. So contact area in the lateral direction is the projected cross-sectional area of the penetrated portion of the tip in the substrate in the sliding direction. This quantity for a spherical tip is given by

$$S = R^2 \cos^{-1}(1 - h_c / R) - (R - h_c)\sqrt{2Rh_c - h_c^2} \tag{15}$$

In the limit of $h_c \ll R$ or for a parabolic tip, this equation is simplified to

$$S = \frac{4}{3}\sqrt{2Rh_c^3} \tag{16}$$

From Eqs. (4), (7), (13) and (16), one obtains,

$$F = 2\mu_0 E_r S \tag{17}$$

It means that the friction force in the elastic region should be proportional to the projected cross-sectional area for a spherical tip. The proportional constant is dependent on the coefficient of friction in the elastic region and reduced modulus, but independent of tip size. It has a physical meaning as the lateral pressure as the tip slides through the substrate surface elastically.

Plotting of friction force versus projected cross-section area is shown in Fig. 14 for silica and silicon substrates. As expected from Eq. (17), a linear relationship is observed for both systems and for the two spherical tips with radii R = 0.5 μm and 1.2 μm. With the 3D profiles of the tips obtained using the replica method, we were able to compute projected cross section area for an arbitrary tip shape, which was done for the ellipsoid tip that has a nominal tip radius of 4 μm. As shown in Fig. 13, the data from this tip follow a common line with that of the spherical tips for each substrate. This result suggests that Eq. (17)

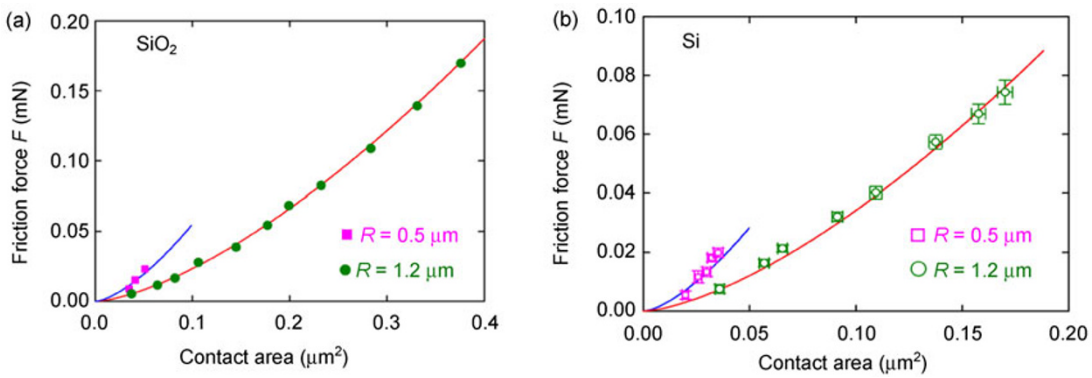

Fig. 13 Friction force as a function of contact area. Only the data in the elastic region are shown.

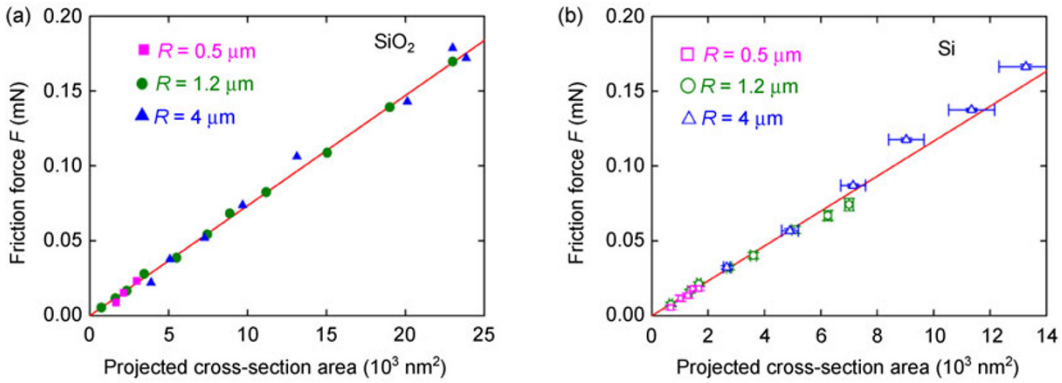

Fig. 14 Friction force as a function of projected cross-section area in the tip sliding direction. Only the data in the elastic region are presented.

might be modified to accommodate other tip geometric shapes, limited by a convex surface with sufficient smoothness.

The behavior of the friction within the elastic limit could be explained by a model, which recasts the Amontons' law into a form showing frictional force to be proportional to the projected cross-sectional areas in the direction of sliding. This dependence on the cross-sectional area reveals that this is a volumetric dependence instead of just the contact area in the normal direction as commonly assumed by other. The volumetric dependence includes depth of penetration, irregular tip shape, and movement during sliding. All these factors have to be taken into account before the data fall into line. This would suggest that stored elastic strain energy during sliding is a significant part, if not the dominant part of the frictional force in the elastic regime. This observation also experimentally verifies the concept proposed by Ref. [31] that nano behavior can be achieved from either tip size or depth of penetration perspective. To sum up, the dependence of COF on tip geometry, as well as plastic plowing in the high load condition, is likely to be the major causes of anomalous behavior in nanofriction measurements.

3.8 Nanofriction formulation

Studies on nanofriction are not complete without understanding the origin and nature of atomic and molecular friction. There are many notable contributions in the literature providing insights into atomic force fields and atom-probe tip interactions [14, 27, 38, 40–43] and to some extent, the intrinsic limitations of instrumentation to probe atomic friction in real time and in sync with the atomic electron movements.

Nanofriction fundamentally is measured by sharp tips mounted on a cantilever approaching a surface. The cantilever stiffness determines the sensitivity of the measured displacement by the laser beam reflected off the cantilever to a photodiode sensor. Because of the measurement sensitivity and the short distance involved, various surface forces are involved, such as van de Waals force, capillary force in air containing trace amount of water, electrostatic charge force in the presence of the semiconductor material, various chemical bonding forces including hydrogen bonding, short range and long range molecular bonding forces. In a measurement, not all forces are present or active. The common method of using magnetic holder for samples invariably introduces electromagnetic force field. Therefore, careful calibration procedures including known sample and known spring constants level are critical in obtaining accurate data (standard reference materials are produced by National Metrology Labs such as the NIST and EU Labs).

Therefore, nanofrictional resistance (force) can be represented as follows:

$$F = \mu L + \tau A + \varepsilon V + \sigma V + \nu B \pm \xi A \pm \psi d \qquad (18)$$

where μL=interfacial resistance due to load;
τA=interfacial resistance due to adhesion;
εV=interfacial resistance as a result of plowing;
σV=interfacial resistance to motion due to meniscus force;
νB=interfacial resistance due to chemical bonding

force;

ξA=interfacial resistance from electrostatic forces;

ψd=molecular attractive or repulsive force depending on distance between surfaces.

We have been conducting long term research to isolate one parameter at a time to obtain quantitative relationship among the various forces.

