Synchronous measurement of tribocharge and force at the footpads of freely moving animals

Black phosphorus as a new lubricant
## Review Article

**Stress-augmented thermal activation: Tribology feels the force** / 1–31
Hugh SPIKES

## Research Articles

**Anti-loosening performance of coatings on fasteners subjected to dynamic shear load** / 32–46
Junbo ZHOU, Jianhua LIU, Huajiang OUYANG, Zhenbing CAI, Jinfang PENG, Minhao ZHU

**Chaotic characteristics and attractor evolution of friction noise during friction process** / 47–61
Cong DING, Hua ZHU, Guodong SUN, Yuankai ZHOU, Xue ZUO

**Thermocapillary migration mechanism of molten silicon droplets on horizontal solid surfaces** / 62–74
Tao SUN, Cunhua JIANG, Jianning DING, Ningyi YUAN

**Synchronous measurement of tribocharge and force at the footpads of freely moving animals** / 75–83
Yi SONG, Zhouyi WANG, Jun ZHOU, Yang LI, Zhendong DAI

**Chemical effects on the sliding friction of Ag and Au(111)** / 84–97
H. E. KO, S. G. KWAN, H. W. PARK, A. CARON

**Transfer of electrons on scratched iron surfaces: Photoelectron emission and X-ray photoelectron spectroscopy studies** / 98–115
Yoshihiro MOMOSE, Daisuke SUZUKI, Keika TSURUYA, Takao SAKURAI, Keiji NAKAYAMA

**Black phosphorus as a new lubricant** / 116–142
Wei WANG, Guoxin XIE, Jianbin LUO
Stress-augmented thermal activation: Tribology feels the force

Hugh SPIKES
Tribology Group, Department of Mechanical Engineering, Imperial College, London SW72BX, UK
Received: 22 September 2017 / Accepted: 24 November 2017
© The author(s) 2018. This article is published with open access at Springerlink.com

Abstract: In stress-augmented thermal activation, the activation energy barrier that controls the rate of atomic and molecular processes is reduced by the application of stress, with the result that the rate of these processes increases exponentially with applied stress. This concept has particular relevance to Tribology, and since its development in the early twentieth century, it has been applied to develop important models of plastic flow, sliding friction, rheology, wear, and tribochemistry. This paper reviews the development of stress-augmented thermal activation and its application to all of these areas of Tribology. The strengths and limitations of the approach are then discussed and future directions considered. From the scientific point of view, the concept of stress-augmented thermal activation is important since it enables the development of models that describe macroscale tribological performance, such as friction coefficient or tribofilm formation, in terms of the structure and behaviour of individual atoms and molecules. This both helps us understand these processes at a fundamental level and also provides tools for the informed design of lubricants and surfaces.

Keywords: stress activation; stress augmented thermal activation; mechanochemistry; friction; EHD friction; wear; Eyring

1 Introduction

Tribology is concerned with the interaction of surfaces in relative motion and the behaviour of interfaces present between these surfaces. Such relative motion is produced by applied shear forces and is opposed by friction.

Despite more than a hundred years of research we still have surprisingly little deep understanding of how the applied forces, the friction, and the relative motion are connected at an atomic and molecular scale. Most tribological experiments are carried out at the macroscale, where, while they provide useful relationships between friction, load, temperature, and sliding velocity, they do not reveal whence these originate. Computer-based, molecular-scale simulations are starting to provide some insights and are certainly the way forward but they have yet to be linked fully to the macroscale laws of friction and sliding derived from experiment.

There is, however, one model or group of models that is able to connect directly molecular-scale behaviour to macroscale friction response and other tribological phenomena. This model was developed independently in quite different contexts by Prandtl between 1913 and 1928 [1, 2] and Eyring in 1936 [3]. It is based on the concept that a mechanical force applied to an atom or molecule couples with the thermal energy possessed by this entity to increase the rate of a process. This process may be sliding, fluid flow or chemical reaction, or indeed any process that involves an atom or molecule traversing an energy barrier. Because it involves a combination of mechanical force and thermal energy this concept is sometimes called stress-promoted or stress-augmented thermal activation but no standard descriptor has yet become widely accepted.

The essential idea starts with the Arrhenius rate equation in which the rate constant $k_o$ is given by:

$$k_o = Ae^{-E_k/k_BT}$$ (1)
where \( A \) is a pre-constant, \( E_a \) is the thermal activation energy for the process, \( k_\text{B} \) is Boltzmann’s constant, and \( T \) is the absolute temperature. This equation can be quite easily understood in terms of probability since according to the Boltzmann energy distribution, \( e^{-E_a/k_\text{B}T} \) is the fraction of atoms or molecules in a large ensemble that have energy greater than \( E_a \) and thus enough to traverse this energy barrier. It can however also be derived from thermodynamic or statistical mechanics principles with \( E_a \) replaced by a free energy [4].

According to stress-augmented thermal activation, when a force, \( f \), is applied to the atom or molecule along the direction of the process, Eq. (1) is modified to

\[
k_f = A e^{\frac{(E_a - f \Delta x)}{k_\text{B}T}}
\]

(2)

where \( \Delta x \) is the distance the force must move the atom or molecule to reach the top of the energy barrier (the activation length). The term \( f \Delta x \) is a work term and has the effect of reducing the energy barrier and so increasing the rate of the process. The impact of this term on the rate can easily be seen from the rearrangement

\[
k_f = A e^{\frac{E_a}{k_\text{B}T}} e^{\frac{f \Delta x}{k_\text{B}T}}
\]

(3)

where it is evident that the rate of the process increases exponentially with the applied force. It should be noted that Eq. (2) is the simplest model of stress-augmented thermal activation and this has been extended and embellished to a considerable degree over the years, as will be discussed later in this paper. This equation and those that follow can, of course, be scaled up from molecular to molar quantities so that

\[
k_f = A e^{\frac{E}{RT}} e^{\frac{f N \Delta x}{RT}}
\]

(4)

where \( E \) is now the molar thermal activation energy, \( N \) is Avogadro’s number, and \( R \) the gas constant.

In Eqs. (2) and (3), \( k_f \) is the forward rate constant of the process, i.e., in the direction of the applied force. The applied force also reduces the backward rate of reaction according to

\[
k_b = A e^{\frac{(E_a + f \Delta x)}{k_\text{B}T}}
\]

(5)

so that in systems where both forward and reverse processes are possible and have the same prefactor, the overall net reaction rate constant becomes

\[
k_{\text{net}} = A \left( e^{\frac{(E_a - f \Delta x)}{k_\text{B}T}} - e^{\frac{(E_a + f \Delta x)}{k_\text{B}T}} \right)
\]

(6)

Upon rearrangement this is

\[
k_{\text{net}} = A e^{\frac{-E_a}{k_\text{B}T}} \left( e^{\frac{f \Delta x}{k_\text{B}T}} - e^{-\frac{f \Delta x}{k_\text{B}T}} \right) = 2k_\text{B} \sinh \left( \frac{f \Delta x}{k_\text{B}T} \right)
\]

(7)

At high applied forces or when no reverse process is possible the second term is negligible, in which case the rate of a process will vary exponentially with the applied force.

The influence of the applied force on the energy barrier is shown schematically in Fig. 1 for a very simple case where, prior to the application of force, the height of the energy barrier is the same in forward and reverse directions.

The above model has been applied to many disciplines including various aspects of tribology, without, however, ever becoming central to the latter. In the author’s view it is surprising since the concept appears uniquely relevant to tribology, which is concerned the kinetic response of systems to large applied mechanical forces.

The aim of this paper is therefore to show how this stress-augmented thermal activation principle applies to Tribology, to raise its general profile as a key descriptor of many tribological processes and so help understand these processes. The paper first provides a brief history of how the model was developed and applied in various forms to different areas of tribology up to the 1950s. This is then followed by a more detailed review of each of the most significant areas of application, plastic deformation, lubricant rheology,
sliding friction, tribochemistry, and wear. Finally, there is a general discussion that attempts to bring together the various ways that stress-augmented thermal activation has been applied across Tribology and to look forward to future developments.

2 Early history

The concept of stress-augmented thermal activation appears to have been first developed by Ludwig Prandtl at the University of Göttingen in the early 1900s, being outlined by his student von Karman in 1913 [1] and described in full by Prandtl himself in 1928 [2]. Prandtl's interest was in crystal plasticity and, as described in the next section, he derived an expression for the rate of plastic flow of a metal in terms of the effect of applied stress on the forward and reverse motion of atoms across the potential field of an adjacent atomic plane. This gave an expression quite similar to Eq. (6) above. Although Prandtl's focus was on metal deformation he noted at the end of his 1928 paper:

“To conclude, it may now be suggested that our conceptual model is also suitable for the treatment of kinetic friction.”

Prandtl’s model and a related one by Becker [6] was applied in a series of subsequent studies of metal deformation to describe dislocation motion, creep, and plastic flow, as described later in this paper. However, it did not appear to reach the attention of researchers in the fields of friction, lubrication, and wear until the early 2000s.

In 1936 Henry Eyring at Princeton developed a model of liquid viscosity using a similar concept [3]. Since his ideas originated from his work on chemical kinetics he described his model as being based on absolute reaction rates. For simple molecular liquids he envisaged liquid flow as originating from the movement of individual molecules over the potential field of an adjacent plane of molecules. In the absence of an applied shear stress there will be no net flow since the forward and reverse rate constants of flow will be the same. However, from Eq. (7) an applied shear stress will impose a force on the molecules so as to produce the difference in the velocity between adjacent layers of the form

$$\Delta \mu = 2k_o \lambda \sin \left( \frac{f \lambda}{2k_o T} \right)$$

(8)

where $\lambda/2$ is the distance the molecules move from the bottom to the top of the energy barrier and is thus equivalent to the activation length $\Delta x$ in Eqs. (2)−(7) and Fig. 1. As described later in this review, Eyring converted this velocity difference to a shear rate and the force to a shear stress and thus derived a rheological equation relating shear rate and shear stress. At moderate shear stress the shear rate varied with the sinh() of the shear stress, but at high shear stress this simplified to an exponential dependence.

Although Eyring developed his equation in terms of individual molecular motion, he was well aware of its general thermodynamic origins. Thus he wrote in 1941 [7]:

“The general equation giving the rate of any reaction which has been modified by some external agency may be written:

$$k' = \left( \frac{k_o T}{h} \right) e^{-\Delta F^* / RT}$$

(9)

Here $\Delta F^*$ is the contribution made by the external agency to the free energy of activation.”

Eyring’s external agency was of course applied force. Eyring and his students went on to apply this model to the prediction of the viscosities of simple liquids [5, 8, 9], polymer solutions [10, 11], and unimolecular films [12], to the rate of deformation of metals [13], and to the flow and creep of polymers [14, 15]. In the latter he treated flow as a balance between the breaking and reforming of molecular bonds and noted the potential applicability of this to tribology, writing [14]:

“The phenomena of friction and lubrication appear to involve this type of equilibrium between rate of breaking and re-formation of bonds.”

There is no evidence that Eyring was aware of Prandtl’s earlier work and indeed the two models originate from very different starting points, Prandtl’s based on mechanical principles and Eyring’s on chemical kinetics. A recent paper has compared the two approaches and the similarities and differences between them in some detail, as well as briefly outlining the general application of stress-augmented thermal activation to Tribology [16].
In 1953, Schallamach studied the kinetic friction of rubber sliding against ground glass and silicon carbide paper and found that the logarithm of the sliding velocity, \( v \), varied linearly with the applied force, \( F \), and inversely with the absolute temperature [17, 18]. Based on Eyring’s approach and assuming negligible reverse process he proposed an equation relating sliding velocity, \( v \), to force, \( F \), and temperature, \( T \), very similar to Eq. (3)

\[
\frac{v}{A e^{-E/F}} \approx \frac{T}{\gamma T}
\]

where \( E \) is an activation energy and \( A \) and \( \gamma \) are constants, the latter being the activation length. Clearly this equation predicts that \( F \) will be proportional to \( \log(v) \) at constant temperature.

In the 1950s, Beuche developed a model to describe the mechanical scission of polymers under shear. His model was initially based on a combined effect of thermal and stored elastic energy in breaking polymer bonds [19], but this was soon modified to one very similar to Eyring’s, with probability of bond breaking, \( p \), given by

\[
p = e^{-(E-F\delta)/\gamma T}
\]

where \( E \) is the tension in the bond and \( 2\delta \) the distance the bond will stretch before breaking [20, 21].

Also in the 1950s, Zhurkov investigated lifetime to fracture in uniaxial tension of a wide range of solid materials including metals, ceramics, and polymers [22–24] and found a dependence on stress and temperature of the form

\[
\tau = \tau_0 e^{(U_0 - \gamma\sigma)/\gamma T}
\]

where \( \tau \) is the lifetime, \( U_0 \) an activation energy for bond dissociation, \( \sigma \) the applied tensile stress, and \( \tau_0 \) and \( \gamma \) are constants. The similarity of this to Eyring’s model is evident, with the exponent in Eq. (12) being positive simply because \( \tau \) is a lifetime, and thus reciprocal to a rate. Another difference from Eqs. (8), (10), and (11) is that the work term in Eq. (12), \( \gamma \sigma \), is based on applied shear stress rather than applied force, so that the activation constant \( \gamma \) has units of volume rather than length; this difference will be discussed later in this paper. It is not clear whether Beuche or Zhurkov were aware in the 1950s of Eyring’s work, though in later papers Zhurkov does reference Eyring’s work on polymer flow [24, 25].

From the above it can be seen that, by the end of the 1950s, the stress-augmented thermal activation approach pioneered by Prandtl and Eyring had been successfully used in models of plastic deformation, liquid viscosity, kinetic friction, and chemical bond dissociation.

In subsequent years these models have been further applied and developed and, as described below, have helped us understand the atomic and molecular origins of key tribological processes. The next five sections of this paper will illustrate this by reviewing, in turn, the application of stress-augmented thermal activation models to five areas of relevance to tribology: (i) plastic deformation of materials, (ii) lubricant rheology, (iii) sliding friction, (iv) tribochemistry, and (v) wear.

### 3 Plastic deformation

#### 3.1 Prandtl’s model

Prandtl’s concept of stress-augmented thermal activation was first described fully in 1928 [2]. Prandtl considered the forces involved when a single, elastically constrained metal atom slides across the sinusoidal force field produced by a stationary plane of adjacent atoms, as shown schematically in Fig. 2(a). A plot of the force experienced by this atom as it traverses the stationary surface is shown in Fig. 2(b). The rest position of the atom is at C but its actual position is at D where the (negative) linear, elastic, constraining force shown as straight line, intercepts and balances the sinusoidal force field at force E. As the atom moves to the right it tracks the sinusoidal force field until, so long as it is at absolute zero temperature, it reaches position \( D' \) where the force is \( E' \). Here it becomes unstable and jumps to the alternative intercept position \( D'' \), corresponding to force \( E'' \). If the motion is now reversed the atom does not jump back to \( D' \) but instead initially tracks the sinusoidal force field and Prandtl explained the hysteresis of plastic deformation process on the basis of this irreversible process. This concept is now known as the Prandtl-Tomlinson model [16].

However, in practice the atom will be at a temperature considerable above absolute zero and so will possess thermal energy. This means that it will...
be able to jump sooner than from D, for example from P, when the force is Q, to P'. To do this it must possess thermal energy $U_1$, which corresponds to the horizontally hatched area in Fig. 2(b). The reverse process is much less likely since the atom needs energy $U_2$, corresponding to the vertically-hatched area, in order to jump from P' to P.

Figure 2(b) shows the variation of force with $x$, the coordinate of the stationary surface, while Fig. 2(c) shows this force as a function of the displacement of the frame (or plane of atoms) in which the single atom is elastically constrained, $\xi$. If it possesses the requisite thermal energy, the atom jumps from force Q to Q' spontaneously at a fixed position of the sliding frame. Indeed, depending on the thermal energy available this jump can take place anywhere between $-a$ and $a$.

Prandtl then considered a very large number of such atoms, irregularly distributed in the sliding frame and non-commensurate with the adjacent force field. The proportion of atoms at a particular location $\xi$ between $-a$ and $a$ in the upper part of the curve in Fig. 2(c) (and thus available to jump down) is $\mu$. The fraction of these atoms having energy $U_i$ sufficient to jump is given by the Arrhenius expression $\exp(-U_i/k_BT)$.

The corresponding proportion of atoms in the lower part of the curve is $1-\mu$ and the fraction with enough energy to jump upward is $\exp(-U_2/k_BT)$. Thus Prandtl derived the net rate at which atoms would move to the upper part of the curve as

$$\frac{d\mu}{dt} = \frac{1}{\tau} \left( -\mu e^{-\frac{U_1}{k_B T}} + (1-\mu)e^{-\frac{U_2}{k_B T}} \right)$$

or, at constant deformation rate, since $d\xi/dt = dx/dt$

$$\frac{d\mu}{dx} = \frac{1}{\tau c} \left( -\mu e^{-\frac{U_1}{k_B T}} + (1-\mu)e^{-\frac{U_2}{k_B T}} \right)$$

where $\tau$ is the atomic oscillation time and $c$ is the speed of deformation, $d\xi/dt$.

The total applied force, $P$, required to produce this deformation rate is then the integral of the local force multiplied by the jump probability over the range $-a$ to $a$

$$P = \int_{-a}^{a} (\mu f_1(x) + (1-\mu)f_2(x))dx$$

where $f_1$ and $f_2$ correspond to the levels of force between which the atoms jump, e.g., Q and Q' at the jump location shown in Fig. 2(c).

Solution of Eq. (14) requires an expression for $U_i$ in terms of the displacement of the atom and for much of his analysis Prandtl assumed a simplest possible, linear dependence, i.e., $U_i/k_BT = A - Bx$ and $U_2/k_BT = A + Bx$. Even with this approximation, Eq. (14) could only be solved analytically if the second, reverse terms in Eq. (14) were assumed negligible, in which case Prandtl showed that $\mu = \exp\left(-e^{-Bx}/(Bc/e^A)\right)$

and that integration of $P = \int_{-a}^{a} \mu f_1(x)dx$ gives applied force, $P$ to be

$$P = \frac{C}{B} \left( \log(c) + A + \log(Br) - 0.5772 \right)$$

where $C = f_1 - f_2$ and was taken to be constant. This equation thus predicts that the applied force $P$ varies with the logarithm of the plastic flow rate $c$ in accord with experimental findings. Prandtl also explored a more realistic parabolic dependence of the jump energies $U_1$ and $U_2$ on displacement and discussed the dependence of plastic flow rate on temperature. Although focussed on plastic deformation of metals he showed that his model was able to predict the effective viscosity of the material under deformation.
from an essentially elastic solid at low temperature right through to viscous-like behaviour at high temperature.

Prandtl’s derivation has been described in some detail above since it has been used in recent work on modelling single contact friction, described later in Section 5. However, it is clearly complex, involving a two stage integration to find the local jump probability and then to combine these at all locations to determine the total force. As such it is analytically soluble in a very limited number of cases and this may be one reason why Prandtl’s model received little contemporary attention. Instead a much simpler thermal-activation model of plastic deformation developed by Richard Becker in 1925 was preferred.

3.2 Becker’s model

Becker considered the effect of combined applied shear stress and temperature on the glide of one plane of atoms over another in a metallic crystal [6]. He defined the displacement fracture stress, \( \sigma_0 \), as the stress required to produce glide in the absence of any thermal motion and the actual applied stress, \( \sigma \), to be smaller than this. He then suggested that for glide to occur, thermal motion would be required within an unknown volume \( V \) to bridge the difference between \( \sigma \) and \( \sigma_0 \) and that the mechanical energy required to achieve this would be \( V(\sigma_0 - \sigma)^2 / 2E \) where \( E \) is the Young’s modulus of the material (in later work this became the elastic shear modulus \( G \)). The probability of this condition being reached is then given by

\[
p = ke^{\frac{V(\sigma_0 - \sigma)^2}{2kT}}
\]

The plastic flow velocity was assumed to be proportional to this probability. Becker tested this equation using data from the deformation of tungsten single crystal monofilaments and thus determined \( \sigma_0 \) and \( V \). Of particular interest was the estimated value of \( V \), which he found corresponded neither to the volume of a single atom nor to a whole plane but was instead about 3000 tungsten atoms. Becker thus deduced that slip originated from the concerted response to stress of a cluster of atoms within the crystal. In the 1930s, Andrade and Roscoe [26] and Becker’s student Orowan [27, 28] combined this model with crystallographic evidence to propose that plastic deformation and creep occur due to the motion of dislocations. Becker’s model was also applied in 1941 to the creep of polymers [29].

3.3 Kauzmann

In 1941 Kauzmann applied Eyring’s model as expressed by Eq. (7) to the creep of metals [13]. This predicts that at high shear stress the underlying rate process should vary exponentially with the applied stress, rather than with the stress squared as suggested by Becker’s Eq. (17). Kauzmann noted the general similarity of his and Becker’s approaches however and, like Becker, found that when tested against creep data his model indicated an activation volume corresponding to several hundred atoms. He suggested that this represented the number of atoms involved in dislocation movement. Figure 3 shows Kauzmann’s view of how applied stress influences the potential energy barrier for dislocation movement. It is elongated because of the multi-atom activation volume.

Kauzmann also noted that at very low stresses in the region of micro-creep, the activation volume was much smaller and close to an atomic volume, suggesting that the glide process had changed from motion of dislocations to that of single atoms.

Both Becker’s and Kauzmann’s models continued to be applied to describe the dependence on stress and temperature of plastic deformation, creep, and dislocation motion throughout the 1940s to 60s, with Kauzmann’s model slowly displacing Becker’s [30–36] (Ref. [31] is one of very few to cite Prandtl). In 1952 Cottrell [37] discussed three existing stress-augmented thermal activation models, Kauzmann’s that suggested that creep varies exponentially with \( \sigma \), Becker’s that indicated an exponential dependence on \( \sigma^2 \) and also

![Fig. 3 Influence of applied stress on the potential energy curve of a moving dislocation. Adapted from [13].](https://mc03.manuscriptcentral.com/friction)
a model by Mott and Nabarro that proposed an exponential $\sigma^{3/2}$ dependence. A detailed review of the state of the subject in the mid-1970s was provided by Kocks et al. [38].

3.4 Since the 1970s

The stress-augmented thermal activation concept has continued to be applied to describe the dependence of plastic flow stress on temperature and strain rate and thus the relationship between applied stress and dislocation velocity to the present day [39]. Its application has even been extended to describe the dependence of crack propagation rate on temperature and stress [40]. Models have become more complex to incorporate dislocation pinning effects and the behaviour of different crystal structures. Most recent models have replaced Becker’s original volume $V$ by the product of an activation area and the Burger’s vector (an activation length) and have allowed this activation area to vary with the applied stress [41].

Although most interest has been in the plastic deformation of metals, both Becker’s [29] and Eyring’s model [42, 43] have also been applied to the deformation of solid polymers. In 1971 Ward analysed tensile and compressive yield stress data of several polymers as a function of temperature and strain rate in terms of an extended Eyring model [43]

$$\dot{\varepsilon} = Ae^{(\Delta U - \tau \cdot \gamma \cdot d \cdot p) / b T}$$

(18)

where $\dot{\varepsilon}$ is the strain rate, $\Delta U$ an activation energy for the flow process, $\tau$ the shear yield stress, $v$ the shear activation volume, $p$ the hydrostatic pressure, and $\Omega$ the pressure activation volume. Ward thus introduced a pressure work term in the rate equation which he assumed was positive, i.e., the rate decreased with increasing pressure. This dependence on pressure had been previously used by Eyring et al. to describe the influence of pressure on viscosity [44] and, as will be described in Section 5 of this paper, was later adopted by Tabor to describe the influence of load on sliding friction. In many ways a pressure term is more familiar to thermodynamicists than a shear stress term, since a PV work term is embedded in rate and equilibrium equations involving gases.