3.9 Summary on the scaling issue

Based on the results we have generated, as mentioned in Ref. [31], nanoscale contact friction measurements are often interfered by plowing and tip shape, penetration depth. When all these factors are properly accounted for, constant interfacial friction can be obtained. In the macroscale, friction measurements, microstructural change from deformation, strain hardening, and microcracks basically change the materials properties and defect populations, hence the friction is governed by the deformed material, and no longer governed by the pristine nanomaterial.

In molecular dynamics models, the pairwise potential energy level is governed by the internal thermodynamic energy states. When the surfaces are in contact, the contact stresses induced various levels of strain in the atoms, molecules, and crystalline domains around the contact area, and this strain energy necessarily changes the energy stored in the atoms and molecules of the affected area. Accommodation of the various deformation strain energy due to contact poses location specific simulations. So there is a gap between the nano- and macro-regimes. When microstructures begin to change, nanometer models based on pristine atoms and molecules would not be able to describe the frictional event or predict the outcome.

Figure 15 illustrates that for friction, there are potential discontinuities in the scaling relationship. Recent studies on mechanical properties of small structures below 100 nm dimensions reveal many systems exhibit size dependent mechanical properties [44–47], which deviates from the conventional continuum mechanics approach. Friction being a complex system function in most cases, therefore, it should exhibit size dependent values.

At the same time, bridging the gap between nanoscale events and macroscale phenomenon has always been a goal for many tribologists. Many investigators have continued to do so [48–52].

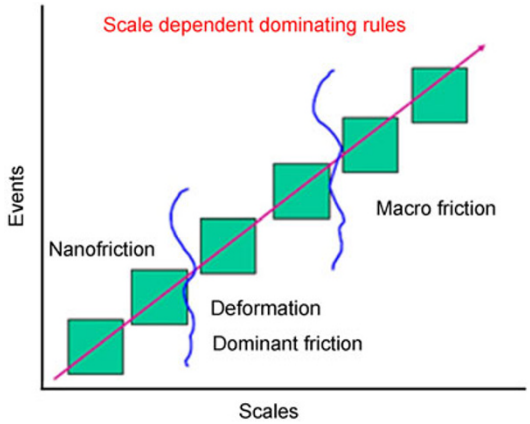

Fig. 15 Scale specific friction levels.

4 Friction under lubricated conditions

In engineering applications, most systems are lubricated with liquid lubricants specially designed for the intended application. Over the years, lubricated systems have been able to deliver reliable performance for the intended duty cycle of the machinery. Since the beginning of the lubricant development, the primary goal for the lubricant is to protect the system from wear and premature failures. Friction has not been a primary focus for lubricant design. After the 1973 oil embargo, friction reduction through the use of friction modifiers was introduced into the modern lubricant formulations for cars and trucks. The fuel economy improvement is small around 1% and oil degradation over time also deteriorates the fuel economy benefits. Today, the impending liquid petroleum shortage as fuels for the internal combustion engines again looms as one of the grand challenges of our time and friction under oil lubricated condition emerges as a vital issue.

Why lubricants have been so effective in preventing wear and damage? Through out this paper, we have been discussing the importance of real area of contact that justifies the Amontons' laws that we learn early in our education. In our study to understand frictional processes, we designed and built a two ball collision apparatus to observe the surface before and after the collision using a high speed video camera (Fig. 16(a))[19]. We used various materials combinations such as

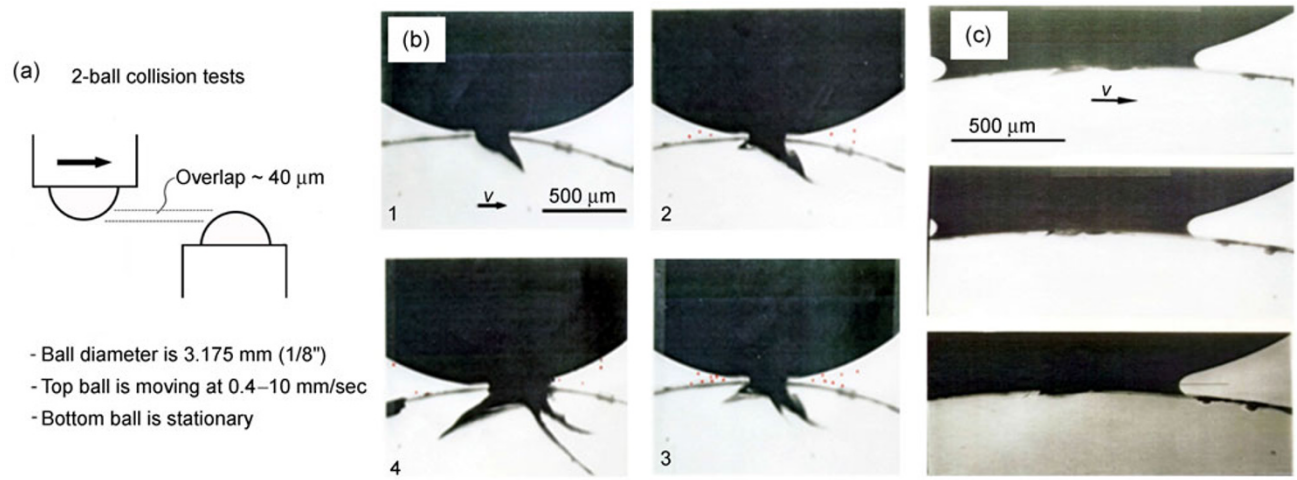

Fig. 16 Steel on quartz under identical one pass collision condition: (a) shows the collision apparatus and operating conditions; (b) is under dry collision condition; (c) is under paraffin oil lubricated condition (without additive). Note the propagation of cracks as the collision progress in (b) and no crack propagation appears in (c). Microcracks appears in (c) but self-healed due to hydrostatic pressure closure forces.

quartz ball and disk (to be able to observe the internal strain fringes), steel on copper, copper on copper, and 52100 bearing steel on 52100 bearing steel, ceramics on ceramics, etc. under dry, paraffin oil lubricated, and fully formulated commercial lubricants. Under dry collision, depending on the materials, real area of contact is small and the asperity stresses are very high (Fig. 16(b)). The stress intensities and associated strain are very high, causing tensile microcracks under low load, and mode 2 and mode 3 fractures at high load. For the same conditions, when a liquid lubricant is present, real area of contact approaches, the apparent area of contact, and the contact pressure are evenly distributed, lowering the contact stress intensities and internal strain dramatically (Fig. 16(c)). At high velocities, the hydrodynamic lift force set in and the stress intensities are lowered further. Details of the equipment design and other studies are shown in Refs. [53–55]. This is the basic lubrication mechanism of lubricated contacts.

4.1 Difference in friction and wear data analysis between dry and lubricated contacts

Tests run under this "dry" sliding condition tend to be dominated by the material properties of the two opposing surfaces and the surface roughness. Dry friction often follows a "linear" dependence as a function of time when the dominant wear mechanism remain the same, as shown in Fig. 17(a). The wear community tends to use pin-on-disc as a primary tool for evaluating wear with accompanying friction data reported under constantly wearing conditions. As such, the wear is relatively severe. Wear data under this kind of sliding conditions exhibit wear volume as a function of time or the distance slid. The use of the wear coefficient, K is a prime example:

$$K = W_v H/LD \tag{19}$$

where W_v is the wear volume, H is the hardness, L is the load, and D is the distance slid. The assumptions are that wear is proportional to load and distance slid and inversely proportional to the hardness of the surface material being removed. In dry wear studies, this equation and the wear coefficient have been used so frequently that sometimes the basic assumptions behind the equation are overlooked.

In the wear characteristics in lubricated contacts, there are two types of "equilibriums" in the tribosystems: (1) the fluid film generates sufficient hydrodynamic film pressure to support the load; (2) the chemical boundary film which is designed to be worn off (sacrificial wear) regenerates rapid enough to balance the shear induced wear motions. Under both circumstances, a constant wear rate (which can be very low level) or a zero wear rate can be maintained, as shown in Fig. 17(b). In this case, expressing wear

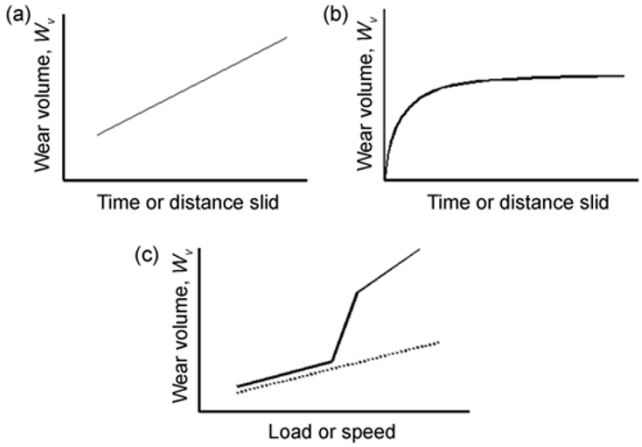

Fig. 17 Illustration of well lubricated wear as contrast to "dry" wear.

volume as a function time or distance slid has no meaning but to make the wear rate smaller and smaller, causing misrepresentations. Wear will increase dramatically when the boundary lubricating film fails to sustain itself due to various breakdown mechanisms, such as additive depletion, lubricant starvation, scuffing, etc. When the film breaks down, wear (and friction) rises rapidly to a new level of equilibrium, as shown in Fig. 17(c). While wear is not the subject being discussed here, friction under lubricated conditions is strongly affected by the wear mechanisms occurring inside the contact.

Friction in lubricated systems is caused by two mechanisms: fluid dynamics dominated load support conditions (hydrodynamic and elastohydrodynamic lubrication regimes); and surface chemical film dominated regime (boundary lubrication conditions). Most practical tribosystems are designed to operate within the fluid film support regimes but the start- and stop-process points will inevitably in the boundary lubrication regime. In the following sections, we will explore briefly the fluid dynamic regime then focus on the boundary lubricating films and their friction characteristics.

4.2 Fluid dynamic controlled friction

Within a sliding bearing, if the surfaces are separated by a continuous fluid, then the relative surface roughness is not a factor. The primary friction is from viscous shear of the fluid or lubricant in the contact. The friction can be described by a Stribeck curve as shown in Fig. 18.

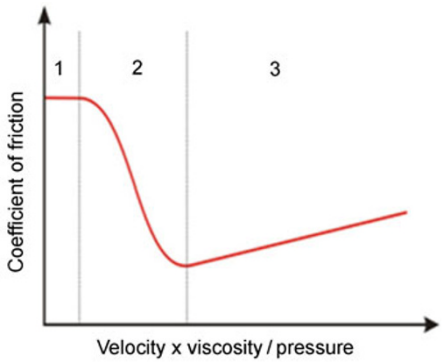

Fig. 18 A schematic diagram of the Stribeck curve.

The Stribeck curve was developed by Richard Stribeck in 1902 using a journal bearing to analyze friction data as a function of viscosity, speed (rotational speed) and load. He found three regimes with different levels of friction. The diagram provided a comprehensive view of the journal bearing operation under various speeds and loads. Over the years, the use of Stribeck diagram has gained in popularity to describe the whole lubrication regime. The specific friction level depends on viscosity, oil formulation, contact geometry, surface roughness, and the nature of contacting materials.

As shown in Fig. 18, regime 1 describes the friction under boundary lubrication conditions, and the COF ranges from 0.08 to 0.15 where the load is supported by asperity contacts. Regime 2 is the elastohydrodynamic lubrication where asperities of the surfaces are in occasional contacts but deform primarily elastically, and COF from 0.05 to 0.08 is typical. Regime 3 is the hydrodynamic lubrication regime where a continuous fluid film separates the two sliding surfaces, and the COE can range from 0.05 to 0.01 or below.

Theories for hydrodynamics and elastohydrodynamic lubrication are well developed [56–61]. The equations are based on Reynolds Equations and they need to be adjusted for the contact geometry, materials properties, and speed and load operating conditions. For elasto-hydrodynamic lubrication, a continuous fluid film is still assumed but occasional elastic asperity–asperity contacts are taking into account. Thermal effects from the occasional contacts will have to be taken into account as the frequency of contact increases towards the boundary lubrication regime.

Under boundary lubrication conditions, most of

the load is supported by asperity contacts and a thin fluid flow meanders through the maze of contact points. Friction becomes independent of the viscosity of the fluid. Chemical reaction products between the lubricant and the surface are controlling the friction and wear processes. Friction in the mixed lubrication to boundary lubrication regimes depends on surface roughness, boundary chemical films, chemical kinetics of the film formation, and frictional property of the film.

The wear scar shown in Fig. 19 is a typical lubricated boundary lubricating film formed by the zinc dialkyl-dithiophosphate (ZDDP) on 52100 bearing steel surface. The picture was taken after one minute of running time, the wear scar was rinsed with hexane to remove excess oil. Already, one can see the brownish coloration indicating the presence of an organic film interspersed with darker spots, described by some people as wear pads formed from ZDDP reactions. Notice the striations scratch marks along the horizontal axis, the high speed video tape showed a single wear particle detached from the edge and scratch horizontally in the sliding direction creating a groove during the initial wear-in process. This fundamental scratching process occurs under dry friction, lubricated friction, and nanofriction plowing. This will be further demonstrated later.