4 Lubricant viscosity and rheology

4.1 Eyring’s model

In 1936 Eyring developed his well-known model of liquid viscosity. This originated from his transition state theory of chemical kinetics that considers a chemical reaction as involving passage of a molecular system over an energy barrier in the process of converting reactants to products. Eyring thus considered liquid flow as resulting from passage of individual molecules over an energy barrier and, as the title of his paper indicates, regarded viscosity as an example of a reaction rate. Figure 4 shows his schematic model in which a molecule moves across an adjacent plane of molecules into a vacancy or “hole” under the application of a shear force $f$[14]. To do this it must traverse an activation energy barrier $E_a$ as shown earlier in Fig. 1.

The applied force on the molecule has the effect of lowering this activation energy by a factor $f \lambda / 2$ in the direction of the force and raising it by this amount in the backward direction. This causes the molecule (and the other molecules in its plane) to move relative to the adjacent plane at velocity given by Eq. (8), i.e.,

$$\Delta U = 2k \lambda \sin \left( f \lambda / 2k T \right).$$

The velocity gradient or shear rate is thus

$$\dot{\gamma} = \Delta U / \lambda_1 = \frac{2k \lambda}{\lambda_1} \sin \left( f \lambda / 2k T \right)$$

(19)

where $k_0$ is the rate constant in the absence of an applied force, $\lambda_1$ is the spacing between adjacent molecular layers and $\lambda$ is the total distance the molecule moves in one jump, i.e., $2 \Delta \chi$ (the distance between the two energy minima in Fig. 1).

The force on a molecule can be related to the applied shear stress by

$$\tau = f / \lambda_2 \lambda_3\lambda_4,$$

where the product

Fig. 4  Eyring’s schematic model of viscous flow of simple molecules. Reproduced with permission from [14]. Copyright ACS, 1940.
$\lambda_2\lambda_3$ is the cross sectional area of the molecule in the plane of shear as shown in Fig. 4.

Replacing $f$ in Eq. (19) by the shear stress $\tau$ yields an expression for the effective viscosity

$$\eta_{eff} = \frac{\tau_{\lambda_1}}{2k_B\lambda_1 \sin h \left( \frac{\tau_{\lambda_2}\lambda_2\lambda_3}{2k_B T} \right)} \quad (20)$$

At low applied shear stress, when $\tau << k_B T/\lambda_2\lambda_3$, $\sin h(\tau_{\lambda_1}\lambda_2\lambda_2/2k_B T) = \tau_{\lambda_1}\lambda_2\lambda_2/2k_B T$, so the Newtonian viscosity is given by

$$\eta_N = \frac{k_B T \lambda_1}{\lambda_2\lambda_3\lambda_1 \lambda_3^2} \quad (21)$$

Eyring used this expression in two ways. Firstly, he evaluated the rate constant $k_o$ by relating the energy of activation $E_a$ to the enthalpy of vaporisation and including a free volume term [8]. Eyring found that for liquids with approximately spherical molecules Eq. (21) predicted the measured Newtonian viscosity quite well, based on $\lambda^2\lambda_2\lambda_3/\lambda_1$ being the molecular volume. For polymeric molecules however, the predicted viscosity was too low and Eyring suggested that this was because in the latter case the flow unit was only a segment of the molecule [14].

Eyring also used Eq. (21) to substitute $k_o$ in Eq. (20) and so obtained a molecular-based model of shear thinning behaviour of liquids

$$\eta_{eff} = \frac{\tau_{\lambda_2}\lambda_2}{2k_B T \sin h \left( \frac{\tau_{\lambda_2}\lambda_2\lambda_3}{2k_B T} \right)} \quad (22)$$

The term $2k_B T / \lambda_2\lambda_3\lambda_1$ has units of stress and is nowadays called the Eyring stress, $\tau_e$. We thus have

$$\eta_{eff} = \frac{\tau}{\eta_N \sin h \left( \frac{\tau}{\tau_e} \right)} \quad (23)$$

Or, since $\eta_{eff} = \tau / \dot{\gamma}$

$$\tau = \tau_e \sin h^{-1} \left( \frac{\eta_N \dot{\gamma}}{\tau_e} \right) \quad (24)$$

This remarkably simple expression was the first theoretically-derived rheological model of liquid shear thinning. It originates directly from the stress-augmented thermal activation concept and its relationship to this can be most clearly seen by the rearrangement of Eq. (24) to

$$\frac{\eta_N \dot{\gamma}}{\tau_e} = \sin h \left( \frac{\tau}{\tau_e} \right) = \frac{e^{\tau_e / \tau} - e^{-\tau_e / \tau}}{2} \quad (25)$$

This has obvious analogies to Eq. (7), showing how the net flow rate is a combination of flow in the forward and backward directions and how these are separately influenced by the applied force and thus shear stress. It also shows that this $\sin h$ relationship possesses the requisite symmetry properties for a rheology equation in that, when the sign of the stress is changed, this results in the shear rate changing sign but not magnitude.

Equation (24) predicts that simple molecular liquids such as most base oils will begin to show significant shear thinning when the value of $\eta_N \dot{\gamma}/\tau_e$ approaches unity. However we now know that for such liquids, $\tau_e$ is typically 5 to 50 MPa and it was not possible in Eyring’s day to reach such values of the product of $\eta_N$ and $\dot{\gamma}$ without excessive shear heating. Instead Eyring’s shear thinning model was applied, arguably less appropriately, to more complex fluids such as polymer melts and solutions and colloidal dispersions that shear-thinned much more readily [45]. Eyring’s original model assumes one type of flow unit and in 1955 Ree and Eyring [46] extended the model to accommodate blends of molecules with very different size flow units so that

$$\tau = \sum_{i=1}^{n} \frac{x_i}{\alpha_i} \sin h^{-1} \left( \beta_i \dot{\gamma} \right) \quad (26)$$

where $n$ is the number of different flow units present, $x_i$ is the fractional area of a plane parallel to the shear stress occupied by the $i$-th flow unit and $\alpha_i$ and $\beta_i$ are defined as $\alpha_i = (\lambda_i\lambda_2\lambda_3)/2k_B T$ and $\beta_i = 1/(2k_B / \lambda_1)$. $\alpha$ is effectively the reciprocal of the Eyring stress for that component. Unfortunately, the complexity of this approach detracts from the simple elegance of the original model and its application has been limited. It has been used primarily to describe the flow of polymer solutions [47] or colloidal dispersions [48] with just two
components although one three component system has been examined [49].

4.2 EHD friction

In the 1960s there was growing interest in the friction of high pressure elastohydrodynamic (EHD) lubricated contacts, as found in gears, cams, and rolling bearings. This friction was much lower than calculated by assuming that the lubricant behaved in a Newtonian fashion, indicating that the combination of the high strain rate and very high pressure (and thus very high viscosity) present in such contacts was resulting in shear thinning, even of quite low molecular weight lubricants. Initially this was ascribed to the existence of a limiting shear stress for the liquid [50, 51] although Plint [51] also noted a dependence of effective viscosity on log(shear rate) in low slip conditions. Then, in the 1970s, Hirst and Moore [52–55] and Johnson and Tevaarwerk [56, 57] showed that disc machine friction measurements could be very well explained by Eyring’s shear thinning model as expressed in Eq. (24). In such measurements, and indeed all measurements at a combination of high viscosity and high shear rate, it is important to eliminate thermal effects and this was done initially by studying very low sliding speed conditions and later by applying an isothermal correction [58, 59].

Figure 5 shows plots of isothermal friction coefficient versus shear rate from Evans and Johnson for a series of disc machine tests on a mineral oil at varies applied loads and thus mean contact pressures [59]. The mean shear stress is simply the product of the friction coefficient and the mean pressure so these plots are equivalent to shear stress versus shear rate.

The results were fitted to

$$\mu = \bar{p} \tau_e \sin h^{-1} \left( \frac{\eta_N \dot{\gamma}}{\tau_e} - \frac{\eta_N}{\tau_e} \frac{d\tau}{dt} \right)$$

(27)

where $\mu$ is the friction coefficient, $\bar{p}$ the mean pressure, and $G$ the combined shear modulus of the surfaces and lubricant. The first term in the brackets is clearly the Eyring viscous response, while the second term represents the reduction in friction due to elastic compliance of the surfaces and the lubricant film in response to the rapidly increasing shear stress in the inlet half of the contact. This elastic term becomes negligible at high shear rate.

At low pressures and high shear rates the curves follow closely Eyring’s $\sinh^{-1}$ relationship between shear stress and shear rate. At high pressures and shear rates these curves appear as straight lines, indicating that shear stress is proportional to log(shear rate). This is fully consistent with Eq. (24) since, when $\tau > \tau_e$, the equation approaches $\tau = \tau_e \log_e \left(2\eta_N \dot{\gamma} / \tau_e\right)$.

Many other experimental studies have confirmed the above findings and the Eyring shear thinning equation has become widely used for modelling the behaviour of liquid lubricants in EHD conditions [60–66]. It has also been extended and applied to glass-forming liquids [67]. It is not without controversy and some researchers have advocated alternative shear thinning equations that were originally developed to study polymer shear thinning [45]. However, polymer shear thinning occurs at relatively low shear stress, while the phenomenon of simple, low molecular weight liquids exhibiting extensive shear thinning is practically unique to tribology. This is because only in EHD lubricated contacts are liquids routinely subjected to the extreme pressure and shear rate conditions required to produce shear stresses significantly in excess of the Eyring stress.

A recent development is the application of non-equilibrium molecular dynamics simulation (NEMD)
to model the molecular behaviour of large ensembles of liquid molecules under shear. Such simulations confirm the logarithmic dependence of shear stress on shear rate predicted by Eyring at high shear stress for most of the liquids studied and also, in some cases, the whole $\sinh^{-1}$ transition from linear Newtonian dependence at low shear rate to logarithmic dependence at high shear rate [68–70]. Figure 6 shows the predicted stress-strain rate behaviour of squalane, which matches the Eyring model over an extremely wide shear rate range [70]. This is not definite confirmation of the validity of the Eyring model however since other models show equally good fit within the limits of accuracy of the simulations [71, 72]. Interestingly some liquids do not follow this type of Eyring dependence, instead showing a levelling out of shear stress with increasing strain rate to reach a limiting shear stress [69].

Finally, it should be noted that, while all of the above has focussed on bulk rheology, Eyring’s model has also been applied to describe the shear rheology of very thin adsorbed films [12, 73]. This is important in terms of the impact of surfactant and polymers on the stability of foam and emulsions but may also have relevance to the rheology of nanoscale, confined films.

5 Sliding friction

5.1 Polymer friction

Although both Prandtl and Eyring noted the applicability of their stress-augmented thermal activated models to kinetic friction, it was only in the 1950s that Schallamach [17] and Bartenev [74] employed it in this context, to explain the speed, load, and temperature dependence of the adhesive sliding friction of polymers and elastomers. Initially both researchers envisaged polymer friction as resulting from the interlocking of the molecular force fields between rubber and track at the real area of contact, a concept very similar to Eyring’s viscous flow. However subsequent work indicated that at very high sliding speeds friction started to decrease, as shown in Fig. 7 [75]. To explain this, Schallamach’s interpretation shifted towards sliding involving the breaking and reforming of bonds, with bond-forming being a thermally activated process but bond-breaking involving stress-augmented thermal activation [76]. He also

Fig. 6 NEMD simulations of squalane showing the dependence of predicted shear stress and viscosity on strain rate and temperature (dashed lines are Eyring fits). Reproduced with permission from [70]. Copyright NAS, 2017.

Fig. 7 Friction versus log(sliding speed) for a vulcanised styrene-butadiene copolymer rubbed against polished steel at 40 °C and 0.65 kg normal load. Reproduced with permission from [75]. Copyright Elsevier, 1965.
considered the bonds to be elastic so that the forces they experienced were proportional to how far they were stretched. The effect of these two assumptions was to introduce a maximum in the friction versus speed curve since friction depended on factors that both increased with sliding speed (bond stretching and thus force), and reduced with sliding speed (lifetime and thus number of bonds).

Bartenev’s explanation for the drop in friction differed from Shallamach’s. He found that the fall in friction at high sliding speeds corresponded to a similar fall as temperature was reduced. He thus invoked time-temperature transformation to suggest that at high sliding speeds, the polymer’s mechanical properties close to the interface changed to become similar to its properties at its glass transition temperature [75].

A considerable problem in developing models of polymer friction lies in fact that the real area of contact can vary considerably with speed and load due to surface deformation and junction growth and this change in contact area can obscure intrinsic variations of interfacial shear strength with load and speed. To address this issue, both Schallamach and Bartenev carried out studies to observe the real area of contact during the sliding of elastomers. Bartenev et al. measured real contact area and friction simultaneously while increasing sliding speed and his findings supported the logarithmic dependence of friction on speed predicted by his model [77]. Schallamach, by contrast, observed the presence and motion of deformation waves now termed Schallamach waves for polymers with low elastic modulus [78].

Schallamach’s initial model of polymer friction was applied by Stejin in 1968 to interpret the measured friction properties of PTFE [79], while his bond-breaking and reforming model was considerably refined in 1986 by Chernyak and Leonov to take account of the actual force–extension properties of polymer chains and the way that momentum is transferred from polymer to surface during this stretching process [80].

In the late 1970s Tabor and co-workers studied the dependence of the friction properties of polymers on load, sliding, and temperature [81, 82]. By using thin polymer films of thickness ca 200 nm on very smooth, rigid substrates, contact areas close to those predicted by Hertz were obtained so that mean shear stress could be determined from the measured friction. They found that shear stress was proportional to pressure and increased linearly with log(sliding speed) at a fixed temperature for most polymers, although one polymer, PMMA, gave shear stress that reduced linearly with speed at low temperature. This anomalous result was ascribed to a viscoelastic response [82]. They developed an empirical equation relating stress to pressure, sliding speed, and temperature closely based on stress-augmented thermal activation.

Much more recently Bouhacina et al. have used an AFM to study the friction properties of polymer films grafted on a silicon substrate [83]. Like Tabor, they found friction to increase with log(sliding velocity) and interpreted this in terms of Eyring’s model.

4.2 Boundary friction

In 1982 Briscoe and Evans studied the influence of load, sliding speed, and temperature on the boundary friction of sliding surfactant monolayers [84]. The test set-up involved a pair of crossed, mica-covered glass prisms and monolayers of fatty acids and fatty acid soaps were deposited on the mica surfaces using the Langmuir Blodgett method. Optical interferometry was used to measure the separation of the mica surfaces and also the precise contact area during sliding. This approach gave very accurate measurement of shear stress of the sliding contact.

For fatty acid monolayers, shear stress, \( \tau \), was found to increase linearly with load, to increase with log(sliding speed) (Fig. 8(a)) and to decrease linearly with temperature (Fig. 8(b)). Similar behaviour was also seen when multilayers were used. With fatty acid soaps formed by depositing fatty acids at high pH, friction was constant or even fell with sliding speed.

Briscoe and Evans interpreted their fatty acid results in terms of the Eyring model with an added pressure work term as introduced by Ward [43], to give the equation for the average velocity of a molecule at the sliding interface

\[
\nu = 2\nu_0 e^{-\frac{Q'}{k_B T}} \sin \left( \frac{\tau_0}{k_B T} \right) \tag{28}
\]

where \( \nu \) is the effective vibration frequency of the sliding molecules, \( b \) is the distance across the energy barrier, \( Q' \) is the thermal activation energy for the flow process, \( \Omega \) is the pressure activation volume,
and $\phi$ is the shear activation volume. Assuming the sliding velocity, $V$ is proportional to the average molecular velocity gives

$$V = 2V_o e^{\left(\frac{Q' + D_p}{k_b T}\right)} \sin h\left(\frac{\tau \phi}{k_b T}\right)$$  \hfill (29)

where $V_o$ is an unknown velocity constant.

At high stresses this becomes

$$V = V_o e^{\left(\frac{Q' + D_p - \tau \phi}{k_b T}\right)}$$  \hfill (30)

or

$$\tau = \frac{k_b T}{\phi} \log\left(\frac{V}{V_o}\right) + \frac{1}{\phi} (Q' + p \Omega)$$  \hfill (31)

This predicts the linear dependence of $\tau$ on $p$ and $T$ and its logarithmic dependence on $V$ found experimentally. Based on this equation, at constant sliding speed and temperature it also predicts the two term Derjaguin equation for friction where

$$F = c + \alpha W$$  \hfill (32)

and the constant $\alpha$ corresponds to $\Omega/\phi$. Based on estimated values of the shear activation volume, $\phi$, Briscoe and Evans concluded that a flow unit comprised several fatty acid molecules. They interpreted the pressure activation volume, $\Omega$, as the local increase in volume needed to permit molecular motion and found that this corresponded to about the volume of an ethyl group.

Subsequently a number of studies have studied the friction of surfactant solutions in hydrocarbons in macro-scale sliding contacts and found that friction depended on log(sliding speed), in accord with Briscoe and Evans’ model [85–87]. This behaviour appears to depend strongly on the ability of the surfactant to form a close-packed monolayer, so that if this layer is disrupted in some fashion, friction tends to become independent of speed. Thus Ingram et al. found that some blends of low and high chain length fatty acids gave speed-independent friction even though solutions of the individual acids gave logarithmic dependence [86]. Campen et al. found that while oleic acid solutions gave friction constant with speed, the same concentration of its trans-isomer, elaidic acid, showed logarithmic dependence, similar to stearic acid [87]. The trans-isomer is able to pack more closely in absorbed films than the cis-oleic acid.

In 2003, Drummond et al. used a surface forces apparatus (SFA) to measure the friction properties of a pair of cationic surfactant-coated mica surfaces in aqueous solution over a wide sliding speed range [88]. They were able to measure the separation and the precise contact area using optical interferometry and thus convert friction to shear stress. Four stages of behaviour were observed, as shown in Fig. 9. At very low sliding speeds shear stress increased with log(speed), before levelling out to a constant value. At still higher speed stick slip was observed with the friction and thus shear stress transiting between high and low values. For some surfactant concentrations at very high sliding speeds, smooth sliding returned, with shear stress again increased approximately with log(speed).

Drummond and coworkers interpreted these results in terms a stress activation model based closely on Shallamach’s 1963 model [76], with the contact consisting of many nano-junctions that can form, stretch,
Fig. 9 Dependence of shear stress on driving velocity while shearing two adsorbed monolayers of trimeric ammonium tribromide-based surfactant from aqueous solutions at $T=20 \, ^{\circ}C$. Reproduced with permission from [88]. Copyright AIP, 2003.

and break independently. Thus the first stage of the shear stress/speed curve is dominated by the Eyring principle of stress-augmented thermally activated breaking of these junctions with the now recognisable $\sin h^{-1}$ or, at higher speed, logarithmic dependence of shear stress on sliding speed. Eventually this process saturates and the shear stress levels out before starting to fall as the number of nano-junctions reduces. The stick slip stage ensues as a result of the well-known instability that occurs in elastic systems when static friction is lower than dynamic friction. Eventually, at very high speeds Drummond et al. suggested that all nano-junctions were broken and friction originated from thin film viscous shear. Mazuyer et al. have recently applied Drummond’s model to the measured friction behaviour of solutions of two organic friction modifiers in polyalphaolefin base fluid in a surface forces apparatus [89].

5.3 AFM friction

In the late 1980s atomic force microscopes (AFMs) started to be used to study friction. Initial work focussed on the stick-slip behaviour generally observed in this system and this was analysed in terms of the Prandtl-Tomlinson model [90]. However, in many cases, both mean and maximum friction also increased with log(speed) and in 1990 this was analysed by Gnecce et al. [91] using a stress-augmented thermal activation model very similar to Prandtl’s 1928 concept [2]. They considered an AFM tip, elastically constrained by a cantilever, sliding up the flank of a sinusoidal force field in a system identical to that shown in Fig. 2. They described the change in probability of it not having jumped (i.e., the tip not having slipped forward), $p(t)$ with time $t$ as

$$\frac{dp(t)}{dt} = -f_o e^{-\Delta E^* / k_b T} p(t) \left( \frac{df}{dt} \right)^{-1}$$

(33)

where $f_o$ is a characteristic vibration frequency and $\Delta E^*$ is the energy needed to overcome the remaining potential barrier in the forward direction (equivalent respectively to the reciprocal of the oscillation time $\tau$ and to $U_i$ in Prandtl’s Eq. (14)). This equation is couched in terms of probability of not having jumped since the jump rate at time $t$ will be proportional to this.

In principle, the product of this equation and the corresponding elastic force at time $t$ could be integrated to relate the total force to the velocity, as was done by Prandtl, but instead Gnecco et al. transformed it in terms of the force $F$ using $dp(t)/dt = (dp(F)/dF). (dF/dt)$, so that

$$\frac{dp(F)}{dF} = -f_o e^{-\Delta E^* / k_b T} p(F) \left( \frac{df}{dt} \right)^{-1}$$

They then expressed this in terms of sliding velocity, $v$ using $dF/dt = (dF/dx). (dx/dt)$ to give

$$\frac{dp(F)}{dF} = -f_o e^{-\Delta E^* / k_b T} p(F) k_{eff} v$$

(35)

where $k_{eff}$ is the effective spring constant of the cantilever.

They assumed a linear reduction of energy with force

$$\Delta E'(F) = \lambda (F' - F)$$

(36)

where $\lambda$ is effectively an activation length (the distance from the bottom to the top of the energy barrier) and $\lambda F^*$ is the energy needed for the tip to jump at absolute zero temperature (the thermal activation energy needed to jump in the absence of an applied force).

To solve Eq. (35), Gnecco et al. assumed that jump would occur at the maximum probability position where $d^2p(F)/dF^2 = 0$ and so obtained the force at the most probably position for the AFM tip to slip

$$F(v) = F' + \frac{k_b T}{\lambda} \log_e \left( \frac{v k_{eff}}{f_o k_b T} \right)$$

(37)

From the above, the close analogy between this approach and Prandtl’s should be evident, with
Prandtl’s single elastically-constrained atom being substituted by a single elastically-constrained AFM tip. It does not, however, appear that Gnecco et al. were aware of Prandtl’s previous work and it was not until about 2002 that Prandtl’s 1928 paper started to be referenced in this context [92].

In subsequent research, Gnecco’s approach was extended to include a more realistic, parabolic dependence of energy on force, i.e., $\Delta E'(F) \propto (F' - F)^{3/2}$ and this predicted a dependence of friction force on $\log(v)^{2/3}$ rather than a $\log(v)$ [93]. Unfortunately it is quite different to test the validity of this without measuring the variation of friction with speed over a speed range of at least two orders of magnitude.

5.4 Friction simulation

As well as the above analytical approach, there have been a considerable number of studies applying molecular dynamics simulation (MDS) to model sliding friction of tethered or adsorbed monolayer boundary films [94–100]. The majority of these have found friction to increase with $\log$(sliding speed) and compared this with Briscoe and Evans’ stress-augmented thermal activation equations. However, some simulations have predicted other types of friction/speed response. Chandross et al. found friction to be proportional to $\log$(sliding speed) at low loads but to be independent of speed at high loads [96]. Chen et al. found that monolayers films, and indeed bare substrates, that terminated with hydrogen acceptor/donor species and could thus form links across the contact gave friction that decreased with speed, while those that could not form such cross links showed friction proportional to $\log$(sliding speed) [98].

In almost all of the above friction studies, both experimental and modelling, the main finding suggestive of stress-augmented thermal activation has been that the friction force or shear stress increases proportional to $\log$(sliding speed). According to both Prandtl’s and Eyring’s models this should, of course, only be the case at high stresses when the backward jump process becomes negligible. As will be discussed later in this paper this is generally true in the context of sliding friction where contact pressures and thus shear stresses are likely to be very high. One exception is Drummond’s study where at very low sliding speeds an initial linear dependence of friction on speed was observed. However this study used a surface forces apparatus where contact pressures are generally much lower than in normal rough surface tribological contacts.

In 2011 Müser used molecular dynamics simulation to predict dry friction over an extremely wide sliding velocity range and found friction to vary with $\sinh^{-1}(\text{velocity})$ as shown in Fig. 10 [101]. Although Müser did not interpret this explicitly in the context of combined forward and reverse transitions, the similarity to the prediction of the Eyring’s model is quite striking.

![Fig. 10](image)

Variation of kinetic friction with sliding velocity and temperature predicted from a molecular dynamics model. Reproduced with permission from [101]. Copyright AIP, 2011.

6 Mechanochemistry

6.1 Background

A fourth area of application of stress-augmented thermal activation to Tribology concerns mechanochemistry and this section will first briefly consider mechanochemistry as applied to chemistry as a whole and then focus on its specific application to Tribology.

The idea that mechanical forces can promote chemical reactions has a long history. Thus, while studying the reaction of silver salts with reactive metals, Faraday in 1820 noted [102]:

“If dry chloride of silver in powder be triturated in a mortar with zinc filings, then the two bodies immediately act, and a heat above that of boiling water is produced.”

The actual term mechanochemistry appears to have been coined by Ostwald in 1909 although the
mechanical force that he had in mind was pressure rather than shear [103].

As described in Section 2 of this paper, the concept that mechanochemistry might originate from a combination of stress and thermal activation causing chemical bonds to break was first developed by Beuche and Zhurkov in the 1950s, both of whom developed reaction rate models that added a work energy term to the Arrhenius equation (Eqs. (11) and (12), respectively). Figure 11 shows some of Zhurkov's data relating the lifetime to fracture, \( \tau \) (the reciprocal of the fracture rate), to the applied tensile stress for a metal, a ceramic, and an organic glass, illustrating both the exponential dependence of lifetime on tensile stress and also the impact of temperature on reducing the fracture lifetime [24].

For metals and ceramics, Zhurkov suggested that his thermal activation energy \( (U_o \text{ in Eq. (12)}) \) represented the binding energy of metallic and ionic bonds. For polymers it was not clear whether covalent bonds within molecules were being broken or van der Waal bonds between polymer molecules, allowing the latter to slip past one another. However, in 1974 Zhurkov and Kursukov used infrared spectroscopy to monitor the breaking of covalent polymer bonds and showed that this process was paramount in determining the dependence of fracture on stress [25]. They also showed that the values of \( U_o \) calculated from the dependence of fracture time on temperature in Eq. (12) were the same as those in thermal degradation experiments for the polymers tested, writing:

"The equality of the activation energies indicates that a connection exists between mechanical, mechanochemical, and thermal destruction of polymers [25]."

Interestingly this appears to be the first time Zhurkov employs the term “mechanochemical” in one of his papers.

In the field of chemistry an important stage in the development of mechanochemistry was made by Bell in 1978 [104]. He applied Zhurkov’s model to describe the cross-linking reactions of biomolecules such as proteins with cell membranes and the consequent effect of this on cell-cell and cell-surface adhesion. He showed that the kinetics of this process was well-described if the debonding process was driven by Zhurkov’s stress-augmented thermal activation lifetime equation (Eq. (12)).

In the last two decades mechanochemistry has become a defined area of research in the field of chemistry and its significance can be gauged from the considerable number of review papers on the subject in recent years [105–110]. Bell’s 1978 paper is generally taken as the seminal starting point, although occasionally earlier work by Zhurkov and Eyring is cited. Mechanochemistry is now being applied to a wide range of areas of chemistry and biochemistry, including solvent free synthesis routes [108, 111], molecular motors [112, 113], single molecule studies, for example using molecular tweezers or AFMs [114, 115], cell differentiation [116, 117], and mechanophores [118, 119]. The latter are molecular groups designed specifically to respond to mechanical force, to enable this force to be measured externally (sensors) or to produce materials that respond in desired ways to applied forces (smart materials).

Before discussing the application of mechanochemistry to Tribology it is important to define what is meant by this term in the context of the current review. The IUPAC gives a very broad definition of a mechano-chemical reaction being a “chemical reaction that is induced by the direct absorption of mechanical energy” [108]. This has led to some confusion about whether the term “mechanochemistry” should encompass indirect chemical effects of mechanical force.
on solids, such as local temperature rise or increase of surface area [108]. Within tribology such a broad definition would, of course, encompass all of tribochemistry, and nowadays most chemists and also most tribologists limit the term mechanochemistry specifically to describe chemical processes where the application of a mechanical force to individual molecules directly promotes bond breaking and thus increases chemical reaction rate. The reaction rate will then depend on applied force and temperature according to a stress-augmented thermal activation rate equation. This focused definition will be the one assumed in the current review.

6.2 Mechanochemistry in Tribology: Polymer degradation

Until very recently the only application of stress augmented thermal activation to chemical reactions in Tribology concerned polymeric viscosity modifier additives. Polymers are especially susceptible to damage by mechanical forces since these forces acting on many monomers in the chain can accumulate to produce large tensile stresses at mid chain or entanglement points and so rupture the chain. This effect became an important bulk polymer processing tool in the 1920s and was patented in 1937 for use in producing viscosity modifiers additives [120]. Then, in the 1940s it was found that very high molecular weight (MW) polymers in solution could show such chain rupture when subjected to high shear rates, resulting in permanent viscosity loss and limiting the molecular weight of polymers that could be used in lubricants [121]. A short review of permanent viscosity loss of viscosity modifiers has recently been provided by Marx et al. [122], while a general review of the mechanochemistry of polymers up to the late 1980s is given by Sohma [123].

In 1940 Kauzmann and Eyring, at the end of their paper on polymer viscosity, suggested that one mechanism of shear thinning was rupture of polymers caused by cleavage of their backbone C-C bonds due to mechanical force [14]. They related this to the influence of force on the potential between the bonded C-C atoms as shown in Fig. 12, which illustrates how an applied force linearly reduces the bonding energy as the bond is stretched. They then considered how this process might be either reversible, to give temporary shear thinning, or irreversible, to reduce viscosity permanently.

In 1955 Beuche developed a model to describe the rupture of polymer chains under stress in bulk polymers above their glass transition temperature [19]. The rate of segment rupture was given by

$$\frac{dn}{dt} = \omega N \sum_{i=0}^{N} e^{-(E-V_i)/k_B T}$$

where $\omega$ is the segment oscillation frequency, $\nu$ the number of active polymer chains per unit volume, $N$ the number of segments in a network chain, $E$ the thermal activation energy to break a segment, and $V_i$ the elastic energy stored in segment $i$ due to an applied force. This elastic energy was initially defined in terms of the bond stretch, calculated from the length of the bond, the applied tensile stress, and Young’s modulus of the polymer, in a way quite similar to Becker’s stored elastic energy model of plastic deformation outlined in Section 3.

However, in 1960 Beuche modified his definition of $V_i$ to one more similar to Eyring’s, with the elastic energy replaced by the product of a force and an activation distance [21]. The probability of chain bond $q$ rupturing was then given by

$$\frac{dn}{dt} \propto \omega e^{-((E-V_i)/k_B T)}$$

where $\delta$ is the approximately half the distance the bond will stretch before breaking (the activation length)
and \( F_q \) is the force on the bond number \( q \) from the centre of an active chain segment having total number of links \( Z \) given by

\[
F_q = F_o \left[ 1 - \left( \frac{4q^2}{Z^2} \right) \right]
\]  

(40)

\( F_o \) is thus the maximal force experienced by the central bond and Beuche showed that this was proportional to the product of the shear rate and the bulk viscosity, i.e., the shear stress acting on the whole active chain. A very similar model to Beuche’s was proposed by de Gennes in 1974 [124] who also showed that various bond force-distance relationships would all result in an approximately linear decrease of activation energy with applied force.

Perhaps fortunately from the point of view of viscosity modifier polymer stability in lubricants, application of laminar shear subjects a fluid to a combination of rotation and deformation and thus does not impose very large tensile forces. For this reason, fundamental studies of polymer scission in solution are often carried out using elongational flow, where degradation occurs at much lower strain rates. Odell and co-workers studied polymer scission in such flow conditions and showed that the rate of polymer rupture followed a thermally-activated bond scission model very similar to that of Beuche [125, 126]. They also suggested that most observations of polymer degradation in laminar shear might be due to local turbulence, for example at a capillary inlet. This highlights a key feature that will be discussed later in this paper—that mechanochemistry is dependent on bond stretching and consequent breaking, so that its fundamental study in rubbing contacts requires estimation of tensile forces on molecules from more easily accessible shear forces.

6.3 Mechanochemistry in Tribology: Recent work

Polymer degradation as outlined above is a process that takes place in bulk polymers or in solutions under shear. It has also long been recognised that the rubbing of contacting asperities in boundary and mixed lubrication conditions can stimulate chemical reactions at surfaces to form tribofilms. This whole field is generally termed tribochemistry and, as will be discussed in the next section, several different mechanisms have been suggested as drivers of tribochemical reactions. Recognition that stress-augmented thermal activation might cause such reactions at surfaces and the description of this as mechanochemistry, is however quite recent. It appears to have originated to describe the shear-induced removal of atoms, ions, or molecules from surfaces when developing models of wear, as will be outlined in the next section, and only in the last three years has it been formally applied to tribochemical reactions. Since 2015, however, progress has been quite remarkable and stress-augmented thermal activation models are now being used to describe antiwear and extreme pressure (EP) additive reactions as well as polymerisation in rubbing contacts.

In an influential paper in 2015, Gosvami et al. measured the rate of film formation by the antiwear additive zinc dialkyldithiophosphate (ZDDP) on silicon and iron-coated surfaces in a liquid cell AFM at elevated temperature [127]. As shown in Fig. 13, the topography mapping capability of AFM enabled film formation to be monitored as a function of applied load and temperature. An exponential dependence of film growth rate on pressure was observed, and the observed temperature dependence was correlated with a stress-augmented thermal activation model

\[
\Gamma_{\text{growth rate}} = \Gamma_0 e^{\frac{\Delta G_{\text{act}}}{kT}}
\]  

(41)

where \( \Gamma \) is the ZDDP film growth rate, \( \Gamma_0 \) a pre-factor and \( \Delta G_{\text{act}} \) the free energy activation energy of the rate growth process.

Fig. 13  ZDDP tribofilm growth rate versus contact pressure on a silicon surface. Topography images are after 2,000 cycles. Reproduced with permission from [127]. Copyright Science ASSS, 2015.
limiting reaction step, given by an expression with a thermal activation and a stress activation term

$$\Delta G_{\text{act}} = \Delta U_{\text{act}} - \sigma \Delta V_{\text{act}}$$  \hspace{1cm} (42)

Here $\Delta V_{\text{act}}$ is the activation volume and $\sigma$ the driving stress. Goswami assumed this driving stress to be pressure but, as will be discussed in the next section, it is now believed to be primarily the shear stress (or rather the tensile stress inflicted on the ZDDP molecules via the shear stress). Generally, of course, in boundary lubrication the shear stress will be approximately proportional to the pressure at a given temperature, at least over a limited pressure range.

Tysoe and co-workers studied the tribochemistry of two model sulphur-based extreme pressure (EP) additives, dimethyl and diethyldisulphide, on copper. They investigated each stage of the reaction process: (i) adsorption/reaction of EP molecules on the surface to form alkylthiolate; (ii) the reaction of this alkylthiolate during rubbing to break C-S bonds and release sulphur atoms; (iii) the forced diffusion of these S atoms into the metal substrate [128, 129]. By measuring the release of hydrocarbon species using mass spectrometry they were able to monitor the second stage of this process and found that the reaction rates measured during rubbing were consistent with Bell’s stress-augmented thermal activation model [129]. Molecular dynamics simulation was also used to examine the likely response of individual methylthiolate (CH$_3$S) species at the surface during shear. The applied force tilted the molecules and the resulting stretching force promoted breaking of the C-S bond, as shown in Fig. 14. Comparison of the measured reaction rate and MD predictions indicated that the rate-limiting step for the reaction was methylthiolate decomposing to adsorbed methyl and sulphur species.

Also in 2015, Felts et al. studied the influence of contact stress on the removal of oxygen atoms bonded to graphene by a sliding AFM tip [130]. Since oxygen-free graphene gives much lower friction than graphene coated with bound oxygen, the kinetics of oxygen removal could be monitored by measuring the decrease of lateral force. The results correlated well with a stress-augmented thermal activation model, with an activation volume of 11 Å$^3$ and a thermal activation energy of 0.75 eV. The authors measured the tip geometry after sliding and were thus able to obtain good estimates of contact pressure and shear stress. Like Goswami however, they appear to have used normal pressure as the driving stress in their model.

Two recent studies have investigated the effect of contact stress on tribopolymerisation. Yeon et al. investigated the rate of formation of solid reaction product in a sliding silicon oxide on glass contact lubricated by allyl alcohol vapour, using an AFM to monitor the volume of polymerised product [131]. They also carried out parallel MD simulations. Both experiment and model found reaction rate to increase exponentially with pressure and shear stress as shown in Fig. 15.

He and Kim studied the adsorption of, and then rate of formation of polymeric deposit by $\alpha$-pinene, pinane, and n-decane during sliding of stainless steel on stainless steel [132]. They used reflection adsorption IR spectroscopy to measure adsorption and AFM to monitor the volume of solid material produced in and around the contact. Like Yeon et al., they found the rate of polymerisation to depend exponentially on load. The reaction rate for n-decane was much slower than for the ring-structured pinene and pinane.

The above experimental work has inspired models that predict tribofilm growth, removal, and consequent friction and wear in rough surface contacts by numerical modeling, with tribofilm formation governed by the local temperature and stress in a time-evolving system [133, 134]. The current main limitation of such models is that while film formation rates can be modelled now with reasonable confidence, the processes
and equations that describe film removal are not yet at all well understood or reliable.

6.4 Proof of mechanochemical reaction

It is clear from the above that the recognition that some tribochemical reactions may be driven by stress according to a stress-augmented thermal activation mechanism is growing very rapidly. A word of caution should be noted however. The principle evidence that a tribochemical reaction is driven by stress, and thus falls into the category of mechanochemistry, is almost always an observed exponential increase in reaction rate with applied pressure or shear stress, in some cases supported by molecular dynamics simulation. This presumes however that other possible drivers are absent and/or such other drivers will not also increase reaction rate exponentially with stress. Historically there are many suggested drivers for tribochemical reactions including, in addition to pressure or shear stress, the following:

- Flash temperature;
- Generation of catalytic surfaces;
- Activation/roughening of surfaces;
- Enhanced mixing due to shear;
- Localised non-Boltzmann energy distributions;
- Triboemission.

The mechanochemical studies described in the previous section [127–132] above all used very low sliding speeds in order to minimise significant flash temperature rises, but it is debatable whether the other possible drivers listed above can be safely discounted. For example, it has already been noted in this review that plastic deformation and fracture of metals and ceramics are exponentially dependent on shear stress, while it has also been found that triboemission stems directly from plastic deformation and/or fracture. Thus the possibility of triboemission may drive the observed reactions that cannot be ignored. Another problem is how to distinguish whether shear stress or pressure is the main driving force since the two are strongly correlated in boundary lubrication.

This problem has recently been addressed by Zhang and Spikes using the additive ZDDP [135]. The underlying problem is that the rubbing contact of asperities is an extremely severe and generally poorly defined process during which it is difficult to know or control the conditions and their effects. The use of an AFM tip alleviates some of these issues but not those relating to nanoscale tip damage processes.

Zhang and Spikes reasoned that if ZDDP reaction is driven by stress then it should occur at very high stress even in the absence of asperity contact: for example, in full film elastohydrodynamic lubrication conditions. They thus monitored ZDDP tribofilm formation in a very high pressure EHD contact where there was no asperity contact. They used a high entrainment speed to generate a thick, separating EHD film, but very low sliding speed to minimise flash temperature. They compared the behaviour of ZDDP in two different solvents, one a polyalphaolefin (PAO) that produced very low EHD friction and thus shear stress and one a traction base fluid with very high EHD friction and shear stress, both giving the same EHD film thickness. As can be seen in Fig. 16,
they found that ZDDP in the traction base fluid DM2H formed a tribofilm on the rolling-sliding surfaces at high shear stress, but ZDDP at PAO formed no such film at low shear stress even though all other conditions, including pressure, were otherwise identical. From the measured friction it was evident that, depending on the temperature, ZDDP started to form a film when the local shear stress exceeded about 200 MPa.

The ZDDP tribofilm increased linearly with time and the dependence of the rate of the film growth on shear stress and temperature agreed well with the stress-augmented thermal activation model.

This study is important because, as well as eliminating all of the alternative possible drivers listed above, leaving only contact stress as the driver, it also demonstrated clearly that the ZDDP reaction was promoted primarily by shear stress and not normal pressure. For the ZDDP in traction base fluid studied, Zhang and Spikes found a thermal activation energy of 53 kJ/mol and an activation volume of 180 Å³.

7 Wear

The final area where stress-augmented thermal activation has been applied in Tribology is in the field of wear. This is closely related to the mechanochemistry outlined above since it is presumed that wear involves the breaking of bonds (ionic, covalent, metallic) and consequent release of one or more particles from a solid sliding surface. In 1997 Dickinson et al. studied the dissolution of calcite into calcium carbonate solution as an AFM tip was rubbed over the calcite surface [136]. He found the dissolution (wear) rate increased exponentially with the applied contact force and fitted the data to a “Zhurkov-Arrhenius” model. Based on the radial tip stress he estimated an activation volume of 44 Å³.

In 2000 Kopta and Salmeron investigated the formation of wear scars on mica surfaces rubbed by a Si tip in an AFM [137]. They found that rubbing caused breakage of Si-O bonds in the mica to form defects that developed into 2Å deep wear scars. They fitted the results to a stress-augmented thermal activation model in which the rate of Si-O bond breaking and thus wear was exponentially dependent on the stress.

Gottsmann and Lanz in 2008 studied the wear of cross-linked polyaryletherketone rubbed by a sharp Si tip in an AFM and applied Briscoe and Evans’ model of boundary friction to explain the influence of applied load on wear [138]. More recently Jacobs and Carpick rubbed a silicon tip against a diamond flat in vacuum in a transmission electron microscope (TEM) and used the TEM images to measure the change in volume of the tip and thus the wear rate [139]. As shown in Fig. 17, they found this wear rate to vary exponentially with applied mean normal contact stress and attributed this to a stress-augmented thermally activated reaction with thermal activation energy 0.91 eV and an activation volume based on contact pressure of 6.7 Å³. This corresponded to an activation volume of 55 Å³ if shear stress were considered to be the driver.

Fig. 16 Comparison of optical interference images for ZDDP solution in PAO and DM2H over a four-hour test at 75 N and 100 °C (U = 3 m/s, SRR = 3%) [135].

Fig. 17 Influence of contact pressure on the wear rate of pure silicon against diamond in a TEM. Reproduced with permission from [139]. Copyright Nature, 2013.
8 General discussion

8.1 Connections

It is striking from all of the research outlined in the previous five sections how the principle of stress-augmented thermal activation links together apparently disparate areas of Tribology. Not only are models of plastic deformation, liquid rheology, friction, tribochemistry, and wear all based on the same underlying concept but also the boundaries between them become blurred. Thus Eyring’s model, originally developed to describe liquid viscosity, has been extended and used to describe plasticity, while Prandtl’s model spans the whole range of behaviour from crystal plasticity to viscosity as temperature is increased. Kauzmann and Eyring’s polymer shear thinning and Shallamach’s polymer friction model are based on the formation and breaking of bonds and are thus precursors of mechanochemistry, suggesting that friction and tribochemistry may be interlinked at a fundamental level. Atomic-scale wear and mechanochemistry are also directly related.

It is not always clear from the literature how Prandtl’s and Eyring’s original concepts spread and the extent to which they were independently rediscovered by subsequent researchers. Figure 18 attempts a chart showing the links from researcher to researcher where these are reasonably clear. It is evident that the studies of Eyring and of Zhurkov are the origins of most recent work.

8.2 Activation constants

One point of some confusion in the literature concerns the significance of shear activation volume. Prandtl, Eyring, Schallamach, and Beuche used an activation distance $\Delta x$ in their versions of the rate equation $k = Ae^{-\frac{(a - \Delta x)}{k_BT}}$, respectively employing the nomenclature, $a$, $\lambda/2$, $\gamma$, and $\delta$ in Eqs. (16), (8), (10), and (11) ($a$ does not appear explicitly in Prandtl’s Eq. (16) but is equal to $A/B$). These all have unit of length and so, when multiplied by the applied force $f$, provide the work term $f\Delta x$ that promotes the process. In practice, the activation length generally approximates to the distance from the initial valley to the top of the potential energy barrier that must be surmounted for the process to occur. It thus corresponds broadly to half molecular spacing or bond length.

Unfortunately, in experimental work the force on a molecule is not generally known; what is known is the applied shear stress calculated from the friction force divided by the contact area. Thus most experimental researchers have employed rate equations of the form; $k = A\exp\left(-\frac{(E_a - \tau\Delta\sigma)}{k_BT}\right)$ where $\tau$ is the shear stress and $\Delta\sigma$ is called the shear activation volume. This has units of volume and so when multiplied by stress forms the required work term. It is very simply determined from the gradient of applied stress versus log(rate), where rate is sliding speed, strain rate, tribofilm formation rate, etc. This gradient is $k_BT/\Delta\sigma$.

It is, however, important to note that $\Delta\sigma$ is not actually a volume but, as should be clear from Eyring’s derivation in Section 4.1, the product of the area over which the stress acts multiplied by an activation distance. If the driving stress is a shear stress, this active area corresponds to the cross-sectional area of a molecule or set of molecules in the plane parallel to the direction of stress. If the stress acts over a tensile stress, the activation area is, of course, orthogonal to the applied stress. This stress may be acting on a single molecule, in which case the active area is the area of one molecule. However, if the interface contains defects, the stress acting on several molecules may focus the resulting force on a single molecule to help drive it over an energy barrier. In this case the active area will be larger, and possibly much larger, than a single molecule, as is the case in the deformation of metals.

---

Fig. 18 Timeline showing connections in application of stress-augmented thermal activation.
It is of interest to compare the various values of shear activation volume determined by researchers. Eyring stresses are typically in the range 5 to 50 MPa. From Eyring’s model, (Eqs. (22) and (23)), Eyring stress is equivalent to \( 2k_B T/\lambda_2\lambda_3 \lambda \) where \( \lambda_2\lambda_3 \) is the area of a molecule over which the stress acts and \( \lambda \) is the distance a molecule or molecular segment moves in a jump (twice the activation length). \( \lambda_2\lambda_3\lambda/2 \) is thus the activation volume and, over the Eyring stress range above, it spans 100 to 1,000 Å\(^3\). For approximately spherical molecules, if the shear activation originates from the force on individual molecules, the shear activation volume will be approximately half that of the cube containing a molecule, so this can be compared to the molecular volume. A hydrocarbon lubricant of average MW 400 will have a molar volume of ca. 350 cm\(^3\) and so a mean molecular volume of ca. 580 Å\(^3\). Of course, because of free volume, the actual occupied molecular volume will be somewhat smaller than this. For non-spherical molecules, if we arbitrarily assume a thickness perpendicular to the shear plane (Eyring’s \( \lambda_i \)) of 10 Å, we obtain a molecular area on which the shear stress acts of 20 to 200 Å\(^2\). As noted by Hirst and Moore [54], these are not unreasonable values, in that for polymeric molecules the flow unit is likely to be a polymer segment, while for molecules having short side chains some interlocking is likely so the flow unit may well be more than one molecule.