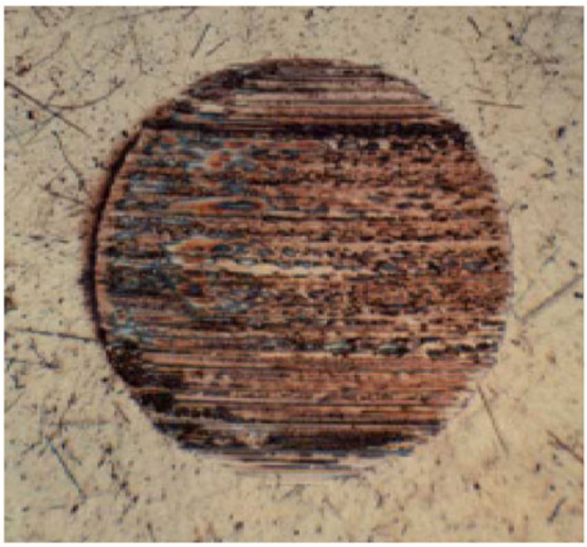

Fig. 19 A lubricated wear scar in a four ball wear test conducted under ZDDP added paraffin oil at 40 kg, 600 rpm, and 25 °C, after one minute of sliding. The wear scar diameter is 0.55 mm.

4.3 Friction under boundary lubrication regime

In boundary lubrication, surface chemistry plays a critical role. Our understanding of the complex interplays among surface roughness, chemical film generation rate, and frictional properties of the film is reasonable, but how properties link to lubrication at the asperity level is not clear at this time. Therefore, it is no surprise that we do not have a friction model capable of predicting boundary friction at this time.

The fundamental processes that generate friction under boundary lubricated conditions are: (1) asperity–asperity contacts or collisions; (2) sliding friction from "asperity sliding in a groove" frictional resistance as shown in Fig. 19; (3) plowing of the detached particle from edge (edge stresses) across the sliding path, as shown in Fig. 20 below; (4) strain resistance to asperity/particle plowing and deformation; (5) viscous drag from lubricant inside the contact (small as compared to others; and (6) elastic and plastic strain of the antiwear pads generated from the antiwear additives.

In an effort to understand the onset of wear and the associated friction increase, a two ball collision test apparatus was used [19]. The system was lubricated with pure paraffin oil and the force traces from the x, y, and z axes were shown. Figure 20 shows a single particle generates from the edge (Fig. 20(b)) and slides across the surface, then it is trapped and starts to scratch the opposite surface (they are mirror image of one another) and then exits the contact. The F_t/F_n curve shows the gradual increase of the friction and at exit, a friction peak was observed for the exit process.

4.4 Influence of wear modes on boundary friction

Boundary friction is also closely linked to wear and its contribution to the friction, compounded by the various chemistry associated with corrosion, fatigue, electrochemistry, and scuffing. Beerbower [62, 63] conducted a comprehensive survey of the literature, consulted with large number of experts to come up a picture of boundary lubrication, its models, mechanisms, and wear failures in 1971. He developed a diagram delineating the various modes of wear and lubrication mechanisms as a function of specific film thickness, as shown in Fig. 21.

Even though this diagram was developed some 40

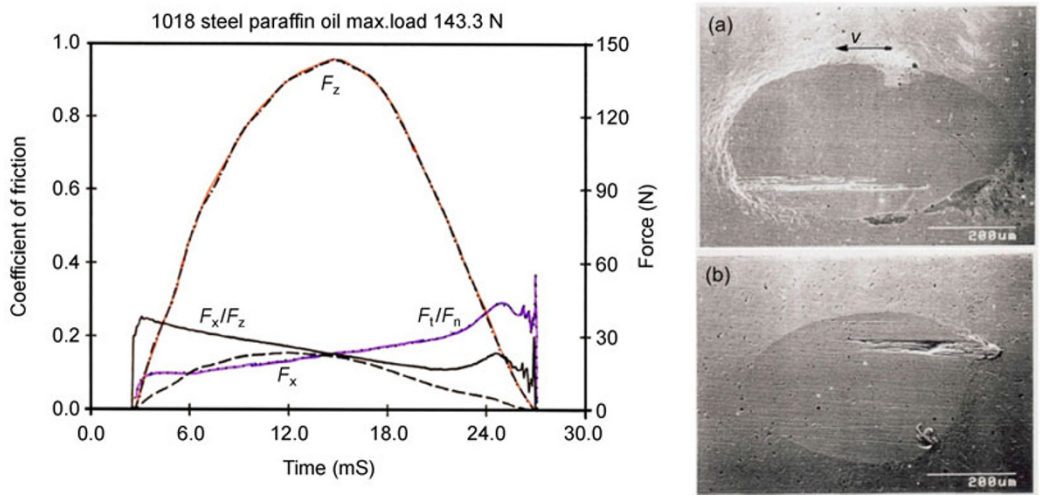

Fig. 20 A 52100 steel ball collides with another 52100 ball under lubricated collision, the force traces are shown in the graph. The two surfaces are reciprocated image. (a) is moving and (b) is stationary. It shows a particle is plowing the contact producing high friction F_t/F_n (F_x/F_z is recorded from the force transducer but needs to be corrected by contact geometry).

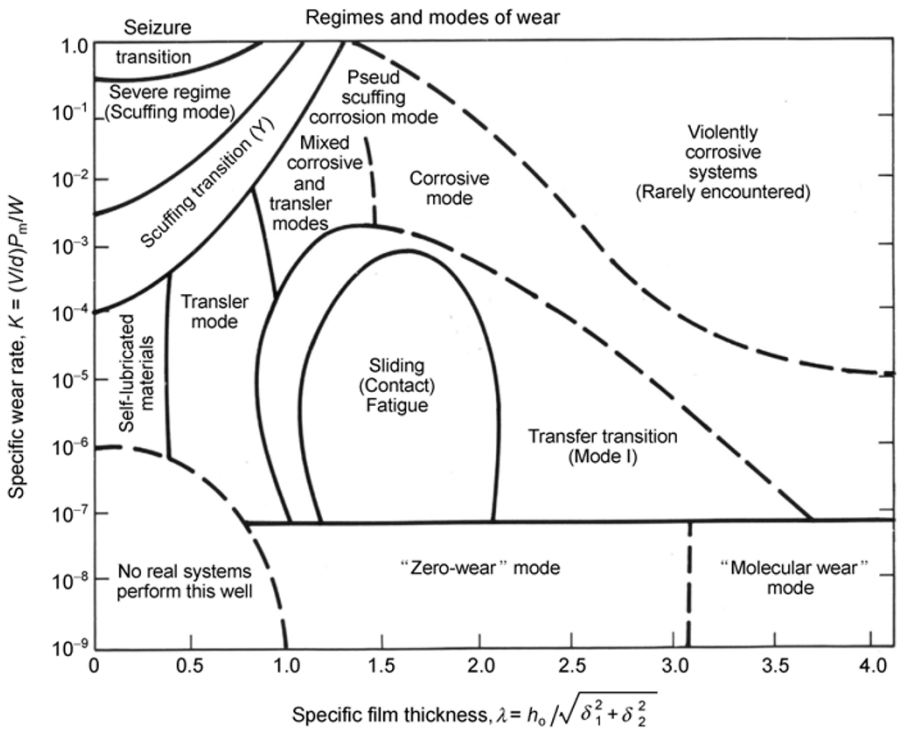

Fig. 21 Various wear mode in boundary lubricated conditions [62].

years ago, it still represents one of the most comprehensive descriptions of boundary lubrication and associated wear modes. Of course, each wear mode has a specific characteristic friction. In the dry wear community, high friction is always associated with high wear, but in boundary lubricated systems, low friction can be associated with high wear, for example, in the corrosive wear regime. Also, in the transmission friction plate system, the system maintains high friction for torque transfer but low wear.

4.5 Boundary friction influenced by film chemistry and tribochemistry

Lubricant chemistry and antiwear additives play a

significant role in controlling friction when the two sliding surfaces are pressed together under high load. Lubrication is achieved by the chemical reaction products between the surface and chemical species in the lubricant. The most commonly used antiwear additives are phosphorus-based chemical structures: ZDDP and tricresyl phosphate (TCP). ZDDP also functions as antioxidant but tend to produce oil insoluble products and is commonly used in engine and motor oils. TCP is effective in wear protection but does not form oil insoluble products and is commonly used in industrial oils. They both form very effective antiwear film but essentially do not modify friction as compared to the baseline cases. This is because if the function of the antiwear film is to provide sacrificial wear to protect the substrate, low frictional films generally do not have strong adhesion with the substrate to resist repeated high shear structural strength. For this reason, friction modifiers are molecules adsorbed on top of the antiwear film to provide lower friction under hydrodynamic or elastohydrodynamic lubrication regime, or to adsorb onto non-contact surfaces to lower the hydrodynamic friction.

Tribochemistry is the chemistry took place inside a sliding contact and is the fundamental science underpinning the modern lubrication science. To achieve effective lubrication under boundary lubrication conditions, organometallic compounds are formed under the sliding conditions and condensed into high molecular weight products [64, 65]. These high molecular weight products are essential to provide a high viscosity near surface layer to support the wear pads formed by the antiwear additives. The question is if the antiwear additives do not modify friction, then what controls the boundary lubrication friction?

4.6 The nature of boundary friction: Thin film friction + plowing/deformation

Similar to our previous analysis of dry friction, boundary lubricated friction can be viewed as it consists two parts: pure sliding friction due to the monolayer film adsorbed or bonded to the surface; and frictional resistance stem from abrasion, plowing, deformation, and fracture processes. The latter is much bigger in magnitudes and the energy dissipated is much larger. The complexity of the plowing and deformation induced energy dissipation makes it difficult to predict since it is material dependent and influenced by the operating conditions and chemical environment. The monolayer film friction is small but is the critical fundamental factor in determining the basic frictional properties of the contact.

Bowden and Tabor [15] first examined this issue by coating surfactants on glass slides and to measure the effect of the monolayer of molecules on friction. Results showed that longer chain surfactant molecules have lower friction and there was a minimum chain length to be effective. Their measurement was limited to some extent by the sensors and instrument available at that time. Since then, many researchers have repeated such experiments using surface force apparatus and atomic force microscope on mica and atomic flat surfaces [23, 24, 66, 67]. Friction of monolayer films turns out to depend on adhesion and cohesion as well as film thickness relative to the composite roughness. Adhesion is the bonding strength of the molecules with the substrate and cohesion relates to cross-linking, molecular order, phase structure or defect populations inside the film. These carefully conducted studies provide insights into the basic molecular structural effects on ideal surfaces. But the chemical bonding between mica and surfactant molecules is very different from lubricant and steel surfaces; as well as the surface topography of engine components is very different from homogeneous atomically flat mica or pyro lytic graphite surfaces.

To bridge this gap to measure monolayer friction on realistic surfaces, we designed and built a high precision apparatus based largely on the apparatus described in Ref. [18] but more precise and sensitive. The design is shown in Fig. 22. The apparatus is designed for the purpose of measuring film adhesion and cohesion characteristics. The concept is to use a highly polished flat plane squeeze against a rigidly mounted ball using a high precision x–y stage. The load is controlled by the inclined angle of the plane. Force transducers were mounted on the stationary ball housing and forces in x, y, z directions are continuously recorded. For monolayer or nanometer thick chemical film, the inclined angle (or the load) has to be controlled very precisely. It is desirable to

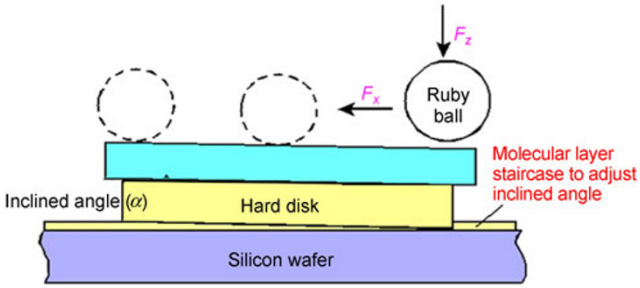

Fig. 22 Schematic diagram of the film shear rupture strength instrument.

be able to control the inclined angle to 0.001 degree consistently depending on the roughness and smoothness of the sample planes. We developed a molecular "staircase" of dilute thermosett polymer dip-coated sequentially to create a staircase step by step. This technique was successful to control the inclined angle to 0.001 degree. For most of the samples, an inclined angle of 0.1 degree is sufficient.

Since the roughness of the test samples (silicon wafer, single crystal of iron, and polished steel plate) is higher than the monolayer of molecular film, so the test design is to have a large area of contact with calculated elastic deformation of asperities for the monolayer to function. Therefore, the diameter of the stationary ball can be from 3 mm diameter to 12.7 mm diameter, depending on a specific system to be measured. Balls of various materials such as ruby, quartz, silicon nitride, and steel bearing balls were used.

The flat samples were polished and cleaned in a class-1000 clean room then transferred to a class 100 clean room for dip-coating of purified organic molecules (percolated through silica gel and activated alumina columns). All solvents and reagents were cleaned similarly. The tests were conducted in the class 100 clean room with vibration isolation platform. These precautions were necessary since in the early stages of the study, the samples were heavily contaminated by dust particles from air, which dominate the friction characteristics masking the effect of the molecules.

The thickness of the dip-coated film on the test sample was calibrated using Fourier Transformed Infrared glancing angle spectroscopy with known thickness standards. Ellipsometer was also used to cross calibrate the thickness. Various substrates such as silicon, iron film deposited on silicon, and polished single crystal iron were used. Only selected data from that study were shown here to provide insight on boundary lubrication friction.

The test proceeds as the load continuously increases. When the film ruptures, the friction force trace jumps up. At that point, the motion stops. A picture was taken at the spot where friction jumps and the contact area is measured. Knowing the normal force, the contact pressure was calculated and the film rupture pressure is reported. Figure 23 shows the measurement of the film rupture strength of six molecular films.

The film rupture strength increases with chain length, size of the molecules, and active functional groups. The film rupture strength reflects both adhesion and cohesion of the film binding to the specified surface, in this case, iron. The data are reasonable and are in agreement with lubrication experience, validating the measurement technique.

When the surface is changed from iron surface to silicon, aluminum, copper, and titanium surfaces, the adhesive bonding strength of the pure paraffin oil film as measured by the film rupture load drops dramatically, as shown in Fig. 24. This is collaborated that iron has much higher chemical reactivity towards hydrocarbon molecules [68].