In studies on boundary friction, Briscoe and Evans [84] and Chugg and Chaudri [85] estimated shear activation volumes of the fatty acids studied of approximately 1,500 to 6,000 Å\(^3\). They suggested that this was equivalent to about 2-6 molecular volumes and that a small number of molecules were therefore combined as flow units. However, the authors compared molecular volume with activation volume and this is probably not appropriate for vertically-oriented molecules. As discussed above, the activation volume is generally considered to be the area of the molecule times its activation length. The area of a fatty acid head group is around 20 Å\(^2\) [140] from which its activation length will be about 2.5 Å in a close packed monolayer. This suggests an activation volume of ca. 50 Å\(^3\), implying that 30 to 120 molecules may be involved in experiencing the driving stress. This raises the possibility of slip resulting from localised dislocation motion, as deduced for plastic flow of metals.

In mechanochemistry, Gosvami et al. [127] calculated an activation volume for ZDDP tribofilm formation of 3.8 Å\(^3\) but this value presumed the driving force was pressure. If it were shear stress then, assuming a friction coefficient of 0.1, the shear activation volume would be 38 Å\(^3\). Zhang and Spikes [135] determined an activation volume value for ZDDP of 180 Å\(^3\). It is, of course, extremely unlikely that the two ZDDPs studied by Gosvami and Zhang were the same. For breaking of the Si-O bond Felts [130] found an activation volume of 11 Å\(^3\) but it is not clear whether this was calculated on the basis of pressure or shear stress as the driver. Yeon et al. [131] calculated an activation volume of 7.8 Å\(^3\) for tribopolymerisation of allyl alcohol. If we assume an activation distance of about 1 to 2 Å, corresponding to a covalent bond length, these values yield activation areas that are reasonably representative of the areas of single molecules.

As already mentioned, the values of activation volume estimated for plastic deformation of metals are much larger than for most other applications of stress-augmented thermal activation. This is because plastic deformation involves the motion under stress of 2D structural entities consisting of many atoms, so the activation dimensions are considerably larger. They can thus be regarded as meso-scale stress-augmented thermal activation models. This is especially evident in Becker’s [6] model which involves a bulk property of the material, its elastic modulus, but is also the case for Kauzmann’s model [13].

### 8.3 Forms of stress

While it may be reasonable to use shear stress in stress-activated models of friction, rheology, and plastic deformation, it is probably less appropriate in the context of mechanochemistry. Here it is generally believed that tensile forces act on atom-atom bonds to promote bond breaking, so rate equations should really be based on tensile stress or force. Unfortunately, we currently do not have reliable ways to measure the tensile forces imparted to lubricant molecules from the shear stresses present in asperity-asperity or EHD contacts. The rate equations employed in Refs. [127–132] all implicitly assume proportionality between shear force and tensile force. The issue is complicated by the fact that we need to consider the tensile forces...
experienced by molecular bonds at surfaces, which may well be different from those in the bulk. Thus, based on their modelling work, Yeon et al. noted [131]:

“It appears that anchoring one molecule to the surface helps transfer the mechanical force or action from the solid surface to the other molecule being reacted.”

Another stress that should be considered is normal pressure. As outlined in Section 5, some researchers include a pressure term in the stress-activated rate equation, as shown in Ward’s Eq. (18) and Briscoe and Evans’ Eq. (31). Here the applied hydrostatic pressure is assumed to influence the rate of the process via a pressure activation volume. For friction this term is generally considered to be negative since slip is assumed to require a local, albeit very small, motion away from the counter-surface against the applied pressure. Hirst and Moore also included a negative pressure activation term in addition to a positive shear activation one in their EHD friction analysis [54]. Nowadays in EHD this tends to be subsumed in the highly pressure-sensitive Newtonian viscosity term as in Eqs. (24) and (27) while the Eyring stress is generally also found to increase linearly with pressure.

In mechanochemistry, the sign of a pressure term should depend on the volume change associated with bond breaking and Gosvami et al. implicitly assume there to be a reduction in volume for ZDDP film formation by their use of pressure as the reaction driver [127]. It should also be noted that although shear stress has been subsequently been shown to drive ZDDP tribofilm formation [135] this does not, of course, mean there might not also be a pressure activation term, and further work needs to be done to establish this. This would be very difficult to do in boundary lubrication conditions where normal pressure and shear stress are coupled, but should be feasible using Zhang and Spikes EHD approach [135], where they can be varied independently by changing the solvent friction.

8.4 Forward and reverse reactions

It should be clear from the above that although both the Prandtl and Eyring models of stress activation include both forward and reverse terms, to give, in Eyring’s case the dependence of rate on \( \sin h(\text{stress}) \), in practice in most experimental studies only the forward reaction appears relevant in that the rate is found to depend exponentially on stress. There are a few exceptions, notable in EHD rheology [54, 59], very high temperature plastic deformation [32, 141], and very low speed friction modelling [101], but generally the reverse term is negligible. This can be understood by considering Eq. (7) based on a stress activation volume

\[
k_{\text{net}} \propto \sin h(\tau \Delta v / k_B T)
\]

If \( \tau \Delta v / k_B T \) is greater than about 1.5 then the \( \sin h(\text{ )} \) term becomes practically equal to \( 0.5 \exp(\tau \Delta v / k_B T) \), so that rate increases exponentially with stress. The \( \sin h(\text{ )} \) dependence will thus only become evident if \( \tau \Delta v / k_B T \) is below this level. At a representative temperature of 60 °C this would imply that \( \tau \Delta v \) should be below about \( 5 \times 10^{-21} \) m. In both EHD and boundary friction, activation volumes are typically 100 to 6,000 Å³, suggesting that then \( \sin h(\text{ )} \) dependence of rate on stress would be evident at shear stresses below somewhere between 1 and 50 MPa. These shear stresses are commonly present in relatively low load EHD contacts, so a \( \sin h \) dependence may be seen, but are not usually found in boundary friction, where shear stresses at asperity contacts are generally much higher.

In mechanochemistry of course, another factor militating against a significant reverse reaction rate is that the thermal activation energy barrier is generally lower in the forward than the reverse direction.

8.5 Force-displacement relationships

So far, almost all of this review has assumed the simplest form of the stress-augmented thermal activation equation, i.e., the applied force reduces the stress activation energy linearly with displacement, \( x \), so \( k_f = A e^{-E(x) / k_B T} \). This is, however, only an approximation and two subtly different refinements have been applied: one by adopting more accurate solutions of the balance between the applied force and the potential field with which it interacts, the second allowing the position of the maximum and minimum of the potential field to be shifted by the applied force.

Prandtl noted that a linear dependence of activation energy on displacement, \( U_i = A - Bx \), was only an approximation in his model and that a more accurate
expression would allow it to vary parabolically so that 
\[ U_x = \text{const.} \left( a - x \right)^{3/2} \] [2]. He showed how an approximate solution could be derived for the jump probability and thus the dependence of total force on sliding velocity and temperature using this relationship, but analysis was complicated and limited [2].

Gnecco et al. [142] applied both linear and parabolic expressions to sliding AFM models and found that the linear form of the stress activation energy (Eq. (36)) was satisfactory for low values of force but that a parabolic form as in Eq. (44) should be used when the force was close to its critical jump value, \( F^* \)

\[
\Delta E^T(F) = \mu \left( F^* - F \right)^{3/2} \tag{44}
\]

As well refinements in the potential versus displacement solution, the effect of applied force on the overall shape of the potential field has been considered, in particular allowing for locations of the minimum and maximum of the potential curve and thus the stress activation distance to vary with applied force. This was considered in the context of plasticity in 1975 when Kocks et al. used a Taylor expansion to examine the impact of allowing the total activation energy to vary non-linearly with both stress, \( \sigma \), and apparent activation area [38]. They concluded that a linear dependence of total activation energy on both stress and activation area was valid so long as the stress did not approach the maximum plane glide resistance.

The influence of applied force on the activation distance has also been considered in the context of mechanochemistry in what is generally known as the extended Bell model [143]. This takes account of the applied force shifting both the potential minimum and maximum as shown schematically in the potential diagrams in Fig. 19, where \( \xi \) is the reaction path coordinate, the solid line is the potential when no force is present and the dashed line shows the effect of applied force. According to extended Bell theory, the total activation energy \( V(F) \) at applied force \( F \) is given by

\[
V(F) = V(0) - FA\Delta x - \frac{F^2}{2} \left( \chi_T - \chi_i \right) \tag{45}
\]

where \( V(0) \) is the thermal activation energy, and \( \chi_i \) and \( \chi_T \) are the compliances of the molecule in its initial and transitions state (i.e., at the minimum and maximum positions of the potential barrier). The product \( F\chi \) is the distance these positions move due to the applied force.

Furlong et al. have explored the impact of the shape of the potential field on sliding friction behaviour [144], while Tysoe has recently provided a detailed discussion of the influence of applied stress on activation barriers and interpreted experimental mechanochemical results using advanced expressions for the activation energy [145].

8.7 Prandtl and Eyring

It is of interest to compare Prandtl’s and Eyring’s models. This was done in some detail in Ref. [16] and so is only briefly summarised here. Although both are based on stress-augmented thermal activation they are fundamentally different in two aspects. Firstly, Prandtl considers the transition across a potential barrier of a finite number of particles. This means that his rate of transition varies as the population of particles on either side of the barrier changes. This gives the \( \mu \) and \( (1-\mu) \) pre-factors in Prandtl’s Eqs. (14) and (15) and it is these that make solution of these equations difficult. By contrast, Eyring essentially assumes a very large ensemble of particles, so large that the number of particles available to transit is invariant and effectively the same in both directions. This results in Eq. (7) where the forward and reverse exponents appear without pre-factors, making the equation much more tractable. It is perhaps a pity
that Prandtl did not realise this simplification but his use of varying population levels either side of the transition did enable him to apply his model to stress relaxation as well as steady state plastic flow.

Models of AFM friction illustrate the extreme of Prandtl's approach in that only one particle (the AFM tip) traverses an energy barrier [91]. Schallamach's revised model which involves the formation and breaking of junctions also has similarities with Prandtl's approach since it takes account of the varying fraction of surviving junctions during sliding [76].

The second main difference is that Prandtl assumes that his sliding particles are elastically constrained so that their displacement is proportional to the applied force. Eyring however implicitly assumes that the constraining force on his particles is independent of their displacement—in effect being constrained by a constant force spring. Eyring's assumption has been embedded in most subsequent work, but Becker's, Schallamach's, and AFM friction models involve elastic constraint.

8.8 A caveat

This review has concerned the application of the concept of stress-augmented thermal activation to develop atomic-scale models of plasticity, friction, tribochemistry, and wear. Before concluding, it is important to note that, while evidently operative in many cases, this should not be regarded as the only mechanism underlying these phenomena. Thus while many processes exhibit the relationship stress \( \propto \log(\text{rate}) \), others do not. For example in EHD at high strain rates most simple liquids show friction that increases with \( \log(\text{strain rate}) \) in isothermal conditions, but traction fluids reach a constant value [146]. In boundary friction Briscoe and Evans found that fatty acid films showed friction proportional to \( \log(\text{sliding speed}) \) but this was not the case for fatty acid soaps [84]. Campen et al. showed that the friction-speed response depended on whether the adsorbing surfactants could form close-packed monolayers [87]. Drummond et al. showed that friction as only proportional to \( \log(\text{sliding speed}) \) up to a critical speed [88]. These suggest that the applicability of stress-augmented thermal activation has a finite scope. Of particular interest is to understand when and why it ceases to be applicable rather than only studying conditions under which it seems to be prevalent.

It is also possible that other mechanisms that control rheology and friction may give observed stress \( \propto \log(\text{rate}) \) behaviour. It has already been noted that network-based rheology models might do this in some cases in the context of EHD friction. The problem then becomes to develop experimental ways of distinguishing between mechanisms. This was done in the field of tribochemistry where a designed experimental approach showed that shear stress (and not many alternative drivers) almost certainly controls ZDDP reaction rate [135]. But this does not mean that all tribochemical reactions are driven by stress. Indeed, almost all the factors listed in Section 6.4 are likely to, and in most cases have been shown to, accelerate tribofilm formation under some conditions, the only exception to date being non-Boltzmann energy distributions.

In summary, stress-augmented thermal activation almost certainly plays a key role in controlling some, and probably many, friction, tribochemical, and wear processes but the range of conditions over which it is important is finite and other processes may compete with or complement it and cannot safely be ignored.

8.9 Future directions

How is the concept of stress-augmented thermal activation likely to develop within Tribology in the near future? Clearly, from the scientific viewpoint, the most important issue is to fully test that the underlying principle of atomic and molecular response due to a combination of temperature and stress is valid. Experiments at the individual atomic and molecular scale are difficult, so the most promising approach is to use molecular dynamics simulation to test whether the patterns of behaviour assumed in stress-augmented thermal activation models do actually occur, i.e., when shear and sliding boundary conditions are imposed, are atoms or molecules with thermal energy induced by shear forces to translate past their neighbours to an extent that matches Prandtl's or Eyring's model? And when a very thin molecular layer is sheared, do the forces transmit to molecules at surfaces stretch and break their covalent bonds in the fashion by mechanochemistry? Crucially molecular dynamics is
able to determine the actual shear and tensile forces experienced by molecules in sheared systems and relate these to the overall ensemble shear stress. A few such simulations of this types have already been made and many more are likely in the near future.

At the experimental level things are more difficult. In the area of EHD friction, one problem has always been that, while we know the shear forces on the two bounding solid surfaces and also the velocities of these surfaces, this does not tell us how the fluid film shears between them. Does the velocity within the fluid vary linearly across the film (Couette flow) as is commonly assumed when estimating the shear rate, or during shear thinning is there shear localisation such as plug flow or a slip band? Without such information it is not possible to test Eyring’s shear thinning model fully, or indeed other rheological models. This issue may be addressed via molecular simulation [69] but a recent, complementary experimental approach is phosphorescence imaging, where the velocity of the fluid across a thin EHD film can be profiled from the convection of a small, laser-excited column of fluid within the film [147].

In terms of tribofilm formation, it is very likely that studies will be made in the next few years on many more additive systems to test whether film formation rate depends on stress and temperature as predicted by stress-augmented thermal activation. Unfortunately, such measurements do not resolve the problem of distinguishing the impact of shear stress and normal pressure since, as discussed in Section 8.3, these are always strongly coupled. They can, however, be decoupled using the EHD film approach outlined in Section 6.4 and this approach should provide useful information about mechanochemistry in general, and the separate influences of shear stress and pressure in particular, over the next few years. Another possibility is to measure covalent bond stretch in high shear stress contacts using a spectroscopic method. This has already been done in some solid-state systems, e.g., Ref. [148], but is technically difficult for very thin films because of the limited number of active centres, and thus the very weak signal. Another possible avenue would be the application of nano-manipulation methods, already developed for chemical studies [109], to investigate the properties of single lubricant additive molecules on surfaces, such as the strength of metal-ligand bonds.

In the context of friction, single contact studies would appear to be a useful way forward, but ideally using stiffer supports than those usually employed in an AFM. Such studies, in parallel with very well-defined surfaces such as Langmuir Blodgett films or single crystals, should help determine the boundaries of stress-induced thermal activation.

9 Conclusions

This paper has described the origins of stress-augmented thermal activation and has outlined its application to Tribology. This concept has provided the basis of a series of closely-related, atomic-, and molecular-based models to explain and predict plastic flow, EHD friction, solid-solid friction, tribochemistry, and atomic-scale wear. The key prediction of these models is that the rate of the process involved (flow, sliding, covalent bond breaking, or atom/ion removal from a surface) increases exponentially with shear stress, so that the stress increases with the logarithm of the rate. For example, it is very often found that friction rises logarithmically with sliding speed in EHD, polymer, and boundary friction while tribofilm formation and polymer scission rates increase exponentially with shear stress.

One area of application that is growing very rapidly is tribochemistry, where several recent studies have suggested that the formation of several types of reaction films on rubbed surfaces is driven by stress-augmented thermal activation. This type of tribochemistry is now often termed mechanochemistry. In principle, both normal stress (pressure) and shear stress might drive mechanochemical reactions and one problem in such studies is to distinguish between the effects of these on reaction rate, since pressure and shear stress are general coupled in dry or boundary lubricated sliding conditions. A recent study has shown how these can be decoupled using a thick film rolling-sliding EHD contact with solvents having very different friction coefficients and this has indicated that shear stress is the key driver in ZDDP tribofilm formation.

The fundamental importance of the stress-augmented thermal activation concept is that it provides a link between the nano-scale behaviour of individual atoms
and molecules and macro-scale responses such as friction and wear. Some future possibilities in research on stress-augmented thermal activation have been outlined. Modern non-equilibrium molecular dynamics simulation methods offer the possibility of validating and exploring further the implications of this mechanism and, indeed, converting such simulations into useful, generalised tools for predicting how molecular structure and surface properties influence macroscale performance. In future, such tools hold the promise of enabling the design of molecules and surfaces for optimised friction and wear performance much more efficiently and rapidly than is currently possible by trial and error.

Finally, it is concluded that stress-augmented thermal activation offers a guiding and unifying principle for explaining many of the processes that take place in sliding tribological contacts.

Open Access: The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References


Hugh SPIKES. He graduated in natural sciences from the University of Cambridge in 1968 and received his PhD degree in mechanical engineering from the University of London in 1972. He joined the staff of the Mechanical Engineering Department at Imperial College in 1978, where in 1996 he became professor and head of the Tribology Group. He is currently a senior research fellow at Imperial College. His research areas cover all aspects of liquid lubrication, from hydrodynamic to boundary, with a particular interest in the influence of lubricant molecular composition on performance. He is a recipient of the ASME Mayo D Hersey Award, STLE International Award, and the Tribology Gold Medal.
Anti-loosening performance of coatings on fasteners subjected to dynamic shear load

Junbo ZHOU¹, Jianhua LIU¹, Huajiang OUYANG², Zhenbing CAI¹, Jinfang PENG¹, Minhao ZHU¹,*  
¹ Tribology Research Institute, Southwest Jiaotong University, Chengdu 610031, China  
² School of Engineering, University of Liverpool, Liverpool L69 3GH, UK  
Received: 02 November 2016 / Revised: 19 March 2017 / Accepted: 24 March 2017  
© The author(s) 2017. This article is published with open access at Springerlink.com

Abstract: This paper investigates the self-loosening of threaded fasteners subjected to dynamic shear load. Three kinds of typical coatings, PTFE, MoS₂, and TiN applied to bolts and nuts, are tested in this investigation. The study experimentally examines the loosening mechanisms of fasteners and assesses the anti-loosening performance of the three tested coatings based on their tightening characteristics, loosening curves, and the damage of thread surface. Additionally, the anti-loosening performance of the three coatings is compared under different load forms. The results indicate that the PTFE and MoS₂ coatings have significant anti-loosening effect, whereas the anti-loosening performance of TiN coating is not satisfactory. It is also found that an appropriate increase of the initial tightening torque can significantly improve the anti-loosening effect. In addition, the microscopic analyses of PTFE and MoS₂ coating reveal that a reduced initial tightening torque leads to fretting wear on the thread contact surfaces of fasteners, thereby aggravating the damage.

Keywords: bolted fastener; loosening curve; fretting, coating; dynamic shear load

1 Introduction

The most common modes of threaded fasteners failure under dynamic loads are fatigue and self-loosening. Self-loosening is often encountered when threaded fasteners are subjected to transverse or shear load, which can cause gradual loss or even complete loss of the clamping force in bolted connections. Ultimately, it is likely to lead to the occurrence of severe safety related accidents. There has been much research on this topic. Self-loosening mechanisms [1–16] and their various influencing factors [17–21] were studied through experiments, theoretical analyses and numerical simulation. A variety of measures were proposed to prevent self-loosening, such as chemical locking [21], step lock bolt (SLB) [22], and double-nut [8, 23]. However, loosening of threaded fasteners is still common. Therefore, an in-depth study of threaded fasteners loosening mechanisms is necessary to improve the anti-loosening performance of the bolted connections.

Recently, the loosening process of threaded fasteners had been extensively studied under different loading conditions. However, there has not been a universal acceptance regarding the self-loosening process mechanisms. The focus of the early research in the field was on loosening due to axial loading (dynamic loads acting along the fastener axis) [1–6]. Goodier et al. highlighted the radial sliding motions between the threads of the bolt and the nut or the interfaces of the clamped bearing surfaces as important factors causing loosening of a threaded coupling structure under axial vibration [1]. Basava and Hess found that the clamping force could remain steady, decrease, or increase when the assembly was subjected to axial vibration, which was related to the vibration frequency, amplitude,

* Corresponding author: Minhao ZHU, E-mail: zhuminhao@swjtu.cn
and the frictional force between the contact surfaces [2]. However, Sakai pointed out that the clamping force might be reduced under the axial vibration [3]. Nassar and co-workers considered the irreversible plastic deformation of bolts under axial vibration as a major loosening mechanism [4, 5]. Additionally, Liu et al. found that the plastic deformation of the structure and fretting wear between contact surfaces were the loosening mechanism of bolted joints under axial excitation [6].

On the other hand, the experimental studies by Junker in 1960s demonstrated that loosening was more severe when a joint was subjected to shear loading [7]. A considerable number of researchers have focused on the responses under the dynamic transverse load or displacement. Sase et al. found that the loosening of fasteners was mainly caused by two factors under a transverse displacement [8, 22]: one factor was the relative slip between the threads of the bolt and nut, while the other factor was the relative slip between the bolt (or nut) surface and the surface of the fastened material. The findings of the loosening analysis of bolted joints subjected to dynamic shear load conducted by Pai and Hess indicated that fastener loosening occurred as a result of complete or localized slip at the thread and head contact surfaces [9, 10]. Fasteners can loosen under lower loads than those previously anticipated because of localized slip at the contact surfaces, which was confirmed by Dinger and Friedrich [11]. Shoji and Sawa found that the relative slip between threads was recurrent, which led to a nut rotation large enough to cause the bolt to loosen [12]. Yang et al. further noted that with the increasing number of load cycles, the relative slip gradually accumulated which continuously reduced the preload and eventually resulted in the failure of the fastening bolts connection [13]. Studies conducted by Jiang and co-workers revealed that the self-loosening process of bolted joints subjected to a transverse displacement could be divided into two distinct stages [14–16]. The first stage did not involve relative rotation between the nut and bolt, and the loosening was caused by material deformation. However, the second stage was characterized by obvious backing-off of the nut and rapid decrease of the clamping force.

Coating and lubrication are often used for reducing the friction coefficients of threaded fasteners [24, 25]. However, the effect of friction coefficients on the loosening process is controversial. On one hand, increased friction coefficient means an increased friction torque for resisting the relative rotation or slip at the thread and head contact surfaces which plays a role on anti-loosening. Daadbin and Chow described a theoretical model for a threaded connection under impact loading [26]. They found that the increase of the friction coefficient decreased the preload loss. Houari [18] and Karamis [27] showed that the increase of the bearing friction reduced the loosening rate. Furthermore, Zaki et al. [19] found that the increase of the coating thickness decreased the friction coefficients of the thread and head bearing, therefore thick coated fasteners would loosen at a faster rate than thin coated fasteners [28]. Sanclemente and Hess discovered that friction coefficients decreased with the application of lubricant which promoted slippage and thereby loosening [20]. However, the study also indicated that when lubrication was applied in combination with higher preload, the resulting effect was beneficial for vibration resistance. On the other hand, the torque–tension relationship for threaded fasteners is highly sensitive to the friction coefficient variation between the turning surfaces at the head/nut interface and threads [29]. The decrease of friction coefficients increases the preload under the same initial tightening torque which leads to a better anti-loosening performance. Liu et al. considered adding MoS2 lubricant to bolted joints as a good method to prevent loosening when the bolted joint was subjected to axial excitation [6]. Hence, it is crucial to investigate the effect of coating on the self-loosening of threaded fasteners.

Solid lubricating coatings are highly regarded in scientific and industrial communities due to their excellent friction reduction and wear resistance properties [30]. Coating used on fasteners can change friction coefficient, and thus different coatings have different anti-loosening performance. However, there is very little research on the anti-loosening performance of coatings. Polytetrafluoroethylene (PTFE) is one of the most commonly used solid lubricants that exhibits an ultra-low coefficient of friction [31]. In addition, molybdenum disulfide (MoS2) is a popular solid lubricant used widely in machinery equipment to reduce or eliminate various wear damage [32]. Titanium nitride (TiN) coating is also widely used due to its high
hardness and high adhesion strength that enhance the surface properties under wear and corrosion conditions [33, 34]. Thus, conducting an experimental research to investigate the anti-loosening performance of these typical coatings will provide theoretical support and engineering guidance to prevent loosening of bolt-connected structures.

In this paper, two kinds of typical anti-friction coatings (PTFE and MoS2) and anti-wear coating (TiN) are applied to fasteners in order to investigate their anti-loosening performance under dynamic shear load. The evolution curves of the bolts clamping force are obtained and the damages of the thread surfaces in contact are analyzed. This study examines the loosening mechanism as fretting of surface coating and assesses the anti-loosening performance of these typical coatings. In addition, this study also explores the differences in the anti-loosening performance of the three coatings under different load forms.

2 Experimental details

2.1 Test device and parameters

Self-loosening experiments of fasteners are carried out using the electro-hydraulic servo fatigue testing machine with a custom designed testing fixture under cyclic shear load. A schematic illustration of the test machine is shown in Fig. 1. The upper fixture is connected to the grip of the fatigue testing machine while the lower fixture is fixed to the test stand. The upper and lower fixtures are made of 25 mm thick 1045 steel and are clamped by a bolt and nut. The load cell is placed between the upper and lower fixtures. It is connected to a data acquisition system to monitor the clamping force in real-time manner. In order to protect the fretting wear of the load cell, two thin washers made of aluminum alloy are placed between the load cell and the bolt testing fixture.

Load-controlled experiments are conducted. The dynamic shear load is the controlling parameter which is applied by the fatigue testing machine and measured by high-precision mechanical sensors.

According to the related standard, an appropriate level of preload corresponds to 60%–70% of the nominal yield strength of the bolt [35]. Accordingly, for M12mm grade-8.8 bolts, the proper preload ranges from 18.2 kN to 21.2 kN. Therefore, the self-loosening experiments in the study are conducted with a preload of 20 kN, as a median value between 18.2 kN and 21.2 kN. However, in practice, a controlling-torque method is adopted to control the preload. An initial tightening torque \( M_0 \) is applied to achieve a specified torque by slowly tightening the bolt using a digital torque wrench. In order to effectively assess the anti-loosening performance of the different coatings, two groups of experimental parameters are selected: (a) applying the same initial tightening torque \( M_0 = 72 \text{ N·m} \); (b) applying the same preload \( P_0 = 20 \text{ kN} \).

Preliminary tests of uncoated fasteners for the self-loosening experiment are performed to determine the test parameters. In the preliminary tests, four different dynamic shear load levels are applied in the experiments, which are ±4 kN, ±6 kN, ±7 kN, and ±8 kN, respectively. It is found that when the shear load is ±4 kN, the self-loosening curve is similar to the loosening curve under the shear load of ±6 kN. When the shear load increases to ±7 kN or ±8 kN, the bolts become completely loosened in less than 5,000 cycles. The reason may be that when the shear load is large enough to overcome the static friction between the two clamping plates, an excessive lateral slippage between them occurs and the clamping force decreases quickly. So the dynamic shear load is selected as ±6 kN in the
coated fasteners experiments. In the preliminary tests, two loading regimes of 200,000 and 1,000,000 cycles are used and it is found that the self-loosening curve yields similar results. To save time, 200,000 cycles of loading are used in the coated fasteners experiments.

A frequency of 10 Hz is selected to carry out the tests due to the following reasons: First, 10 Hz is the frequency commonly used in the field of mechanical engineering. For example, fasteners on train bogies are subjected to an excitation of frequency of about 10 Hz. Secondly, the frequency of 10 Hz is relatively easy to achieve when using the electro-hydraulic servo fatigue testing machine. Hence, all experiments are conducted with 10 Hz frequency at room temperature.

### 2.2 Preparation and characterization of coatings

The bolts and nuts used for the purpose of this study are M12 × 1.75 mm of grade 8.8. In order to reduce the dispersion of the preload and loosening values, a number of high-quality bolts and nuts are machined using 1045 steel. Relevant parameters and thread profile of the testing bolts can be found in Ref. [35]. The composition and main characteristics of 1045 steel are listed in Tables 1 and 2, respectively.

Three kinds of typical coatings, PTFE, MoS₂, and TiN, are applied to the tested blots and nuts in order to study the effect of these coatings on bolt loosening. As previously mentioned, Zaki et al. [19] found that the increase of the coating thickness decreased the friction coefficient between threads and the friction coefficient between the bolt (or nut) and the bearing surface. Therefore, the coating thickness may affect the loosening behavior and it is necessary to choose the appropriate coating thickness before coating preparation. On one hand, if the coating thickness is too thin, the coating may be worn out quickly during the pre-loading process. On the other hand, if the coating is too thick, the fit between the bolt and the nut would be affected, leading to difficulty in applying tightening torque. Thus, a coating thickness of 15 μm is selected for the PTFE and MoS₂ coatings. However, since the TiN coating is typically considered as an anti-wear coating, the thickness of TiN coating can be appropriately reduced. Thus, the selected thickness for the TiN coating is 5 μm.

The preparation of the three coatings is as follows: (1) PTFE coating is prepared using a spraying technique. The fine PTFE powders are uniformly dispersed in epoxy adhesive. After descaling, rusting, and sand blasting, the mixture is sprayed onto the surfaces of fasteners by a spray gun. Then, the coating is solidified by heat curing at 200 °C for 1 hour. The coating thickness is 15 ± 1 μm. (2) MoS₂ coating is prepared in a similar way to PTFE coating. The thickness of the coating is also 15 ± 1 μm. (3) The PVD TiN coating, with thickness of 5 ± 1 μm, is deposited on the fasteners by an ion-plating equipment (MIP-800) at a bias voltage of ~100 V to ~150 V and a mixed gas (N₂ + He) pressure of 0.6 Pa. The characteristic parameters of coatings are listed in Table 3.

Figure 2 shows the morphology of the fasteners after the coating treatment. Figure 2(a) shows a cross-sectional morphology of the coating/substrate. It can be seen from Fig. 2(a) that the thickness of coatings does not change significantly in the observation range. Figure 2(b) shows the thread surface morphology of the three kinds of coated fasteners. It can be seen that there is no spalling in any of the tested coatings. Based on the above observations, we believe that the three coatings are sufficiently uniform.

All test specimens, screws, washers, and fixtures are cleaned with acetone to remove surface contamination before the self-loosening testing. Some uncoated fasteners are also tested for comparison purposes.

Each self-loosening experiment is repeated five times for coated fasteners and uncoated fasteners at identical

### Table 1 The chemical composition of 1045 steel (wt.%).

<table>
<thead>
<tr>
<th>Material</th>
<th>C</th>
<th>Si</th>
<th>Mn</th>
<th>Ni</th>
<th>Cr</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1045 Steel</td>
<td>0.45</td>
<td>0.27</td>
<td>0.65</td>
<td>0.25</td>
<td>0.25</td>
<td>≤0.04</td>
<td>≤0.04</td>
</tr>
</tbody>
</table>

### Table 2 The main characteristics of 1045 steel.

<table>
<thead>
<tr>
<th>Material</th>
<th>Yield strength (MPa)</th>
<th>Tensile strength (MPa)</th>
<th>Elastic modulus (GPa)</th>
<th>Poisson’s ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1045 Steel</td>
<td>650</td>
<td>850</td>
<td>210</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### Table 3 The characteristic parameters of coatings.

<table>
<thead>
<tr>
<th>Coatings</th>
<th>Preparation</th>
<th>Thickness (μm)</th>
<th>Hardness (HV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTFE</td>
<td>Spraying</td>
<td>15 ± 1</td>
<td>23 ± 2</td>
</tr>
<tr>
<td>MoS₂</td>
<td>Spraying</td>
<td>15 ± 1</td>
<td>60 ± 5</td>
</tr>
<tr>
<td>TiN</td>
<td>PVD</td>
<td>5 ± 1</td>
<td>2,500 ± 200</td>
</tr>
</tbody>
</table>

test parameters. After conducting the self-loosening tests, the morphologies of wear scar are examined by a scanning electron microscope (SEM) and the chemical compositions of the damaged zone are analyzed using energy dispersive X-ray (EDX).

3 Results and discussion

3.1 The anti-loosening performance of three coatings

The loosening degree ($\phi$) is defined as the ratio of the loss of preload and the initial preload as expressed by Eq. (1) [20].

$$\phi = \frac{P_0 - P_i}{P_0} \times 100\%$$  \hspace{1cm} (1)

where $P_i$ is the remaining preload after the test, $P_0$ is the initial preload.

$\phi$ is a dimensionless variable which is simply a measure or indication of preload loss. A low $\phi$ value means a slight loosening of bolted fasteners, and thus a better anti-loosening performance, and vice versa.

Figure 3 shows the experimental results of the self-loosening performance of three coated fasteners tested under the same initial tightening torque $M_0 = 72$ N·m. The experimental results are the calculated median value of the repeated test data. It is worth mentioning that the PTFE coated fastener begins slipping when the tightening torque is 60 N·m. So the initial tightening torque is selected as 60 N·m for the PTFE coated fastener.

Figure 3 illustrates that applying different coatings result in quite different anti-loosening performances under the same initial tightening torque. As shown in
Fig. 3 Self-loosening experimental results of three coatings under the same initial tightening torque: (a) self-loosening curves, (b) loosening degree. ($f = 10 \text{ Hz}$, $M_0 = 72 \text{ N\cdot m}$, $N = 2 \times 10^5$)

Fig. 3(a), virtually no change in the clamping force of the PTFE coated fasteners is reported, and the loosening degree of PTFE coated fasteners is close to zero (Fig. 3(b)). The clamping force of MoS$_2$ coated fasteners does not obviously decline; with an average loosening degree of only 8.08%, this is considered very slight compared to the 70% decrease in uncoated fasteners. More importantly, adding PTFE and MoS$_2$ coated fasteners can reduce the scatter of the loosening degree test data, thereby improving the reliability of bolted structures. Additionally, the coatings allow reducing the variance of loosening degree by nearly 85% compared to uncoated fasteners. These results prove that the PTFE and MoS$_2$ coatings have significant anti-loosening performance.

On the other side, an opposite result is achieved from testing the TiN coating on fasteners. The preload is only 12 kN when the initial tightening torque is 72 N\cdot m. Although it is shown in Fig. 3(b) that the obtained loosening degree of TiN coated fasteners is around 80%, however, the actual observations during the laboratory experiments indicate a higher loosening degree than 80%. It is observed that when the clamping force decreased to 4 kN, the dynamic shear load can no longer be applied and loosening tests were aborted due to the excessive displacement between the upper fixture and lower fixture. It can be concluded that the TiN coating allows complete loosening under the initial tightening torque ($M_0 = 72 \text{ N\cdot m}$). Therefore, a fastener with TiN coating has an insignificant effect on anti-loosening performance under this initial tightening torque.

The anti-loosening performance among the different coatings under the same initial tightening torque is very obvious, which is attributed to the difference in the tightening characteristic of various coatings. It is assumed that the threads friction coefficient $\mu_t$ equals to the friction coefficient between the bolt (or nut) and the bearing surface $\mu_b$ under various experimental parameters. The tightening characteristics of various coatings can be theoretically calculated using Eqs. (2) and (3) [36, 37].

$$M_0 = KP_0d$$  \hspace{1cm} (2)

$$M_0 = P_0\left(\frac{P}{2\pi} + \frac{\mu_t r_t}{\cos \beta} + \mu_b r_b\right)$$  \hspace{1cm} (3)

where $M_0$ is the tightening torque applied to the bolt head/nut; $P_0$ is the preload; $K$ is the nut factor; $P$ is the thread pitch; $\beta$ is half of the thread flank angle; $d$ is the basic pitch diameter of thread; $\mu_t$ is the friction coefficient between threads; $\mu_b$ is the friction coefficient between the bolt (or nut) and the bearing surface; $r_t$ is the effective contact radius between threads; $r_b$ is the effective bearing radius of the bearing contact area under the turning head or nut. The dimensions of the bolts used in this series of tests are listed in Table 4.

Figure 4 shows the tightening characteristics of various coatings. It is observed from Fig. 4 that coatings have a significant effect on the nut factor and friction coefficients in threaded fasteners. PTFE coating and MoS$_2$ coating can effectively reduce the nut factor and friction coefficients which decreases by nearly 50% compared to uncoated fasteners. Thus, as shown in
Fig. 5, under the same initial tightening torque, the initial preload of fasteners with PTFE coating or MoS\textsubscript{2} coating is approximately double that of uncoated fasteners; and applying a higher preload leads to reduced loosening or no loosening [8, 15, 20]. While the initial preload of the TiN coating fasteners is only half of that of uncoated fasteners. Under a lower preload, the friction force decreases and the shear force easily overcomes the friction force and the occurrence of slippage increases, causing the rapid decrease of the clamping force and even complete looseness of fasteners (Fig. 3(a)).

Figure 6 shows the experimental results of self-loosening performance of the three coated fasteners under the same preload $P_0 = 20$ kN. It should be noted that the nut factor of TiN coating is so high that the fasteners are distorted before the preload reaches 20 kN. In several experiments, the preload can only be applied to approximately 19 kN. Thus, the initial preload of 19 kN is selected for TiN coated fasteners.

Figure 6(a) shows the clamping force reduction in relation to the number of cycles. As shown in Fig. 6(a), it is experimentally observed that a typical self-loosening process due to dynamic shear load can be divided into three distinct stages including a stable stage, a rapid declining stage, and a slow declining stage.

1. The first stage: at the beginning of the experiment up to around the 50\textsuperscript{th} cycle, the clamping force is nearly constant. The duration of this stage is not long (only about 5 s), as it takes a few seconds for the testing machine to achieve 6 kN (the amplitude of the dynamic shear load). From the beginning of the experiment to obtaining the amplitude of shear load takes a short time which corresponds to this stage.

2. The second stage: from 50\textsuperscript{th} cycle to about 2,000\textsuperscript{th} cycle, the clamping force decreases significantly. This stage corresponds to the second stage of the self-loosening process of bolted joints subjected to a transverse displacement investigated by Jiang and
co-workers [14–16]. Liu et al. [6] and Yu et al. [38] found that the self-loosening process of bolted joints under axial excitation also experienced this stage. Previous research showed that the obvious backing-off of the nut [14, 15] and the removal of asperities on contact surfaces in this stage led to reduction of the clamping force. In addition, the dynamic shear load generates slippage between the two clamped plates which reduced the clamping force.

(3) The third stage: from 2,000th cycles to 200,000th cycles, the clamping force slowly decreases. This may be because of the occurrence of localized slip, so-called partial slip in fretting, at the thread and head contact surfaces under dynamic shear load [10, 11]. Fretting leads to a stable friction coefficient and wear, thereby the clamping force declines slowly.

Figures 6(b) and 6(c) show the average and variance of the loosening degree and the initial tightening torque. It can be noticed in Fig. 6(b) that applying different coatings result in very different anti-loosening performance under the same preload. In terms of loosening degree, PTFE has the best anti-loosening performance while TiN is the worst. The reported average of loosening degree of PTFE and MoS2 coated fasteners is respectively 14.17% and 19.20%, which is nearly 40% lower than that of uncoated fasteners. While the average of loosening degree of TiN coated fasteners is 22.16% which is slightly decreased compared to uncoated fasteners. More importantly, adding PTFE and MoS2 coating can reduce the scatter of the loosening degree test data. The variance of loosening degree of PTFE and MoS2 coated fasteners decreases nearly 30% compared to uncoated fasteners.

Considering the difficulty of generating the same preload \( P_0 = 20 \text{kN} \), as shown in Fig. 6(c), the preload of the PTFE coated fasteners can be obtained at a low initial tightening torque level, whereas the TiN coated fasteners cannot obtain the same preload at such a low level of initial tightening torque.

Therefore, applying PTFE or MoS2 coating not only reduces the loosening degree but also reduces the scatter of test data of the loosening degree thereby improving the reliability of bolted structures. This may be because adding PTFE or MoS2 coating causes the friction coefficients and asperity to decrease, thereby the contact area increases and the plastic deformation decreases under the same dynamic shear load. Thus, the residual axial clamping force increases and the loosening degree decreases. Previous studies in the field have shown that a steel surface coated with a PVD TiN coating has a higher roughness [34]. Hence, the contact areas of TiN coated fasteners are relatively small; thereby, the plastic deformation increases under the same dynamic shear load. In addition, the nut factor of TiN coating is so high that it needs to be applied with a relatively high initial tightening torque in order to generate the required preload, which results in increasing the plastic deformation of the threaded surface. Thus, the residual axial clamping force decreases while the loosening degree increases.

Furthermore, when comparing Fig. 3(b) and Fig. 6(b), it can be seen that the anti-loosening performance is more significant under a higher initial tightening torque. When the initial tightening torque is 60 N·m, as shown in Fig. 3(b), the loosening degree of the
PTFE coated fasteners is close to zero which indicates that the anti-loosening performance of PTFE coating is extremely significant. As shown in Fig. 6(b), when the initial tightening torque decreases to 30 N·m, an increased loosening degree is reported at 14.17%, which indicates that the anti-loosening performance decreases under a lower initial tightening torque. When the initial tightening torque \( (M_0) \) is 72 N·m, the degree of loosening of MoS\(_2\) coated fasteners is only 8.08% (Fig. 3(b)). However, when the initial tightening torque decreases to 35 N·m, the degree of loosening is reported at 19.20%, more than double that reported under a higher tightening torque. TiN coated fasteners exhibit complete loosening under the lower initial tightening torque \( (M_0 = 72 \text{ N·m}) \). When the initial tightening torque increases to 110 N·m, the loosening degree of TiN coated fasteners is reported at 22.16%, which indicates that adding TiN coating on fasteners has a certain anti-loosening performance under higher initial tightening torque.

3.2 SEM investigation

It is reported that the distribution of axial load along a fastener is highly uneven and that the first three threads bear about 70% of the total axial load while the first thread carries more than 30% [39, 40]. Past research indicated that the degree of thread surface damage decreased with the increase of working thread ring [38, 41]. Thus, the damage of the first thread is analyzed using SEM and EDX.

Figure 7 shows the SEM morphologies and EDX patterns corresponding to the wear scar of the first thread surface of the uncoated fasteners. As shown in Fig. 7, it can be seen that the regions near the top of the thread show serious furrow while other regions show significant delamination. The EDX analyses of wear scars show that point B presents a higher O-element peak than that of point A. Therefore, the main wear mechanism of the uncoated fasteners is abrasive wear, delamination, and a slight oxidation wear which are typical characteristics of fretting [42].

Figure 8 shows the SEM morphologies and EDX patterns corresponding to wear scars of the first thread surface of the PTFE coated fasteners under different initial tightening torques.

As can be seen in Fig. 8, when the initial tightening torque decreases, the damage of PTFE coated fasteners exacerbates. While Fig. 8(a) shows that under a higher initial tightening torque, the damage of PTFE coated fasteners is reported as slight abrasion and exfoliation in the edge regions. When the initial tightening torque decreases to 30 N·m, serious wear is observed, as shown in Fig. 8(b). Nearly half of the coating may be removed. Seriously damaged areas are observed on the thread edge and it is found that these regions show serious furrow and significant delamination. The EDX analyses of wear scars show that a small amount of O-element is present at point A and point B. This may be due to the slight oxidation reaction occurs at the threaded surface before the application of coating. It may also be caused by the oxidation wear during experiments. Therefore, the main wear mechanisms of PTFE coated fasteners are fatigue wear, abrasive wear, and slight oxidation wear, which are in accordance with the typical characteristics of fretting that means fretting wear occurs on the bolt.
contact surfaces. Fretting exacerbates damage and leads to a higher loosening degree which indicates that loosening of bolts is closely related to fretting.

Figure 9 shows the SEM morphologies and EDX patterns are corresponding to the wear scar of the first thread surface of the MoS$_2$ coated fasteners under different initial tightening torques. Similar to PTFE coated fasteners, with the decrease of initial tightening torque, the damage of MoS$_2$ coated fasteners also exacerbates. However, Fig. 9(a) shows that under a higher initial tightening torque, slight abrasion is observed in the MoS$_2$ coated fasteners, while small areas exhibit exfoliation, mainly at the edge. When the initial tightening torque decreases to 35 N·m, discontinuous damage of MoS$_2$ coated fasteners is reported, as shown in Fig. 9(b). In the parts, denoted as area (I), the material components of point (A) are analyzed. It is found that the region contains only iron and carbon which suggests that coating of the bolt has been completely removed. The damaged surface in this area shows obvious plastic furrow and delamination phenomenon. In other parts, denoted as area (II), the coating displays relatively minor abrasion. The EDX analyses of wear scars show that point (B) presents Mo-element and S-element which indicates that MoS$_2$ coating has a good self-lubricating effect. Meanwhile, a small amount of O-element is present at point (B). This may be due to oxidation wear during the experiments. Therefore, the main wear mechanism of MoS$_2$ coated fasteners is fatigue wear, abrasive wear, oxidation wear, and delamination which are in accordance with the typical characteristics of
fretting. Fretting exacerbates damage and leads to a higher loosening degree which indicates that the loosening of bolts is closely related to fretting.

With regard to TiN coatings, it is reported that some TiN coated bolts are distorted during the test. The reason of distortion of TiN coated bolts is that the nut factor of TiN coated fasteners is so high that in order to achieve the same preload, the required tightening torque needs to be higher than that for uncoated fasteners. As can be seen in Fig. 6, an approximate preload of only 19 kN can be applied to the TiN coated fasteners, while the initial tightening torque is approximately 110 N·m. Such high tightening torque may result in a greater frictional stress between the thread contact surfaces. In fact, the bolts undergo serious plastic deformation under such loading condition or may even fracture. It is observed that some TiN coated fasteners are difficult to unscrew after loosening testing because of the distortions or fractures (Fig. 10). Therefore, adding TiN coating on fasteners gains a certain anti-loosening performance under the same preload (Fig. 6), however it is not an ideal method to prevent loosening.

3.3 Effects of loading form

Many studies show that loosening is more severe when a joint is subjected to transverse alternating loading than axial alternating loading [7, 17]. However, the anti-loosening performance of coated fasteners under the two alternating loading forms has not been reported in the open literature. The authors’ research team systematically studied the dynamic behavior of
coated fasteners under axial excitation using experimental and numerical methods [43]. The aim of this work is to explore the differences in the anti-loosening performances of three coated fasteners under dynamic shear load and axial load. Figure 11 shows the loosening degrees of three coated fasteners under the two different loading forms.

It is obvious in Fig. 11 that different load forms result in different anti-loosening performance of coatings. On one hand, when coated fasteners are subjected to shear load, three kinds of coatings demonstrate obvious anti-loosening performances. PTFE coating has the best anti-loosening performance while TiN coating shows the worst. On the other hand, when the coated fasteners are subjected to axial loading, the anti-loosening performance of MoS\(_2\) coating is somewhat encouraging, while TiN coating is still the worst with its loosening degree being reported even higher than those of uncoated fasteners. PTFE coating possesses relatively poor anti-loosening capability compared to uncoated fasteners.

In addition, as shown in Fig. 11, it can also be concluded that the loosening degree is more severe when the fasteners are subjected to shear load than axial load, especially for uncoated and MoS\(_2\) coated fasteners. This is possibly because the shear load is acting in the same direction as the relative slip, and thus is more likely to cause relative slip, which may lead to fretting damage.