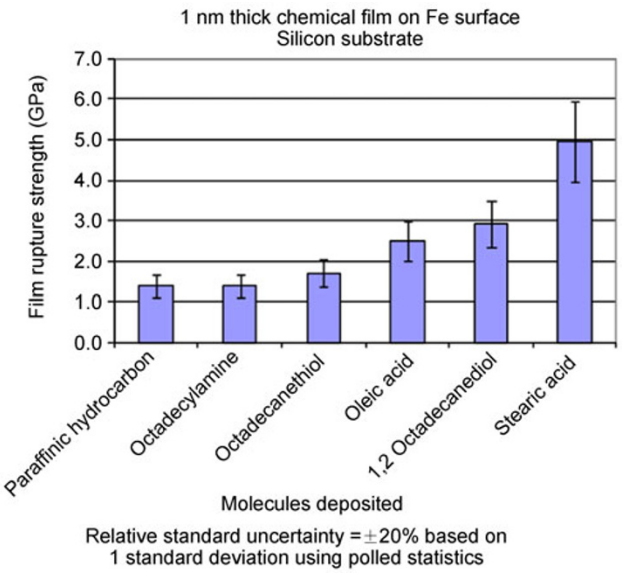

Fig. 23 The film rupture strength of six one nanometer thick films consists of different functional groups, using a diamond tip with spherical shape of 3mm diameter.

To test the effect of chain length and cross-linking, we tested the same material system and varied the chain length and the cross-linking of the molecules on the surface. These experiments were conducted using the dip-coating process and depositing the same thickness of film on the substrate. Results are shown in Fig. 25.

When we tested the antiwear additives such as tricresyl phosphate and zinc dialkyl dithio-phosphate, on this one pass test, they do not show much film strength and the friction levels remain the same as the baseline case (paraffic oil film). However, when the films are heated or rubbed, the film strength increases dramatically. This agrees with the known mechanism of these additives that antiwear films are formed when the antiwear additive is thermally decomposed forming acid phosphates and phosphites, which react with the iron surface forming tenacious sacrificial antiwear films.

4.7 Summary on lubricated boundary friction

Under lubricated conditions, when the surfaces are separated by a full fluid film, friction is dominated by viscous shear resistance [59]. Predictive equations are available. When the load is supported largely by surface contact, friction depends on the boundary lubricating films formed from the lubricant/additives and the contacting surfaces. Since the primary objective

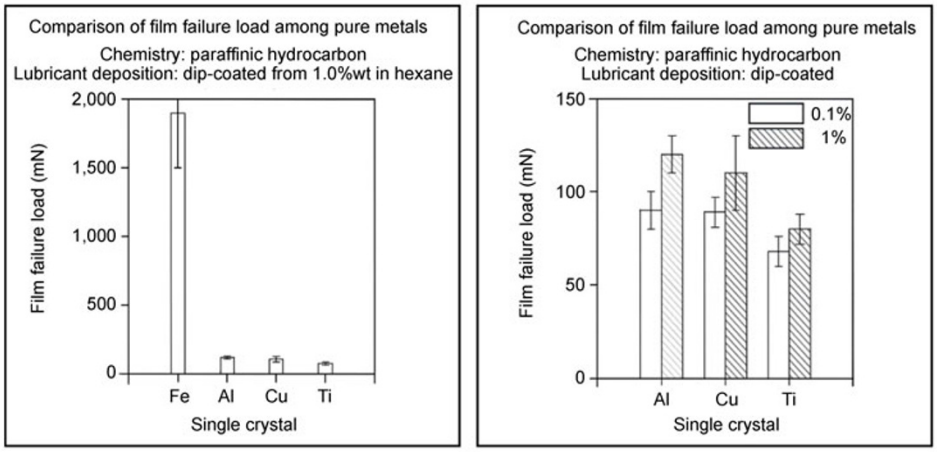

Fig. 24 Comparison of film rupture load among pure metals.

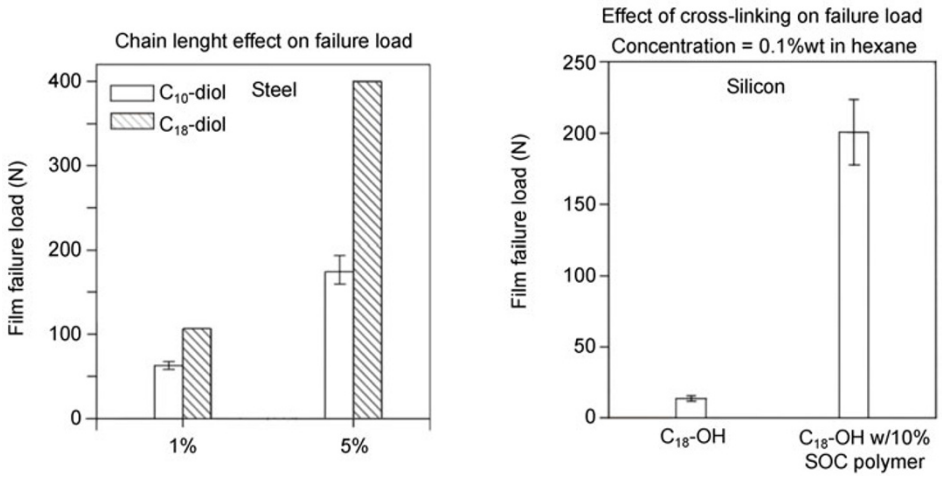

Fig. 25 Effect of chain-length and cross-linking on failure load.

of the boundary lubricating film is to protect the surface from wear and seizure via sacrificial wear, friction is not the dominant objective of the lubricant chemistry design. At the same time, by examining the boundary lubricated surfaces and by video observation of the two ball collision sequence, the presence of plowing and abrasion appears to dominate the boundary lubrication high friction events, as described above. Based on this observation, we propose there are two components to boundary lubricated friction: one is the interfacial friction with asperities in contact; one is the plowing, deformation, and other energy dissipation processes associated with wear and fracture, similar to the dry lubricated sliding case. The difference between the two cases is the real area of contact: under dry sliding condition, the real area of contact is extremely small; under boundary lubricated sliding, the real area of contact can range from 20%–70% [4].

In order to understand the baseline friction from asperity supported load conditions, we designed and built several ball on inclined plane testers to measure lubricated films from monolayer to multilayer, and up to micron-meter scale lubricating films below or at the composite roughness of the two contacting surfaces. Under the asperity load-supported conditions, coefficient of friction of paraffin oil, alcohols, acids, etc. (typical lubricating oil species) ranges from 0.08 to 0.12 on a range of surfaces, including quartz, silicon, single crystals of iron, copper, and titanium, and polished 52100 steel surfaces. Multilayer film or thick films tend to have higher durability but the frictional characteristics remain the same.

Modeling of boundary lubrication and friction has been attempted by many [7, 69–71] but has so far been unsuccessful towards a universal predictive equation. This paper, hopefully will spark some new thinking towards that goal.