In actual situations, fasteners are likely to withstand both axial and shear loads. In order to investigate the effect of the direction of the externally applied load, Zhang et al. [17] designed an experimental fixture which could change the angle between the directions of the applied force in relation to the contact surface of the two clamped plates. An angle of 0° represents the pure shear or transverse loading while an angle of 90° represents axial loading. A 30° angle applies to both axial and shear loads. The results suggest that an axial load component greatly enhances the resistance of the bolted joints to self-loosening. However, their work shows that fatigue dominates the failure under axial load component. Hence, fatigue strength may become a great concern when there is a large axial load component.

In summary, compared to fasteners subjected to shear load, when fasteners withstand both axial and
shear loads at the same time, the risk of loosening may be reduced; however, fatigue risk is possible. Such a phenomenon is worth further investigation in future research.

4 Conclusions

Based on the results obtained by a series of experiments on fasteners with three different kinds of coatings under different initial tightening torques and dynamic load forms, the following conclusions can be drawn.

(1) PTFE coating and MoS₂ coating have significant anti-loosening capabilities which are further enhanced under higher initial tightening torques. However, the anti-loosening performance of the TiN coating is not entirely satisfactory.

(2) The loosening curve can generally be divided into three stages: a stable stage, a rapidly declining stage, and a slowly declining stage.

(3) The loosening mechanism of the coated fasteners is related to fretting under dynamic shear load. Fretting aggravates wear which can readily cause loosening.

(4) The anti-loosening performance of the coatings exhibits obvious differences under shear load and axial load, respectively. The loosening degree is more severe when the fasteners are subjected to shear load than axial load, especially uncoated fasteners and MoS₂ coated fasteners.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the National Science Funds for Distinguished Young Scholars (No. 51025519), the Changjiang Scholarships and Innovation Team Development Plan (No. IRT1178), and the Self-Topic Fund of Traction Power State Key Laboratory (No. 2016TPL-Z03).

Open Access: The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References


Junbo ZHOU. He received his master degree in mechanical engineering in 2009 from Wuhan University of Technology, Hubei, China. After then, he worked in Chengdu University of Technology, Sichuang, China. In 2011, he was a PhD student in the Southwest Jiaotong University. His research interests include fretting wear and design of fastener connection.

Jianhua LIU. He received his bachelor degree and PhD degree from Southwest Jiaotong University, Sichuang, China, in 2010 and 2016, respectively. He is currently a post-doctor at School of Mechanics and Engineering, Southwest Jiaotong University. His research interests include friction, wear and design of fastener connection.

Huajiang OUYANG. He received his bachelor degree, master degree and PhD degree from Dalian University of Technology, China, in 1982, 1985 and 1989, respectively. He is a full professor in the School of Engineering, University of Liverpool, UK. His main research areas are structural dynamics and control (particularly on friction-induced vibration and noise, and moving-load dynamics), and structural identification.

Zhenbing CAI. He received his bachelor degree and PhD degree from Southwest Jiaotong University, Sichuang, China, in 2003 and 2009, respectively. He is currently a professor at School of Mechanical Engineering, Southwest Jiaotong University, deputy director of Tribology Research Institute. His research interests include friction and wear, surface engineering and instrument development.

Jinfang PENG. He received his bachelor degree and PhD degree from Southwest Jiaotong University, Sichuang, China, in 2006 and 2012, respectively. After then, he worked in the same university and became an assistant researcher. His research interests include fretting fatigue and failure analysis.

Minhao ZHU. He received his bachelor degree, master degree from Southwest Jiaotong University, Sichuang, China, in 1990 and 1993, respectively. Then he worked in Southwest Jiaotong University until now. From 1996 to 2001, he received his PhD degree from Southwest Jiaotong University, Sichuang, China. His current position is a Professor, Ph.D. student advisors, Cheung Kong Scholars and dean of School of Materials Science and Engineering. His research areas cover fretting wear, fretting fatigue, surface engineering and design of fastener connection.
Chaotic characteristics and attractor evolution of friction noise during friction process

Cong DING\textsuperscript{1}, Hua ZHU\textsuperscript{1,*}, Guodong SUN\textsuperscript{1}, Yuankai ZHOU\textsuperscript{2,3}, Xue ZUO\textsuperscript{1}

\textsuperscript{1} School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China
\textsuperscript{2} School of Mechanical Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China
\textsuperscript{3} Jiangsu Provincial Key Laboratory of Advanced Manufacturing for Marine Mechanical Equipment, Jiangsu University of Science and Technology, Zhenjiang 212003, China

Received: 01 December 2016 / Revised: 07 March 2017 / Accepted: 27 March 2017
© The author(s) 2017. This article is published with open access at Springerlink.com

Abstract: Friction experiments are conducted on a ring-on-disk tribometer, and friction noise produced during the friction process is extracted by a microphone. The phase trajectory and chaotic parameters of friction noise are obtained by phase-space reconstruction, and its attractor evolution is analyzed. The results indicate that the friction noise is chaotic because the largest Lyapunov exponent is positive. The phase trajectory of the friction noise follows a “convergence-stability-divergence” pattern during the friction process. The friction noise attractor begins forming in the running-in process, and the correlation dimension $D$ increases gradually. In the stable process, the attractor remains steady, and $D$ is stable. In the last step of the process, the attractor gradually disappears, and $D$ decreases. The friction noise attractor is a chaotic attractor. Knowledge of the dynamic evolution of this attractor can help identify wear state changes from the running-in process to the steady and increasing friction processes.

Keywords: friction noise; phase trajectory; chaotic parameters; Lyapunov exponent; chaotic attractor

1 Introduction

Relative motion between two bodies causes friction and contact surface wear, which may exert negative effects on the reliability, security and usage of equipment \cite{1, 2}. Distinguishing and predicting wear states is thus important to extend the life of a mechanical system. However, identifying the wear states of complicated nonlinear dynamic processes is difficult.

In recent years, many domestic and foreign researchers have published studies on the mechanisms and factors affecting friction noise \cite{3}. Chen et al. \cite{4}, for example, performed measurements of the noise induced by friction, and used the waveform and power spectrum to analyze the friction noise sound pressure level. The results of this indicated that friction-induced noise was generated by relative sliding friction and vibration motion. Chen and Zhou \cite{5} applied the concept of friction coefficient and observations of scar topography to analyze the mechanism of friction noise. The results of this work showed that fluctuations in friction force were the main mechanism of this noise. Lars and Staffan \cite{6} studied a spiral-shaped modification of the surface topography of a brake disk to reduce noise; results suggested that a spiral pattern could strongly reduce squealing. A basic study of friction noise caused by fretting was conducted by Jibiki et al. \cite{7}, and the results of this research indicated that friction noise was generated by certain cycles of fretting. The sound pressure level increased with increasing fretting stroke and frequency, which is related to the average sliding velocity and the wear loss. Chen et al. \cite{8} studied

* Corresponding author: Hua ZHU, E-mail: zhuhua83591917@163.com
the relationship between friction noise and friction surface characteristics under reciprocating sliding. The group found that the worn morphology with high intensity level noise presented obvious flake fractures, and that these fractures were important causes of friction noise. Le Bot et al. [9] carried out an experiment on friction noise during dry contact under light pressure, and results demonstrated that friction noise increased with sliding speed and roughness increasing. Disk surfaces with different groove structures were investigated on a pad-disk tester by Wang et al. [10, 11], who demonstrated that a disk surface with 45 degree grooves could strongly reduce friction noise.

The main purpose of studying friction signals is to identify wear states in the friction process. However, the friction process is extremely complex, and accurately identifying and predicting wear conditions from previous studies on friction noise is difficult. Therefore, several researchers have sought to determine the nonlinear dynamics of friction signals to recognize wear states. Zhu et al. [12] extracted the friction force and vibration signals from a pin-disk experiment to realize the chaotic characteristics of a tribological system, and utilized the spectrum and fractal dimension methods to study these signals quantitatively. The results of this research showed that friction signals presented chaotic natures, and that the tribological system was a chaotic system. Zhou et al. [13] studied the chaotic characteristics of friction temperature in the friction process and found that the temperature signal featured chaotic characteristics that could be utilized to recognize changes in wear states. Sun et al. [14, 15] applied chaos theory to study the friction vibration signals extracted from a pin-disk test, and results revealed that these signals demonstrated chaotic characteristics. Wear states can be identified by the evolution of friction vibration attractors. Liu et al. [16] conducted spherical-on-disk running-in tests, and chaotic attractors were used to analyze cross correlations between tangential and normal vibrations. The results of this work demonstrated that the chaotic attractors of vibration signals converged as running-in continued and could be utilized to describe changes in wear states. Oberst and Lai [17] utilized a recurrence plot to reveal the chaotic characteristics of brake squeal noise.

Friction noise contains information reflecting wear states, and friction noise signals can be collected in real time without affecting the normal friction process. Therefore, nonlinear dynamics theory may be utilized to prove that friction noise presents chaotic characteristics and that chaotic attractors undergo dynamic evolution. Friction noise thus presents a new route through which wear states can be monitored on-line, identified and predicted.

Based on the rotational movement of a ring-disk under oil lubrication, the aim of this paper is to illustrate the complex chaotic characteristics and attractor evolution of friction noise signals, knowledge of which is instructive in revealing wear states. In Section 2, the experimental apparatus and specimens are introduced, the tests are conducted, and the original time series of the friction noise signals are obtained. In Section 3, the processed time series of the friction noise signals are obtained by means of reducing sampling and de-nosing. In Section 4, the evolution of the phase trajectories of friction signals is presented by reconstructing the phase space. In Section 5, the correlation dimension $D$ and largest Lyapunov exponent are calculated. Finally, in Section 6, a discussion and analysis are provided, and the main conclusions of this work are given.

## 2 Experiments

### 2.1 Tribometer description

The friction experiments were carried out on a rotating tribometer. Friction noise signals can be extracted in real-time via the ring-disk rotating process. The experimental device is composed of a power system, a loading system, a clamp system, and a data acquisition and analysis system. The equipment is shown in Fig. 1. The upper sample ring was installed on a ring holder by a locking pin and a locking bolt, while the lower sample disk was mounted on a disk holder by a locating pin illustrated in Fig. 2. The friction torque was measured by a torque sensor, which was also attached to the disk holder. Friction noise signals were measured by a microphone PCB fixed 35 mm from the center of the friction pairs. The sensitivity of the microphone PCB was 45 mV/Pa, and its dynamic range was 15-122 dB. Signals were extracted and stored by a CoCo80 analyzer.
2.2 Test samples and test conditions

The test samples included ring and disk specimens. The upper sample was a ring made of GCr15 bearing steel and with a Rockwell hardness of 61 HRC after quenching treatment. Its outside diameter was 34 mm, and its inner diameter was 24 mm. The surface of this ring was first processed by turning, and then ground and polished using sandpapers of 800#, 1200#, 1500#, and 2000# in sequence. The roughness Ra of this specimen was measured, and the mean of three measurements was considered the final Ra of the specimen. The ring specimens showed a roughness of Ra=0.040−0.043 μm. The lower sample was a disk made of AISI 1045; its Rockwell hardness was 44 HRC without heat treatment and its diameter was 46 mm. The disk had an initial roughness of Ra=5.520−6.000 μm after turning. The equivalent radius of the friction pair was 14.5 mm, and the nominal contact area was 455.53 mm². Table 1 shows the experimental conditions of five tests.

The tests were conducted under atmospheric conditions (at 16–22 °C; relative humidity, 48%−58%). Prior to testing, the specimens were cleaned with ethanol (97% pure) using an ultrasonic cleaner. A volume of 0.2 ml of 15W-40 lubricating oil was dropped onto the working surface of the lower sample, and good contact of the specimen surfaces was ensured.
during installation. The sampling frequencies of the sound level sensor and the friction torque were 12.8 kHz and 300 Hz, respectively, and the background noise was measured before the tests. To ensure the repeatability of the test results, five tests were carried out at least three times to account for the randomness of the friction noise.

2.3 Test results

Figure 3 shows the original time series of friction noise signals from five tests. Here, the noise signals are presented as voltage values. The values consistently reflect minimal changes over time but increase sharply in the last process of each test.

3 Experimental signal processing

The voltage signals in the experiments were converted to sound pressure levels in decibels using the equation

\[ P = \frac{U}{S_{0}}, \quad \rho_{0} = \frac{U_{0}}{S_{0}} \]  

(1)

![Fig. 3  Friction noise signals in the friction process.](https://mc03.manuscriptcentral.com/friction)
where, \( U \) is the voltage of the test data, \( U_0 \) is the voltage of the background noise, \( P \) is the sound pressure level of the test data after conversion, \( P_0 \) is the sound pressure level of the background noise after conversion, and \( s_y \) is the sensitivity of the microphone. Here, \( s_y = 45 \text{ mV/Pa} \).

\[
L_P = 20 \log_{10} \left( \frac{P}{P_c} \right), \quad L_{P_0} = 20 \log_{10} \left( \frac{P_0}{P_c} \right)
\]

where \( P_c \) is the reference value of sound pressure, \( P_c = 2 \times 10^{-5} \text{ Pa} \), \( L_P \) is the sound pressure level of the test data, and \( L_{P_0} \) is the sound pressure level of the background noise. The equation used to calculate the sound pressure level after removal of the background noise is

\[
L_{P_1} = 10 \log_{10} \left( \text{abs}(10^{0.1 L_P} - 10^{0.1 L_{P_0}}) \right)
\]

where \( L_{P_1} \) is the sound pressure level of the friction noise without the background noise.

Generally speaking, the friction noise of the friction system includes other insignificant noise that may influence the dynamic characteristics of the former. Therefore, filtering and denoising of the friction noise signals was conducted by empirical mode decomposition (EMD).

EMD was put forward by Huang et al. [18] to filter nonlinear and unstable time series. Using EMD, time series can be decomposed into a set of finite and intrinsic mode functions (imf) with different components. Then, specific imf components selected from the original components are reset to a new time series to analyze and calculate the data during follow-up work [19, 20]. Details of the EMD algorithm are as provided as follows.

The maximum and minimum values of a time series \( x(t) \) are determined, and the upper envelope curve \( x_{\text{max}}(t) \) and lower envelope curve \( x_{\text{min}}(t) \) of the original time series are calculated by the cubic spline functions. The mean of the upper and lower envelope curves is denoted as \( m(t) \). The original time series \( x(t) \) minus the mean \( m(t) \) yields a new series \( h(t) \) without low-frequency components. The relevant functions are shown in Eqs. (4) and (5).

\[
m(t) = \frac{x_{\text{max}}(t) + x_{\text{min}}(t)}{2}
\]

\[
h(t) = x(t) - m(t)
\]

In general, the first \( h(t) \) is not necessarily a stationary data series. Therefore, it should be calculated using Eqs. (4) and (5) repeatedly until it meets the criteria of Eq. (6) with a standard value of SD=0.2–0.3. imf\(_i\) = h(t), where imf\(_i\) is the first component of imf.

\[
SD = \sum \left[ \frac{\text{abs}(h_{i-1}(t) - h_i(t))^2}{h_{\text{max}}^2(t)} \right]
\]

\[
r_i = x(t) - \text{imf}_i
\]

Assuming \( r_i \) is the new \( x(t) \), imf\(_1\), imf\(_2\), ..., and imf\(_n\) are calculated successively until the last time series \( r_n \) remains undecomposed. Thus, the original time series can be presented in terms of imf and a residual component \( r \), as shown in Eq. (8).

\[
x(t) = \sum_{i=1}^{n} \text{imf}_i + r_n
\]

where \( r_n \) is the residual component representing the trend or mean of the original time series \( x(t) \).

The imf component is an oscillating function with different amplitudes and frequencies. Each imf component presents two characteristics [21]. In the data domain of each imf component, the numbers of the maximum values must be equal to the number of the zero crossings or present a difference of one at most. The average value of the envelope curves must always be equal to zero.

Taking the friction noise signal in test 4 as an example, the original signal is decomposed into 11 different imf components and a single \( r \). Figure 4 shows the time- and frequency-domains of the original signal, imf\(_1\), imf\(_2\), imf\(_3\), \( r \), and the reconstructed signal. Several peaks in the power spectra of the original signal and imf\(_1\) and imf\(_2\) components may be observed. By contrast, the power spectrum of imf\(_3\) is relatively smooth. Therefore, the new signal is reconstructed by components from imf\(_3\) to imf\(_{11}\) and \( r \). The power spectrum of the reconstructed signal is smoother than that of the original signal.

Figure 5 displays the time series of friction noise obtained by EMD. Although, some differences may be observed in the test data of different experiments, the data follow a similar variation. The friction noise
Fig. 4  IMF components in time-domain and frequency-domain.

Fig. 5  Friction force and friction noise signals in the rotating friction process.
increases in the running-in process, remains relatively stable in the steady-state process, and then increases sharply in the increasing friction process. At this point, the specimens were damaged and the tests were ended.

4 Evolution of phase-space trajectories and attractor

4.1 Phase-space reconstruction

The phase space is a geometric space that reveals the states of a system. In general, a nonlinear dynamic system presents a very high phase-space dimension. However, in practice, data from the tests are considered to present a single-variable time series obtained from the interaction of different parameters in the system. Thus, the test data should be reconstructed into a high-dimensional space to gain more dynamic information. Takens [22] presented a method wherein a 1D chaotic time series is extended into a 3D or higher-dimensional phase space by phase space reconstruction. The aim of this method is to expose more information on the system hidden in the time series.

Time difference method is usually utilized to reconstruct the phase space for 1D time series of \( x_1, x_2, x_3, \ldots, x_n \). A number selected from the original time series is as one of the components of the vector every \( \tau \) times to construct a group of vectors, i.e,

\[
X_i = [x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}], \quad i = 1, 2, \ldots, N \quad (N = n-(m-1)\tau)
\]

where \( m \) is the embedding dimension (the dimension of the reconstructed phase space), \( X_i \) is the vector of the reconstructed phase space, \( \tau \) is the delay time, \( n \) is the length of the original time series, and \( N \) is the number of vectors in the reconstructed phase space.

Selecting an appropriate \( \tau \) and \( m \) is very important for reconstructing the phase space. When \( \tau \) is too small, \( x(t) \) and \( x(t+\tau) \) cannot be independent of each other because their values are close to each other. When \( \tau \) is too large, the relationship between \( x(t) \) and \( x(t+\tau) \) becomes random, and the chaotic attractor cannot be accurately determined. In this regard, the autocorrelation function method is used to determine an optimal \( \tau \) [23]. For a set of single-variable time series \( \{x(i)\} \), the definition of the autocorrelation function method is

\[
C(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} \left[ x(i+\tau) - \bar{x} \right] \left[ x(i) - \bar{x} \right]
\]

where \( \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x(i) \) (10)

The function constructed from \( \tau \) and \( C(\tau) \) is given by Eq. (10), where \( \tau \) is the optimal delay time when \( C(\tau) \) falls to the value of \((1-1/e) \cdot C(0)\). Taking the friction noise of test 4 as an example, Fig. 6 illustrates the relationship between \( \tau \) and \( C(\tau) \). Here, \( C(\tau) \) falls with increasing \( \tau \). For example, the coordinates of the first point are \((22, 0.6184)\), the \( \tau \) of the time series is 22 s.

The precondition of choosing the \( m \) is \( m \geq 2d + 1 \) (\( d \) is the dimension of the dynamic system). For an infinite time series without noise, \( m \) is just larger than the smallest integer value of \( D \). In an finite time series Fig. 6 Autocorrelation function for the time delay from test 4 \((P = 0.258 \text{ Pa}, \quad v = 0.911 \text{ m/s})\).
with noise, however, $m$ should be much larger than $D$. If $m$ is too small, the attractor could undergo self-intersection because of folding. Thus, selecting a relatively large $m$ is necessary in theory. Unfortunately, increases in $m$ also increase the calculation burden for geometric invariants (e.g., $D$ and the Lyapunov exponent) in practical applications. Moreover, the influences of noise and rounding errors significantly increase.

Therefore, an optimal $m$ is selected by the saturated correlation dimension method [13]. The advantage of this method is that both $D$ and $m$ can be calculated at the same time. The dimension is calculated by means of the G-P algorithm, which was proposed by Grassberger and Procaccia [24]. The formula of the G-P algorithm is

$$C(r) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} H(r - ||X_i - X_j||)$$

(12)

where $H(\cdot)$ is the Heaviside step function. That is

$$H(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

(13)

when $r$ approaches infinity, $D$ is defined as

$$D = \lim_{r \to 0} \frac{\ln C(r)}{\ln r}$$

(14)

Reconstructing the phase space of a 1D time series is necessary prior to this calculation. The $\tau$ can be obtained by the autocorrelation function method described above, and the $\ln C(r)$-$\ln r$ curves are plotted in double-logarithmic coordinates for each $m$. Ideal-linear intervals are then selected from the curves and fitted by the least-squares method [25], the slopes of which reflect $D$. Then, $m$ is plotted as the horizontal coordinate, and $D$ is plotted as the vertical coordinate. The slope values are the $D$ corresponding to the $m$ plotted in the curved figure. Finally, the optimal $m$ and $D$ are obtained when $D$ becomes stable.

Taking the friction noise in test 4 as an example, $D$ and $m$ are calculated by the saturated correlation dimension method. Figure 7(a) shows that double-logarithmic curves are obtained when the integers of $m$ are from 11 to 30. Figure 7(b) displays the relationship between $D$ and $m$ obtained by fitting through the least-squares method. When $m$ is less than 25, $D$ increases with increasing $m$. When $m$ is greater than or equal to 25, $D$ remains stable over a small range. The optimal $m$ and $D$ of the friction noise in test 4 are 25 and 0.8977, respectively. Table 2 shows $\tau$ and the optimal $m$ of friction signals from five tests.

### Table 2 Embedding dimensions and delay times of friction noise signals.

<table>
<thead>
<tr>
<th>Test run</th>
<th>Signal</th>
<th>$\tau$</th>
<th>$m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Noise</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>Noise</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>Noise</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>Noise</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Noise</td>
<td>11</td>
<td>22</td>
</tr>
</tbody>
</table>

![Fig. 7](https://mc03.manuscriptcentral.com/friction)  
Solution of embedding dimension of friction noise from test 4 ($v = 0.911$ m/s, $P = 0.301$ Pa).
### 4.2 Phase-space trajectory and attractor evolution

The \( \tau \) and optimal \( m \) of friction noise signals were calculated by Takens’ theorem [22] in Section 4.1. However, observing the evolution of phase-space trajectory visually in 3D space is impossible because of the high dimension of the evolution process. Therefore, the high-dimensional space is projected onto a 3D space by principal component analysis (PCA; Appendix A) [17, 26]. Three main vectors selected from the reconstructed high-dimensional space are drawn in the 3D graph to observe the evolution of the phase-space trajectory expediently. The reconstructed vector \( \mathbf{X} \) is given in Eq. (9).

The inner product matrix \( \mathbf{Y} \) of the vector matrix \( \mathbf{X} \) is presented as follows:

\[
\mathbf{Y} = \mathbf{X}' \mathbf{X}
\]

(15)

The three main eigenvalues \( \lambda_1, \lambda_2, \) and \( \lambda_3 \) are chosen from all of the eigenvalues of matrix \( \mathbf{Y} \) and then calculated and arranged in descending order as \( \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N \geq 0 \). Then, a new \( N \times 3 \) order vector matrix \( \mathbf{a} \) is obtained by projecting the reconstructed m-dimensional vector matrix \( \mathbf{X} \) to the three matrix directions \( V_1, V_2 \) and \( V_3 \) which are calculated according to \( \lambda_1, \lambda_2 \) and \( \lambda_3 \), respectively.

\[
\mathbf{a} = \begin{bmatrix} U,V,W \end{bmatrix} = \mathbf{X} \begin{bmatrix} V_1,V_2,V_3 \end{bmatrix}
\]

(16)

Each row of matrix \( \mathbf{a} \) represents the coordinate of a point. A total of \( N \) points, which are drawn in 3D coordinates to present the 3D phase-space trajectory, are reconstructed.

Taking the friction noise signals in test 4 as an example, the \( \tau \) and \( m \) of the friction noise are 22 and 25, respectively. The friction noise signal is divided into sections every 20,000 data points, and the points are drawn in 3D space by phase-space reconstruction. All of the phase trajectories are divided by continuous and non-overlapping signals. Because of space limitations, only some figures of the trajectories are selected in this work to present the evolution law of the friction noise signal. The phase trajectories of the friction noise are given in Fig. 8.

During the running-in process of the friction noise signal, the trajectory of the friction noise begins to converge, and the volume of the phase trajectory is very large. The radius of the trajectory is very large in the initial friction stage of 0–15 min, as shown in Fig. 8(a). In the 60–75 min stage (Fig. 8(b)), the curvature radius of the trajectory gradually converges to a central point as the friction process continues. The attractor of the friction noise forms in this stage. In the 75–315 min stage, the trajectory of the friction noise converges to a smaller point, and the trajectory circles reciprocally. The curvature radius remains steady within a small range, as shown in Figs. 8(c)–8(g), and the attractor of the friction noise is stable. In the final process, the phase trajectory of the friction noise begins to diverge, and the curvature radius increases. Thus, the phase trajectory escapes from the space presented in Fig. 8(h), and the attractor of the friction noise disappears. The evolution of the phase-space trajectory of the friction noise can be defined as “convergence-stability-divergence”, which corresponds to the “forming-keeping-disappearing” pattern of the evolution of the friction noise attractor. This pattern also corresponds to the complete friction process of the system. The friction noise first increases, and then declines to a stable value in the 0–75 min stage; this stage is considered the running-in process. The friction noise remains steady in the 75–315 min stage, which is also known as the steady-state process. Finally, the friction noise increases rapidly in the final process. During the complete friction process, the friction noise attractor gradually forms, then remains stable for a long period of time, and then finally disappears.

The evolution of the attractor can be determined from the evolution of the phase trajectories. In the running-in process, the phase trajectory becomes convergent, and the attractor begins to form. During the stable-state process, the trajectory is maintained in a specific space, corresponding to the stable stage of the attractor. Finally, the phase trajectory diverges and escapes from the specific space, which corresponds to the disappearance of the attractor. Thus, the phase trajectory evolution of the friction noise is highly consistent with the evolution of the attractor. The attractor may be considered a running-in attractor because it forms in the running-in process.
5 Chaotic parameters

5.1 Evolution of the correlation dimension

$D$ is a type of fractal dimension and is sensitive to the time behavior of the system. To reflect the characteristics of the chaotic dynamic system, $D$ is usually used to quantitatively describe the complexities of the chaotic signals on a small scale [17]. In the present case, $D$ can be utilized to characterize the complexity of friction noise signals [17, 27].

Coinciding with the time stages of the phase trajectory described in Section 4.2, the time series of the friction noise signal is divided into continuous and non-overlapping windows every 15 min (including 20,000 data points). Figure 9 illustrates the evolution of $D$ of the friction noise in five tests as calculated by the saturated correlation dimension method discussed in Section 4.1.

In the attractor-forming process, $D$ increases from a low value. In the attractor-keeping process, $D$ remains stable. In the attractor-disappearing process, $D$ declines. The evolution curve of $D$ in the friction process corresponds to an “inverted bathtub curve”, and can be generalized as “increasing-steadying-declining” consistent with the “forming-keeping-disappearing” formation process of the friction noise.
attractor and the “running-in, steady process, increasing friction” pattern of the complete wear process.

5.2 Evolution of the Lyapunov exponent

The basic characteristic of chaotic motion is that the system is extremely sensitive to the initial value. The trajectories generated by two initial values that are close to each other separate exponentially with increasing time. This phenomenon can be quantitatively described by the Lyapunov exponent. The Lyapunov exponent represents the average exponential rate of convergence or divergence between adjacent orbits in a phase space. When the Lyapunov exponent of a system is less than zero, the phase volume of the system is contractive in this direction. When the Lyapunov exponent is greater than zero, the phase volume is expansive and folding in this direction. The long-term behavior of an uncertain system is unpredictable. Therefore, the system is chaotic [28].

If a system has a chaotic attractor, it presents three features: (1) There is at least one positive Lyapunov exponent, (2) at least one of the exponents is zero, and (3) the sum of the exponential spectrum is negative.

In practice, the computational burden for all Lyapunov exponents is very large when the dimensions of the system are very high. Thus, determining chaotic characteristics through the largest Lyapunov exponent is appropriate [28].

The exact Lyapunov exponent of a general time series cannot be obtained according to the dynamic equation. The Wolf reconstruction method [28] is thus used to calculate the maximum Lyapunov exponent of such a time series. The calculation process of this measure is as follows:

The $\tau$ and $m$ are respectively obtained by the autocorrelation function and the saturated correlation dimension methods. According to Takens' theorem, the new time series $Y(t) = (x(t), x(t+\tau), \cdots, x(t+(m-1)\tau))$, $(i = 1, 2, \cdots, N)$ is obtained by reconstructing the phase space of the original time series.

Assuming that $Y(t_0)$ is the initial point in the phase space, $Y_0(t_0)$ is the point nearest to $Y(t_0)$, and $L_0$ is the distance between these two points, the time evolution of these two points is tracked from $t_0$ until the distance is larger than $\varepsilon$ at $t_1$.

$$ L_0 = |Y(t_1) - Y_0(t_0)| > \varepsilon, \quad \varepsilon > 0 $$

Fig. 9 Correlation dimensions of friction noise in the wear process.
Then, a point $Y_1(t_1)$ near the point $Y(t_1)$ is determined to ensure that the distance between the two points satisfies the function.

$$L_1 = |Y(t_1) - Y_1(t_1)| \leq \epsilon, \epsilon > 0 \quad (18)$$

To ensure that the angle between $L_1$ and $L_1'$ is as small as possible, the process described above is repeated until the end time of $Y(t)$. Assuming $M$ is the iteration number, the largest Lyapunov exponent [29] is given as Eq. (19).

$$\lambda_1 = \frac{1}{t_M - t_0} \sum_{k=0}^{M} \ln \frac{L_k'}{L_k} \quad (19)$$

Coinciding with the time stages of the phase trajectory in Section 4.2, the time series of the friction noise signal is divided into continuous and non-overlapping windows every 15 min (including 20,000 data points). The largest Lyapunov exponent of every stage is calculated by Eq. (19). Figure 10 illustrates changes in $\lambda_1$ of the friction noise over five tests. Because the $\lambda_1$ of each test is greater than zero, the friction noise signals are chaotic and the attractors of the friction noise signals can be considered chaotic attractors.

6 Conclusions

Experiments with different rotating speeds and loads were performed on a ring-on-disk tribometer, and friction noise signals produced during the friction process were extracted by a microphone PCB. The phase trajectories were projected onto a 3D space by PCA, and $D$ and the largest Lyapunov exponents were calculated based on phase-space reconstruction. The following conclusions were confirmed.

(1) The chaotic attractor is a set of infinite points in phase space. The system state of chaotic motion always converges to a certain attractor in that phase space. Thus, the evolution of the phase trajectory can reflect the evolution of the chaotic attractor. The processes of convergence, stabilization and divergence characterize the evolution of the phase trajectories of friction noise signals, and the evolution of these phase trajectories corresponds to the evolution of the attractors via a pattern called “forming-keeping-disappearing”. The formation process of the attractor can characterize the chaotic behavior of the friction system.

(2) $D$ increases during the running-in process, remains relatively steady during the stable process, and then decreases during the friction increasing

![Fig. 10 Largest Lyapunov exponent of friction noise in the friction process.](image-url)
process. The change process of $D$ conforms to the pattern of an “inverted bathtub curve”. Changes in $D$ are consistent with the evolution processes of chaotic attractors and the phase trajectories.

(3) The attractors of friction noise signals are also called running-in attractors because they form during the running-in process. In addition, the attractors are chaotic at the points where the positive largest Lyapunov exponents are obtained during the friction process.

(4) Friction noise includes information that can be utilized to characterize the dynamic behavior of a friction system. Therefore, the largest Lyapunov exponent can describe the chaotic characteristics of the friction process, and the phase trajectory and $D$ of the friction noise can describe the friction process and evolution of attractors. The results of this study help reveal changes in wear states. As the evolutionary consistency of the chaotic characteristics of friction force and friction noise was not examined in this work, future research on the friction process will include this topic.

**Acknowledgments**

This project is supported by the National Natural Science Foundation of China (Grant No. 51375480), the Graduate Scientific Research Innovation Projects of Jiangsu Higher Education Institutions (Grant No. KYLX16_0527), and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

**Appendix A Proof of concept of principal component analysis (PCA)**

PCA is a statistical process that utilizes an orthogonal transformation to transform a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. It is one of the simplest methods based on true eigenvector multivariate analyses and can reveal the inner structure of data. For a multivariate data set in a high-dimensional space, PCA can provide a lower-dimensional picture or a projection of this object when observed from its most informative viewpoint by applying only the first few principal components so that the dimension of the transformed data is decreased.

Mathematically, the definition of this transformation is a set of $p$-dimensional vectors of $w_{(k)} = (w_1, \ldots, w_p)_{(k)}$ that map each row vector $x_{(i)}$ of $X$ to a new vector of principal component scores $t_{(i)} = (t_1, \ldots, t_m)_{(i)}$ shown by

$$t_{(i)} = x_{(i)} \cdot w_{(k)}, \quad i = 1, \ldots, n, \quad k = 1, \ldots, m \quad (A.1)$$

where $t$ is the maximum possible variance of $x$ and $w$ is a vector.

The full principal component decomposition of $X$ can be given as

$$T = XW \quad (A2)$$

where $W$ is a $p$-by-$p$ matrix with columns representing the eigenvectors of $X^TX$.

Principal component transformation can be performed in accordance with the singular value decomposition (SVD) of $X$

$$X = U\Sigma W^T \quad (A3)$$

where $\Sigma$ represents the singular values of $X$, $U$ represents the left singular vectors of $X$, and $W$ represents the right singular vectors of $X$.

In terms of this factorization, the matrix $X^TX$ is

$$X^TX = W\Sigma U^T U\Sigma W^T = W\Sigma^2 W^T \quad (A4)$$

$$T = XW = U\Sigma W^TW = U\Sigma \quad (A5)$$

Similar to the eigen decomposition, a truncated $n \times L$ score matrix $T_L$ can be gained by the first $L$ largest singular values and their singular vectors:

$$T_L = U_L\Sigma_L = XW_L \quad (A6)$$

Truncation of a matrix $M$ or $T$ by SVD produces a truncated matrix that is the nearest possible matrix of rank $L$.

Therefore, PCA can concentrate most of the signals into the first few principal components, which can be obtained by dimension reduction; later principal components may be affected by noise and disposed of without great loss of information.

**Open Access:** The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License
References


Cong DING. She received her bachelor degree in mechatronic engineering in 2014 from China University of Mining and Technology, Xuzhou, China. After then, she is a Ph.D. student in mechatronic engineering at the same university. Her research interests include friction noise and chaotic theory.

Hua ZHU. He received his Ph.D. degree in mechatronic engineering from China University of Mining and Technology, Xuzhou, China. His current position is a professor, the supervisor of Ph.D. students and the deputy director of the Mine Robot Research Institute. His research areas cover the fractal and chaos theory, tribology theory and application, noise and vibration control and mine robot research.
Thermocapillary migration mechanism of molten silicon droplets on horizontal solid surfaces

Tao SUN¹, Cunhua JIANG¹, Jianning DING¹,²*, Ningyi YUAN¹,²*
¹School of Mechanical Engineering, Changzhou University, Changzhou 213164, China
²Collaborative Innovation Center of Photovoltaic Science and Engineering, the Breeding Construction Point of State Key Laboratory of Photovoltaic Engineering Science, Changzhou 213164, China
Received: 02 March 2017 / Revised: 15 April 2017 / Accepted: 23 April 2017
© The author(s) 2017. This article is published with open access at Springerlink.com

Abstract: Effective lubrication under extreme conditions such as high temperature is of considerable importance to ensure the reliability of a mechanical system. New lubricants that can endure high temperatures should be studied and employed as alternatives to traditional oil-based lubricant. In this paper, a thermocapillary model of a silicone-oil droplet is developed by solving the Navier–Stokes and energy equations to obtain the flow, pressure, and temperature fields. This is accomplished using a conservative microfluidic two-phase flow level set method designed to track the interface between two immiscible fluids. The numerical simulation accuracy is examined by comparing the numerical results with experimental results obtained for a silicone-oil droplet. Hence, the movement and deformation of molten silicon droplets on graphite and corundum are numerically simulated. The results show that a temperature gradient causes a tension gradient on the droplet surface, which in turn creates a thermocapillary vortex. As the vortex develops, the droplet migrates to the low-temperature zone. In the initial stage, the molten silicon droplet on the corundum substrate forms two opposite vortex cells, whereas two pairs of opposite vortices are formed in the silicone fluid on the graphite substrate. Multiple vortex cells gradually develop into a single vortex cell, and the migration velocity tends to be stable. The greater the basal temperature gradient, the stronger the internal thermocapillary convection of the molten silicon droplet has, which yields higher speeds.

Keywords: molten silicon; droplet; temperature gradient; thermocapillary migration

1 Introduction

Effective lubrication is a key factor in ensuring the stability of a high-temperature machine such as a gas turbine or silicon pulling machine [1, 2]. In such machine systems, temperatures T higher than 1,000 °C can develop between the components. In these scenarios, traditional oil-based lubricant is not appropriate for use, because of its disadvantages such as unfavorable viscosity–temperature μ–T characteristics and flammability [3, 4]. Thus, new lubricants suitable for such applications should be found and studied. For example, melttable materials are promising lubricants that can function under the above conditions. Recently, Li et al. [5] exploited the high thermal diffusivity of liquid metal to utilize this material as a lubricant under extreme pressure p.

Thermocapillarity refers to phenomena that arise because of variation in the interfacial tension at a fluid–solid interface, which is caused by a T gradient. Thermocapillarity occurring between a friction pair will affect the lubrication performance by changing the surface tension σ and meniscus shape. Previously, Brozska et al. [6] described the thermocapillary migration behavior of droplets on solid surfaces using a lubrication model, proving that the migration velocity was linearly

* Corresponding authors: Jianning DING, E-mail: dingjn@cczu.edu.cn; Ningyi YUAN, E-mail: nyyuan@cczu.edu.cn
dependent on the radius \( r \) and that a certain critical \( r \) existed. In addition, Brochard [7] assumed an approximately wedge-shaped droplet with a small static contact angle \( \theta \), and found that the droplet motion varies with \( \theta \). Tseng et al. [8] analyzed the change in the flow and \( T \) fields inside the droplet; however, good consistency between experimental and numerical results was not obtained because of differences in the initial and experimental conditions. Finally, Chen et al. [9] studied the effect of contact-angle hysteresis and slip length on the migration velocity during droplet migration, showing that the magnitude of the slip coefficient has little effect on the migration velocity.

A lubrication system will generate heat and, therefore, thermocapillary motion; then, the \( \mu, p \), and capillary motion of the lubricant will change during the droplet migration. Therefore, the thermocapillary migration of molten-material-based lubricant has a significant influence on the lubrication efficacy [10, 11]. However, few works report the flow characteristics of molten materials under high \( T \). Among the investigated materials, silicon has a high melting point and is easily oxidized at high \( T \) values. However, it is difficult to study the thermocapillary migration phenomena of molten silicon droplets via an experimental approach.

In this work, we employ the finite element method with the two-phase level set technique, which is ideally suited to the tracking of interfaces between two immiscible fluids, to examine thermocapillary migration in molten silicon droplets. To verify the accuracy of the conserved level set method, the numerical simulation results obtained for a silicone-oil droplet are compared with experimental results. Hence, the movement and deformation of molten silicon droplets on graphite and corundum substrates are numerically simulated.

Graphite crucibles and corundum crucibles are two commonly used base materials for molten silicon. In this study, high-\( T \) contact-angle equipment is used to measure the \( \theta \) of the silicon liquid on the graphite substrate and corundum, in an argon atmosphere. The evolution of the internal \( T \) field and the flow field of the molten silicon droplets is studied based

---

**List of symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H, W )</td>
<td>Rectangular width and height</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Static contact angle</td>
</tr>
<tr>
<td>( T_{H} )</td>
<td>The temperature of high temperature side</td>
</tr>
<tr>
<td>( x )</td>
<td>Abscissa displacement</td>
</tr>
<tr>
<td>( T_{c} )</td>
<td>The temperature of low temperature</td>
</tr>
<tr>
<td>Bo</td>
<td>Bond</td>
</tr>
<tr>
<td>Ca</td>
<td>Capillary</td>
</tr>
<tr>
<td>Re</td>
<td>Reynolds</td>
</tr>
<tr>
<td>Oh</td>
<td>Ohnesorge numbers</td>
</tr>
<tr>
<td>( \rho )</td>
<td>Density</td>
</tr>
<tr>
<td>( G )</td>
<td>Gravitational acceleration</td>
</tr>
<tr>
<td>( V )</td>
<td>Droplet velocity</td>
</tr>
<tr>
<td>( L )</td>
<td>Characteristic length</td>
</tr>
<tr>
<td>( D )</td>
<td>The droplet contact diameter under the solid surface.</td>
</tr>
<tr>
<td>( k^{-1} )</td>
<td>Capillary length</td>
</tr>
<tr>
<td>( \text{ml}, \text{ua} )</td>
<td>The dynamic viscosity of droplet and air</td>
</tr>
<tr>
<td>( C_{pD}, C_{pa} )</td>
<td>Specific heat of droplet and air</td>
</tr>
<tr>
<td>( k_{D}, k_{a} )</td>
<td>Thermal conductivity of droplet and air</td>
</tr>
<tr>
<td>( F_{x}, F_{y} )</td>
<td>Surface tension forces in the ( x ) and ( y ) directions</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>The surface tension coefficient</td>
</tr>
<tr>
<td>( \sigma_{0} )</td>
<td>Surface tension at the reference temperature ( T_{0} )</td>
</tr>
<tr>
<td>( \Omega_{1} )</td>
<td>Air subdomain</td>
</tr>
<tr>
<td>( \Omega_{2} )</td>
<td>Droplet subdomain</td>
</tr>
<tr>
<td>( \phi )</td>
<td>Level set function</td>
</tr>
<tr>
<td>( \varepsilon )</td>
<td>Interface thickness</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Reinitialization parameter</td>
</tr>
<tr>
<td>( u )</td>
<td>Velocity vector</td>
</tr>
<tr>
<td>( b_{ls} )</td>
<td>Slip length</td>
</tr>
<tr>
<td>( u_{l} )</td>
<td>Velocity of the fluid slip on the boundary</td>
</tr>
<tr>
<td>Ma</td>
<td>Marangoni number</td>
</tr>
<tr>
<td>( \Delta T )</td>
<td>Temperature difference between two contact lines</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>The thermal diffusivity</td>
</tr>
<tr>
<td>( \Delta p )</td>
<td>The droplet pressure difference</td>
</tr>
<tr>
<td>( R )</td>
<td>Droplet interface curvature</td>
</tr>
<tr>
<td>( \sigma_{LG}, \sigma_{SG}, \sigma_{SL} )</td>
<td>Gas–liquid surface tension, solid–air interfacial tension, solid–liquid interfacial tension</td>
</tr>
</tbody>
</table>
on the experimental value of $\theta$. The thermocapillary migration behavior of the silicon droplets on the graphite and corundum substrates under the argon atmosphere is investigated via numerical simulation, which provides a theoretical basis for control of the thermocapillary convection.

2 Physical model and meshing

In the numerical simulation conducted in this study, a small liquid droplet was placed at the bottom solid wall in a rectangular cross-sectional area with dimensions of $1.5 \text{ mm} \times 8 \text{ mm}$ (height $H \times$ width $W$). The liquid droplet was initially assumed to have a cylindrical cap shape, with a static $\theta$ which is defined as the angle formed between the droplet/air interface and the droplet/solid interface at the intersection of droplet-air-solid phase (Fig. 1(a)). A uniform temperature gradient $G$ was imposed on the bottom wall, and the $T$ distribution followed the relation $T = T_{\text{Hi}} - G \times x$. Where $T_{\text{Hi}}$ is the temperature of high temperature side, $x$ is the abscissa displacement. Here, if $x = 0$, $T = T_{\text{Hi}}$ and if $x = W$, $T = T_{\text{C}}$. Here, $T_{\text{C}}$ is the temperature of low temperature side. As the droplet was considered to be very small, the density of the liquid within it could be assumed to be constant and the influence of the body force could be neglected. Note that both the droplet and air can be regarded as incompressible and Newtonian fluids. Further, $\sigma$ and $\theta$ are functions of $T$, whereas the physical properties of the droplet and air are independent of $T$.

A total of 18,912 grids were used for calculation in the subsequent simulation, and grid encryption was used at the interface, as shown in Fig. 1(b). The Bond, capillary, Reynolds, and Ohnesorge numbers, which are expressed as follows, were used to determine the cylindrical cap assumption:

$$
Bo = \frac{\rho g r^2}{\sigma}; \quad Ca = \frac{\mu \cdot v}{\sigma}; \quad Re = \frac{\rho v L}{\mu}; \quad We = \frac{\rho v^2 L}{\sigma};
$$

$$
Oh = \frac{\mu}{\sqrt{\rho \sigma L}} = \sqrt{\frac{We}{Re}}
$$

(1)

For silicone-oil droplets with $r = 0.5 \text{ mm}$, the migration velocity has an order of magnitude of $1 \text{ mm/s}$. We obtained $Bo = 0.115$, $Ca = 0.01$, $Re = 10$, $We = 10^{-5}$, $Oh = 10^{-4}$ where $\rho$ represents the droplet density; $g$ is the gravitational acceleration; $v$ represents the droplet velocity; $L$ is the characteristic length; and $D$ represents the droplet contact diameter under the solid surface.

If the droplet $D$ is smaller than the capillary length $k^{-1}$, $k^{-1} = \sqrt[3]{\frac{\sigma}{\rho g}}$ and the shape of the side surface of the droplet ridge has a circular cross section. However, if $D > k^{-1}$, the shape of the liquid drop ridge is flat.

In the silicone-oil case examined in this study, thermocapillary migration was simulated under the condition that $\theta = 90^\circ$. For the molten silicon on graphite and corundum substrates, the thermocapillary migration of the droplets was simulated under the conditions that $\theta = 45^\circ$ (hydrophilic condition) and $\theta = 90^\circ$.

3 Mathematical formulation and boundary conditions

3.1 Gas–liquid two-phase flow of thermocapillary mathematical model

The two-dimensional equations for the conservation of mass, momentum, and energy for incompressible and Newtonian fluids are expressed as

$$
\rho \left[ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right] = 0
$$

(2)
\[
\rho_i \left[ \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \right] = -\frac{\partial p}{\partial x} + \mu_i \left[ \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right] + F_x,
\]

(3)

\[
\rho_i \left[ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} \right] = -\frac{\partial p}{\partial y} + \mu_i \left[ \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right] + F_y,
\]

(4)

\[
\rho_i C_p \left[ \frac{\partial T}{\partial t} + u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} \right] = k_i \left[ \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right],
\]

(5)

where \( u_i \) and \( v_i \) are the velocity components in the \( x \) and \( z \) directions, respectively; \( \rho_i \) is the fluid density; \( \mu_i \) is the dynamic viscosity; \( C_p \) is the specific heat; and \( k_i \) is the thermal conductivity. The subscripts \( i = l, a \) are used to represent liquid and air, respectively. \( F_x \) and \( F_y \) are the surface tension forces in the \( x \) and \( y \) directions, respectively.