5 Conclusions

Friction as an indicator of energy efficiency today has revived interest in understanding its origin and means to control the process through predictive models. Yet because of its complexity and multidisciplinary nature, progress has been slow. This paper reviews the historical perspective highlighting the significant advances in the past decades, and attempts to put things into an integrative perspective, in the hope of identifying the gaps of knowledge, provoking future work to be conducted to bring predictive models into existence.

Acknowledgment

The authors wish to acknowledge many past students and postdocs who contributed to this effort over the years, especially Dr. Frank Yin, Dr. Jun Zhang, Dr. Richard Gates, Dr. Yinglon Wong, Dr. Ken Zhang, and Dr. Gunter Bartelt who are significantly involved in this research. Financial support from the Office of Naval Research, DOE ECUT program is gratefully acknowledged.

References

[1] Greenwood J A, Williamson J B P. Contact of nominally flat surfaces. *Proc Roy Soc Series A* **295**: 300–319 (1966)

[2] Whitehouse D J, Archard J F. The properties of random surfaces of significance in their contact. *Proc Roy Soc Series* **316**: 97–121 (1970)

[3] Bowden F P. Friction. *Nature* **166**: 330–334 (1950)

[4] Wang F X, Lacey P, Gates R S, Hsu S M. A study of the relative surface conformity between two surfaces in sliding contact. *J Tribol* **113**: 755–761 (1991)

[5] Hsu S M. Fundamental mechanisms of friction and lubrication of materials. *Langmuir* **12**(19): 4482–4485 (1996)

[6] Urbakh M, Klafter J, Gourdon D, Israelachivili J. The nonlinear nature of friction. *Nature* **430**: 525–528 (2004)

[7] Luengo G, Israelachvili J, Granick S. Generalized effects in confined fluids: New friction map for boundary lubrication. *Wear* **200**: 328–335 (1996)

[8] Gao J, Luedtke W D, Gourdon D, Ruths M, Israelachvili J, Landman U. Friction forces and Amontons' law: From molecular to the macroscopic scale. *J Phys Chem B* **108**: 3410–3425 (2004)

[9] Cieplak M, Smith E D, Robbins M. Molecular origin of friction: The force on adsorbed layers. *Science* **265**: 1209–1212 (1994)

[10] Carpick R, Ogletree D, Salmeron M. A general equation for fitting contact area and friction vs load measurements. *J Colloid Interf Sci* **211**: 395–400 (1999)

[11] Carpick R. Contolling friction. *Science* **313**: 184–185 (2006)

[12] Persson B N J. Sliding friction. *Surf Sci Rep* **33**: 83–119 (1999)

[13] Tabor D. Friction-the present state of our understanding. *J Lubr Technol* **103**: 169–179 (1981)

[14] Tabor D. Future directions of research in adhesion and friction-status of understanding. In *Proceedings of NASA Lewis Research Center Tribology in the 80's*, 1984: 119–139.

[15] Bowden F P, Tabor D. *The Friction and Lubrication of Solids*. Gloucestershire (UK): Clarendon Press, 1986.

[16] Buckley D H. *Surface Effects in Adhesion, Friction, Wear, and Lubrication*, Vol. 5. Elsevier, 1981.

[17] Briscoe B J, Tabor D. Shear properties of thin polymeric films. *J Adhesion* **9**: 145–155 (1978)

[18] Wang LY, Ying F Z, Zhang J, Chen C, Hsu S M. Strength measurement of thin lubricating films. *Wear* **237**: 155–162 (2000)

[19] Ying T. Wear mechanism for ductile and brittle materials in micro-contacts. PhD thesis. Maryland (USA): University of Maryland, College Park, 1993.

[20] Drexler K E. In *Nanosystems: Molecular Machinery, Manufacturing, and Computation*. John Wiley & Sons, 1992: 23.

[21] Mate C M, McClelland G M, Erlandsson R, Chiang S. Atomic-scale friction of a tungsten tip on a graphite surface. *Phys Rev Lett* **59**: 1942–1945 (1987)

[22] Gnecco E, Bennewitz R, Gyalog T, Loppacher Ch, Bammerlin M, Meyer E, Güntherodt H J. Velocity dependence of atomic friction. *Phys Rev Lett* **84**: 1172–1175 (2000)

[23] Yoshizawa J, Chen Y-L, Israelachvili J. Fundamental mechanisms of interfacial friction. 1. Relation between adhesion and friction. *J Phys Chem* **97**: 4128–4140 (1993)

[24] Yoshizawa J, Israelachvili J. Fundamental mechanisms of interfacial friction. 2. Stick-slip friction of Spherical and Chain molecules. *J Phys Chem* **97**: 11300–11313 (1993)

[25] Liu H, Bhushan B. Adhesion and friction studies of mic-roelectromechanical systems/nanoelectromechanical systems materials using a novel microtriboapparatus. *J Vac Sci Technol A* **21**: 1528–1538 (2003)

[26] Dienwiebel M, Verhoeven G S, Pradeep N, Frenken J W M, Heimberg J A, Zandbergen H W. Superlubricity of graphite. *Phys Rev Lett* **92**: 126101 (2004)

[27] Enachescu M, van den Oetelaar R J A, Carpick R W, Ogletree D F, Flipse C F J, Salmeron M. Observation of proportionality between friction and contact area at the nanometer scale. *Trib Lett* **7**: 73–78 (1999)

[28] Richetti P, Drummond C, Israelachvili J, In M, Zana R. Inverted stick-slip friction. *Europhys Lett* **55**: 653–659 (2001)

[29] Overney R M, Takano H, Fujihira M, Paulus W, Ringsdorf H. Anisotropy in friction and molecular stick-slip motion. *Phys Rev Lett* **72**: 3546–3549 (1994)

[30] Socoliuc A, Bennewitz R, Gnecco E, Mayer E. Transition from stick-slip to continuous sliding in atomic friction: Entering a new regime of ultralow friction. *Phys Rev Lett* **92**: 134301 (2004)

[31] Niederberger S, Gracias D G, Komvopoulos K, Somorjai G A. Transition from nanoscale to microscale dynamic friction mechanisms on polystyrene and silicon surfaces. *J Appl Phys* **87**: 3143–3150 (2000)

[32] Liu E, Blanpain B, Celis J P, Roos J R. Comparative study between macro-tribology and nanotribology. *J Appl Phys* **84**: 4859–4865 (1998)

[33] Aoike T, Uehara H, Yamanobe T, Komoto T. Comparison of macro- and nano-tribological behavior with surface plastic deformation of polystyrene. *Langmuir* **17**: 2153–2159 (2001)

[34] Bhushan B, Li X. Atomic-scale and microscale friction studies of graphite and diamond using friction force microscopy. *J Mater Res* **12**: 54–63 (1997)

[35] Bhushan B. Adhesion and stiction: Mechanisms, measurement techniques, and methods for reduction. *J Vac Sci Technol B* **21**: 2262–2296 (2003)

[36] McGuiggan P, Zhang J, Hsu S M. Comparison of friction measurement using the atomic force microscope and the surface forces apparatus: The issue of scale. *Tribol Lett* **10**: 217–223 (2001)

[37] Pethica J B, Hutchings R, Oliver W C. Hardness measurement at penetration depths as small as 20 nm. *Philos Mag A* **48**: 593–606 (1983)

[38] Bennewitz R, Gnecco E, Gyalog T, Meyer E. Atomic friction studies on well-defined surfaces. *Tribol Lett* **10**: 51–56 (2001)

[39] Fang H W. Ultra-High molecular weight polyethylene wear particle effects on bioactivity. PhD thesis. Maryland (USA): University of Maryland, College Park, 2004.

[40] Mate C M. On the road to an atomic- and molecular-level understanding of friction. *MRS Bull* **27**: 967–971 (2002)

[41] Gyalog T, Bammerlin M, Luthi R, Meyer E, Thomas H. Mechanism of atomic friction. *Europhys Lett* **31**: 269 (1995)

[42] Sang Y, Dube M, Grant M. Thermal effects on atomic friction. *Phys Rev Lett* **87**: 174301 (2001)

[43] Evstigneev M, Schirmeisen A, Jansen L, Fuchs H, Reimann P. Force dependence of transition rates in atomic friction. *Phys Rev Lett* **97**: 240601 (2006)

[44] Price W, Leigh S, Hsu S, Patten T, Liu G. Measuring the size dependence of Young's modulus using force modulation atomic force microscopy. *J Phys Chem A* **110**: 1382–1388 (2006)

[45] Chen C Q, Shi Y, Zhang Y S, Zhu J, Yan Y J. Size dependence of Young's modulus in ZnO nanowires. *Phys Rev Lett* **96**: 075505 (2006)

[46] Kim J, Greer J. Size-dependent mechanical properties of molybdenum nanopillars. *Appl Phys Lett* **93**: 101916 (2008)

[47] Greer J, Oliver W C, Nix W D. Size dependence of mechanical properties of gold at the micron scale in the absence of strain gradients. *Acta Materialia* **53**: 1821–1830 (2005)

[48] Johnson K L. The contribution of micro/nano-tribology to the interpretation of dry friction. *Proc IMechE, Part C: J Mech Eng Sci* **214**: 11–22 (2000)

[49] Aoike T, Uehara H, Yamanobe T, Komoto T. Comparison of macro- and nanotribological behavior with surface plastic deformation of polystyrene. *Langmuir* **17**: 2153–2159 (2001)

[50] Luan B, Robbins M O. The breakdown of continuum models for mechanical contacts. *Nature* **435**: 929–932 (2005)

[51] Luan B, Robbins M O. Contact of single asperities with varying adhesion: Comparing continuum mechanics to atomistic simulations. *Phys Rev E* **74**: 026111 (2006)

[52] Luan B, Robbins M O. Hybrid atomistic/continuum study of contact and friction between rough solids. *Tribol Lett* **36**: 1–16 (2009)

[53] Ying T N, Hsu S M. Asperity-asperity friction as measured by a two-ball collision apparatus. *Tribol Trans* **39**: 33–42 (1996)

[54] Ying T N, Hsu S M. Effect of friction on subsurface strain distribution of steel. *Tribol Trans* **40**: 420–435 (1997)

[55] Hsu S M, Shen M C, Ying T N, Wang Y S, Lee S W. Tribology of silicon-based ceramics. *Ceram Trans* **42**: 189–205 (1994)

[56] Fuller D. *Theory and Practice of Lubrication for Engineers*. New York: Wiley, 1956.

[57] Booser E R, Ed. *CRC Handbook of Lubrication: Theory and Practice of Tribology, Volume II: Theory and Design*. CRC Press, 2010.

[58] Szeri A Z. *Fluid Film Lubrication: Theory and Design*. Cambridge: Cambridge University Press, 2005.

[59] Dowson D, Higginson G R, Whitaker A V. Elasto-hydrodynamic lubrication: A survey of isothermal solutions. *J Mech Eng Sci* **4**: 121–126 (1962)

[60] Sibley L B, Orcutt F K. Elasto-hydrodynamic lubrication of rolling-contact surfaces. *ASLE Trans* **4**: 234–249 (1961)

[61] Hamrock B J, Dowson D. Isothermal elastohydrodynamic lubrication of point contacts: Part 1—Theoretical formulation. *J Lubr Technol* **98**: 223–228 (1976)

[62] Beerbower A. Boundary lubrication-scientific and technical forecast report. US Army Report AD747336, 1972.

[63] Beerbower A. A critical survey of mathematical models for boundary lubrication. *STLE Trans* **14**: 90–104 (1971)

[64] Hsu S M, Gates R S. Effect of materials on tribochemical reactions between hydrocarbon and surfaces. *J Phys D: Appl Phys* **39**: 3128–3137 (2006)

[65] Hsu S M, Gates R S. Boundary lubrication and boundary lubricating films. In *CRC Handbook of Modern Tribology*, Bhushan B Ed. New York: CRC Press LLC, 2001: 455–492.

[66] Sambasivan S, Hsieh S, Fischer D, Hsu S M. Effect of self-assembled monolayer film order on nanofriction. *J Vac Sci Technol A* **24**: 1484–1488 (2006)

[67] Liu Y, Evans D F, Song Q, Grainger D W. Structure and frictional properties of self-assembled surfactant monolayers. *Langmuir* **12**: 1235–1244 (1996)

[68] Clark D B, Klaus E E, Hsu S M. The role of iron and copper in the oxidative degradation of lubricants. *Lubr Eng* **41**: 280–287 (1985)

[69] Hsu S M, Klaus E E, Cheng H S. A mechano-chemical descriptive model for wear under mixed lubrication conditions. *Wear* 128: 307–323 (1988)

[70] Hélouvry A, Brian P D, De Wit C C. A survey of models, analysis tools and compensation methods for the control of machines with friction. *Automatica* **30**: 1083–1138 (1994)

[71] Persson B N J. Theory of friction and boundary lubrication. *Phys Rev B* **48**: 18140–18158 (1993)

Stephen HSU. He obtained his PhD degree at Pa State University under Prof. EE Klaus on Tribology. He worked four years at Amoco Chemicals then joined NIST in 1978. In 2007, he joined City University of Hong Kong as a Chaired Professor and Head of Manufacturing Engineering Department. Since 2009, he is Professor of Engineering and Applied Science at George Washington University.

His current research interests are energy efficient buildings, wind energy, energy efficiency, fuel economy, surface textures, self-repair, and advanced lubricant development.

Fei ZHAO. He received his PhD degree in Physical Chemistry from Lanzhou Institute of Chemical Physics, CAS in 2010. Then he joined the Henan Key Laboratory of Materials Tribology at Henan University of Science and Technology. He has been working with Prof. Hsu as a post doctor in the George Washington University since 2012. His research interests include superlubricity of DLC coatings by PECVD/PVD, current-carrying tribology and fuel economy improvement by surface texturing.