The force generated by the \( \sigma \) differences at the gas–liquid interface is given by the equation

\[
\mu \frac{\partial u}{\partial y} = \gamma T \frac{\partial T}{\partial x}
\]

(6)

Here, \( \mu \) is the dynamic viscosity and \( \gamma T \) is the surface-tension coefficient, which can be assumed to vary linearly with temperature [12] such that

\[
\sigma = \sigma_0 - \gamma T (T - T_0)
\]

(7)

where \( \sigma_0 \) is the surface tension at the reference temperature \( T_0 \). The physical properties of the liquid and gas are listed in Table 1.

### 3.2 Gas–liquid interface tracking method

The conservative level set [13, 14] method is commonly used to manage the deformation of the free interface during droplet motion. In this method, the air \( \Omega_1 \) and droplet \( \Omega_2 \) subdomains are separated by the interface \( \varphi \) with the level set function \( \varphi = 0.5 \). The value of \( \varphi \) smoothly increases from 0 to 1 with \( 0 \leq \varphi < 0.5 \) in the liquid-droplet subdomain, with \( 0.5 < \varphi \leq 1 \) in the air subdomain. The reinitialized convection of the interface can be expressed as

\[
\frac{\partial \varphi}{\partial t} + u \cdot \nabla \varphi = \gamma \nabla \cdot \left( \nabla \varphi \left( \varphi (1 - \varphi) \right) \frac{\nabla \varphi}{|\nabla \varphi|} \right)
\]

(8)

where \( \epsilon \) is the interface thickness, \( \gamma \) is the reinitialization parameter, and \( u \) is the velocity vector. Note that \( \gamma \) determines the numerical calculation of the convergence in this case and should be of the same order of magnitude as the maximum flow rate in the flow field.

#### 3.3 Boundary conditions

The left and right walls are nonslip adiabatic walls. The upper wall is isothermal with no sliding property; the bottom wall is a wetting surface with a uniform \( G \). The Navier slip condition is applied at the liquid–solid boundary, such that

\[
u_s = b_w \frac{\partial u}{\partial y}
\]

(9)

where \( b_w \) is the slip length and \( u_s \) is the velocity of the fluid slip on the boundary.

### 4 Results and discussion

The droplet velocity \( v \) is defined as the velocity of the droplet motion relative to the solid surface. The method of numerical simulation used to calculate \( v \) gives the average velocity, which is obtained by integration of the velocities of each point inside the droplet, followed by division by the volume integral of the entire droplet.

To verify the accuracy of the conservative level set method, the numerical simulation results obtained for a silicone-oil droplet were compared with the experimental results. On the basis of this numerical calculation method, the thermocapillary migration processes of molten silicon droplets on graphite and corundum substrates were studied. The thermocapillary
migration process for droplets primarily includes a dynamic change in the gas–liquid two-phase flow field, a $T$ change of the droplets and the surrounding gas, and deformation of the gas–liquid interface.

4.1 Analysis and validation of silicone-oil droplet migration model

4.1.1 Slip length effect

Figure 2 illustrates the change in the silicone-oil migration rate over time $t$ for slip lengths $b = 1, 3, \text{ and } 5 \text{ nm}$. In the initial stage ($t < 0.5 \text{ s}$), the greater the value of $b$, the greater the droplet migration rate has. When the migration velocity becomes stable, the effect of $b$ on the silicone-oil migration rate decreases. The stable migration velocity of the silicone oil tends to be consistent for all three examined $b$ values. Previous research has shown that the $b$ of this model has little effect on the thermocapillary migration rate [8], and the computational model used to obtain the results presented below was unified for all cases using $b = 1$.

![Fig. 2](image)

Effect of slip length on silicone-oil migration velocity.

4.1.2 Temperature gradient effect

Figure 3(a) illustrates the $G$ effect on the silicone-oil migration, as determined for $G = 1.96, 5, \text{ and } 20 \text{ K/mm}$ (the corresponding Marangoni (Ma) numbers are 28.8, 73.4, and 293.6, respectively), $\theta = 90^\circ$, and the droplet contact diameter $D = 1 \text{ mm}$. It is apparent that the average migration rate changes with $t$. The Ma effect (i.e., the thermocapillary convection intensity) is usually expressed by the Ma numbers [15], where

$$Ma = \frac{\gamma \Delta TL}{\mu \alpha}$$

Here, $\Delta T$ is the temperature difference between two contact lines, $L$ is the contact length between the droplet and bottom wall, and $\alpha$ is the thermal diffusivity. A larger Ma number indicates greater internal thermocapillary convection intensity. As shown in Fig. 3(a), the silicone-oil droplet thermocapillary migration velocity first increases rapidly, before gradually decreasing to a steady process. The greater the value of $G$, the greater the maximum $v$ and the shorter the time taken to reach the maximum $v$; this behavior indicates that the instantaneous acceleration of the droplet migration increases as $G$ increases. This is because the variation in $\sigma$ from lower values on the warm side of the interface to higher values on the cool side generates a flow in the droplet, which exerts a hydrodynamic force on the solid surface that points in the direction of the applied $G$. The solid surface, which is held immobile, exerts an equal and opposite reaction on the droplet that propels it in the opposite direction [16], i.e., toward cooler regions. Under suitable conditions, quasisteady motion, for
which the net hydrodynamic force on the droplet is zero, can be achieved.

Equation (6) shows that the shear stress produced by the thermocapillary effect on the gas–liquid interfacial surface is proportional to \( G \). These trends are consistent with experimental phenomena [17].

In the numerical simulation, the boundary conditions were chosen such that the static \( \theta = 40^\circ \) and the substrate \( G \) was 2 K/mm; these conditions are identical to those reported for a previous experiment conducted by Tseng [8]. Figure 3(b) illustrates the numerical and experimental results for the thermocapillary migration of the silicone-oil droplet. A certain degree of consistency is apparent between the different sets of results, which verifies the accuracy of the conservation level set method. Therefore, the numerical simulation can provide basic and detailed physical quantities for the droplets, such as the flow, \( p \), and \( T \) fields during the migration. However, because the numerical simulation does not consider the effect of buoyancy convection and gravity, certain inconsistencies exist between the numerical and experimental results.

4.1.3 Detailed flow and temperature fields inside silicone-oil droplet

Figure 4 illustrates the variation in the internal \( T \) distribution of the silicone-oil droplet over time, for \( M_a = 146.8, b_s = 1 \) nm, \( \theta = 90^\circ \), \( G = 20 \) K/mm, and \( D = 1 \) mm. The colored lines indicate isotherms, and the \( T \) difference between two adjacent isotherms is 0.8 K. The initial stage of the droplets and the \( T \) of the surrounding gas constitute the reference temperature. With increasing \( t \), heat is transferred from the substrate to the droplets and air. Because the thermal diffusion coefficient of silicone oil is greater than that of air, heat is transmitted faster in the droplets and eventually forms the obtained \( T \) distribution. As shown in Fig. 5 (\( t = 0.01 \) s), in the high-\( T \) region (left side), the inner vortex is greater; this is attributed to the greater \( G \). Two vortices are formed in the gas area around the droplets, and development of this asymmetric vortex pushes the droplet toward the low-\( T \) region.

In contrast to the internal flow field of the droplet shown in Fig. 6, the \( T \) and \( V \) fields are almost decoupled at 0.01 s, and the \( T \) field does not reflect the disturbance.

Fig. 4 Temperature \( T \) changes in silicone oil with time \( t \) at \( \theta = 90^\circ \) and \( G = 20 \) K/mm.
caused by convection. With increasing \( t \), the \( T \) no longer decreases linearly, and Fig. 4 shows that the convective torsion caused by the flow is distorted. Figure 6 demonstrates that with increasing \( t \), the development of the inner turbines on the high-\( T \) side increases. On the low-\( T \) side, however, the final two vortices develop into a vortex and the silicone-oil droplet migration velocity is stable.

Figure 7 illustrates the droplet internal and gas–liquid interface \( T \) distributions at different times. In the initial stage, the highest \( T \) is at the lowest part of the droplet. Then, as the front and rear contact points move, the position of the advancing angle contact point becomes the low-\( T \) point of the free interface. At a certain \( t \), the lowest point of the interface \( T \) is obtained at the forward angle contact point and remains constant.

4.1.4 Deformation and variation of dynamic contact angle during migration process

Figure 8 shows the deformation process for droplets.
under boundary conditions of $\theta = 90^\circ$ and $G = 20$ K/mm. The non-uniform distribution of the gas–liquid interface generates differences in the gas–liquid surface $p$ that deform the droplet.

From the current simulation, $\theta$ increases during the migration process and then stabilizes at a fixed value, while the advancing angle is always greater than the receding angle. It can be seen from Fig. 9 that the advancing and receding angles [18] increase before $t = 0.01$ s, because the droplet $\sigma$ gradient produces shear stress that deforms the interface. Moreover, changes in the interface curvature vary the internal and external differential $p$ values, thereby altering $\theta$. The gas–liquid interfacial deformation yields capillary convection within the droplet. Note that the capillary flow is caused by the different curvatures at the droplet interface. According to the Young–Laplace equation,

$$\Delta p = \frac{2\sigma}{R}$$

where $\Delta p$ represents the droplet pressure difference and $R$ is the droplet interface curvature. The droplet deformation during the migration induces the difference in the interface curvature at the droplet interface. The smaller the curvature, the greater the difference is between the inside and outside $p$ values. Therefore, the capillary flow occurs from the smaller side of the curvature to the larger side, and the droplet interface shape affects the capillary convection direction. Figure 8 demonstrates that the capillary convection direction is opposite and resistant to the thermo-capillary convection (Marangoni convection); this behavior ultimately stabilizes the advancing and receding angles at a constant value.

### 4.2 Numerical simulation of thermocapillary migration of molten silicon droplets

Figure 10 shows the changes in the static $\theta$ values of the silicon liquid on the graphite and corundum surfaces in response to changing $T$. Here, $\theta$ decreases rapidly as $T$ increases from the silicon melting point to 1,700 K.

The Young equation [19] is given by:

$$\sigma_{LG} \times \cos \theta = \sigma_{SG} - \sigma_{SL}$$

Fig. 8 Change of interface shape during silicone oil migration.

Fig. 9 Receding/advancing dynamic contact angles of silicone oil droplet.
Fig. 10 $\theta$ curves of molten silicon on polished graphite and corundum surface (Ar).

This equation illustrates the relationship between the static $\theta$, the gas–liquid surface tension ($\sigma_{LG}$), the solid–air interfacial tension ($\sigma_{SA}$), and the solid–liquid interfacial tension ($\sigma_{SL}$). With increasing $T$, $\sigma$ of the silicon liquid decreases linearly [20] and the liquid surface of the silicon liquid in the argon atmosphere decreases, thereby decreasing $\theta$. The penetration of the graphite interior by the silicon liquid passing through the graphite surface pores is another important factor that contributes to the $\theta$ reduction of the molten silicon on the graphite substrate [21]. This behavior yields greater reduction of the $\theta$ value on the graphite substrate than that on the corundum substrate. Further, at 1,700 K, the molten silicon density changes abnormally. The coefficient of body expansion in this $T$ region is eight times ($8 \times 10^{-4} ^\circ \text{C}^{-1}$) that in the $T$ region beyond 1,700 K, and is obviously caused by the change in $\theta$. Once $T$ reaches and exceeds 1,700 K, $\theta$ remains almost constant.

Figures 11(a) and 11(b) illustrate the variation trend of the average migration velocities of the molten silicon droplets on the corundum and graphite substrates, respectively, for $G = 2$, 5, and 20 K/mm. The initial $T$ is 1,700 K and the static $\theta$ values of the molten silicon on the graphite and corundum substrates are 45° and 90°, respectively. With increasing substrate $G$, the average migration velocity increases and the instantaneous velocity on the graphite substrate ($t = 0.009 \text{ s}$) is as high as 1.8 m/s. Compared with the silicone-oil droplets, the differences in the thermal diffusivity, $\mu$, and $v$ values of the molten silicon droplets are greater.

Figures 12 and 13 illustrate the changes in the internal $T$ and flow fields of the droplets, respectively, during molten silicone droplet migration on the graphite substrate ($G = 20 \text{ K/mm}$). In these figures, the colored lines represent isotherms, and the $T$ difference between each adjacent pair of lines is 6 K. The substrate has a linear $G$; therefore, as the substrate heat is transferred inside the silicon droplets, a $\sigma$ surface tension gradient is generated at the gas–liquid interface that causes the droplets to move from the region with lower $\sigma$ to the region with higher $\sigma$. Further, a shear stress is generated from the high- to low-$T$ regions on the silicon liquid surface, as indicated by Eq. (6). The $\mu$ effect causes the formation of reflux in the silicon liquid.

Figure 13 shows that the thermocapillary convection influences the silicon liquid and forms two vortices with opposite directions. As the vortices change, the clockwise vortex undergoes greater development, which pushes the droplets toward the low-$T$ side. The
development of the vortex affects the distribution of the internal $T$ of the silicon liquid. It can be seen from Fig. 12 that the distribution of the contours reflects the vortex shape, indicating coupling of the $T$ and flow fields. As the interfacial $T$ increases while the $\sigma$ decreases, the droplet has a tendency to spread.

Figures 14 and 15 show the change in the internal $T$ and the flow fields of the molten silicon droplets at $G = 20$ K/mm on the corundum substrate. The $T$ difference between each pair of isotherm lines is 4 K.
Compared with the graphite substrate, the silicon liquid on the corundum substrate produces two vortices during the migration process. As the silicon liquid migrates to the cold side, the two vortices gradually disappear and form a stable laminar flow from the high- to low-T regions.

Fig. 14 Variation of T field of molten silicon droplet on corundum (θ = 90°) substrate during motion process.

Fig. 15 Change of flow field of molten silicon droplet on corundum substrate (θ = 90°) during motion process.
5 Conclusion

The thermocapillary actuation mechanism of a molten silicon droplet in a microchannel was studied via numerical computation, and the results were compared with experimental results. It was found that the variation in the surface tension from lower values on the hot side of the interface to higher values on the cool side generates a flow in the droplet, which exerts a hydrodynamic force on the solid surface in the direction of the applied temperature gradient. The solid surface, which is held immobile, exerts an equal and opposite reaction on the droplet that propels it in the opposite direction, i.e., toward cooler regions. Under suitable conditions, quasisteady motion can be achieved, wherein the net hydrodynamic force on the droplet is zero. Four horizontal vortex arrangements were generated during the migration process of the molten silicon droplet on a graphite substrate: two clockwise and two counterclockwise. During the same migration process, two vortices were formed on a corundum substrate. The vortices disappeared and gradually formed a stable laminar flow, and the droplet velocity became stable.

These results indicate that the actuation behavior of the droplet is strongly influenced by the temperature gradient. A higher temperature gradient generates higher droplet actuation velocity.

Acknowledgements

The authors appreciate the support from the Key Program of the National Natural Science Foundation of China (Grant No. 51335002) and Key support projects of strategic emerging industries in Jiangsu Province (Grant No. 2015-318).

Open Access: The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References


---

**Tao SUN.** He received his bachelor degree in mechanical engineering in 2014 from Changzhou University, Changzhou, China. After then, he was a master of Collaborative Innovation Center of Photovoltaic Science and Engineering at the same university. He will obtain his master degree of engineering in mechanical engineering at Changzhou University in June, 2017. His research interests include manufacture and equipment of new energy materials and devices.

**Jianning DING.** He received his Ph.D. degree in mechanical engineering from Tsinghua University, China, in 2001. He joined Jiangsu University between 1991–2007. He is a research fellow in City University of Hong Kong from 2002 to 2003. His current position is a professor and vice president of Changzhou University. He is the director of Center for low-dimensional materials, micro-nano devices and system, director of Jiangsu Collaborative innovation center of “Photovoltaic science and engineering” in 2011, director of Jiangsu Key Laboratory for Solar Cell Materials and Technology, the head of Science and Technology Innovation Team in Jiangsu Universities, and the chief scientist in the first level of young and middle-aged in “333 project” of Jiangsu province. His research areas cover the research of new energy materials, low-dimensional materials, micro-nano devices and system, and tribology. Above 500 academic papers and 6 books were published.
Synchronous measurement of tribocharge and force at the footpads of freely moving animals

Yi SONG1,2, Zhouyi WANG1,3, Jun ZHOU1, Yang LI1, Zhendong DAI1,3,*

1 Institute of Bio-inspired Structure and Surface Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
2 College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
3 College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Received: 08 January 2017 / Revised: 13 April 2017 / Accepted: 02 May 2017
© The author(s) 2017. This article is published with open access at Springerlink.com

Abstract: Hypothesis on electrostatic attraction mechanisms involving the hairy adhesion of climbing animals has been a matter of controversy for several years. The detection of tribocharge and forces at attachment organs of animals is a practical method of clarifying the dispute with respect to electrostatic attraction in the attachment of animals. Nonetheless, the tribo-electrification is rarely examined in the contact-adhesion of animals (especially in their free and autonomous attachment) due to the lack of available devices. Therefore, the present study involves establishing a method and an apparatus that enables synchronous detection of tribocharge and contact forces to study tribo-electrification in the free locomotion of geckos. A type of a combined sensor unit that consists of a three-dimensional force transducer and a capacitor-based charge probe is used to measure contact forces and tribocharge with a magnitude corresponding to several nano-Coulombs at a footpad of geckos when they climb vertically upward on an acrylic oligomer substrate. The experimental results indicate that tribocharge at the footpads of geckos is related to contact forces and contact areas. The measured charge allows the expectation of an exact attraction with magnitude corresponding to dozens of newtons per square meter and provides a probability of examining tribo-electrification in animal attachment from a macro level.

Keywords: tribocharge; forces; synchronous measurement; animal; free locomotion

1 Introduction

An understanding of adhesion mechanisms in the hierarchical attachment organs of animals is of immense scientific significance and engineering potential [1−5]. However, the adhesion mechanism of animals’ hairy attachment organs continues to constitute a controversial matter. Although extant studies over several years confirmed that van der Waals interaction [6, 7] and capillary forces [8, 9] play dominant roles in setae adhesion, electrostatic attraction is increasingly invoked [10, 11] to interpret the adhesion of hairy systems [9, 10]. It is easy to positively tribo-electrify hair [12, 13], and thus animals’ hairy attachment organs are expected to become charged and subsequently generate electrostatic forces [10, 14] as long as they interact with the environment to obtain forces to drive their locomotion. Unfortunately, there is a paucity of definite evidence to confirm the electrostatic attraction hypothesis in animal dry adhesion (especially involving free movement) due to the lack of feasible equipment.

Measurement of the tribocharge (the charge resulting from contact and friction) and forces at the feet-substrate interfaces is a practical and effective method to investigate tribo-electrification and electrostatic interaction in animal attachments. Additionally, it is important to synchronously implement the measurements to reduce the error caused by charge dissipation and/or neutralization. Although facilities for measuring force [15, 16] and charge [17−21] are
well developed, the synchronous measurement of both remains a challenge. Chiou et al. measured tribo-voltage between two metals through a voltmeter and obtained contact forces by using a lever system [22]. Budakian et al. developed a type of rotary vane apparatus to study tribo-electrification between a sphere and a plate and obtained corresponding forces through a deformed cantilever [23]. Furthermore, a previous study used atomic force microscopes to measure tribocharge and forces between micro/nano objects [24]. However, to the best of the authors’ knowledge, extant studies to date have not applied these techniques with respect to moving animals.

Based on electrostatic induction, Izadi et al. [14] measured tribocharge and shearing forces at the footpads of geckos that were compelled to slide on polytetrafluoroethylene AF (PTFE AF) film and polydimethylsiloxane (PDMS) film. However, in their experiments, animals were only compelled to interact with a substrate instead of attaching autonomously, and this, for one thing, may cause a deviation in the results, and for another, limited the practicability of their apparatus.

Therefore, the present study involved developing a technique and an apparatus with the aim of examining tribo-electrification in the attachment of moving animals. The apparatus allowed the synchronous measurement of tribocharge and forces at the footpads of geckos when they climbed vertically upwards. Electrostatic attraction and its possible contribution were also preliminarily discussed based on the experimental outcomes. The successful synchronous measurement provided a method to conduct an in-depth investigation of tribo-electrification in animal attachment.

2 Materials and methods

2.1 Animals

Four healthy Gekko geckos (Linnaeus) with an average mass of 86.8 ± 7.2 g (mean ± s.d.) that originally inhabited Guangxi province (China) were employed in the experiments. They were raised in a room with a supply of live crickets and under simulated natural conditions including decorations involving rockworks, a water pool, a lighting system (with a cycle of 12 h of light and 12 h of darkness), air conditioning (23 ± 2 °C), and a humidity regulator (60%−70%), before and after the experiments.

2.2 Techniques and apparatus

In summary, the system comprised of five subparts and included sensor units, a rotatable aisle, an image recorder, a data acquisition (DAQ) module, and a computer (PC) as shown in Fig. 1.

Specifically, a sensor unit (as shown in Fig. 1, a) that detected both the tribocharge and forces was embedded in a rotatable aisle (as shown in Fig. 1, b)
that was well-grounded to guide possible charge at the animals’ feet into the earth prior to contact with the probes. The packaged aisle was fully protected by a metal cage (Fig. 1, c) to prevent environmental electromagnetic noises from affecting the test. A high-speed camera (Fig. 1, d) (i-speed 3, Olympus, Japan) was placed in front of the aisle with its lens perpendicular to the aisle to record the animal-substrate interaction throughout the experiment. The charge signals were converted through a digital coulometer (ES111, ESDEMC Technology, Rolla, Missouri, USA, Fig. 1, e) prior to collecting the same with a DAQ system (National Instruments, TX, USA, Fig. 1, f). In contrast, the force signals were collected straight through the DAQ module and were subsequently transmitted to the PC (Fig. 1, g). The video recording and storing were directly completed on the PC.

In hardware terms, the probe simultaneously functioned as a charge sensor and force carrier to ensure synchronous detections of force and charge signals. In software terms, a trigger signal was generated and subsequently divided into three to trigger the subsystems such that the tribocharge, contact forces, and videos were conjointly recorded.

The sensor unit consisted of a three-dimensional force sensor with scale range of 1.5 N and a resolution of 2 mN (as shown in Fig. 1, h) and a capacitor-based charge probe (as shown in Fig. 1, i). In order to accurately measure charge, it is necessary for the capacity of the probe to be sufficiently small in conjunction with a sufficiently large resistance [25], and this results in a large mass of the probe that can dramatically weaken the dynamical performance of the force sensor [26]. As a compromise, a probe was fabricated by laminating copper foils (25 µm in average thickness, Fig. 1, j) on an industrial FR-4 epoxy glass fiber plate (as shown in Fig. 1, k) with a size corresponding to 30 mm × 30 mm × 0.8 mm, a resistance corresponding to $10^{14} \, \Omega$, a capacity corresponding to 70.2 pF (smaller than one thousandth of that of the coulometer), and a mass corresponding to 2.06 g. Additionally, a thin acrylic oligomer film (as shown in Fig. 1, l) was coated at the top surface of the probe as a substrate to interact with the animals’ feet.

Animals’ feet move unpredictably, and thus it is difficult to track the object accurately in real time. Hence, the charge at the substrate that is equal to albeit with an opposite sign to the charge at the objects [12, 27] was alternatively obtained.

The system was calibrated before and after the experiments. The force sensor was calibrated with a dead weight (Fig. 2(a)) [16] while the coulometer was calibrated based on the user’s manual [28] (Fig. 2(b)). If the total capacity of the system corresponded to $C$, the system obtains a total charge of $Q = CV_D$ when being charged by a voltage $V_D$. Thus, it is possible to calibrate the coulometer by comparing the measured charge and the theoretical charge (as shown in Table 1).

### 2.3 Experiments and data process

Prior to the animal experiments, the surface profile of the polymer substrate was observed by performing super field depth 3D microscopy (VHX-600E, Keyence, Japan) to estimate its roughness.

The tribocharge and three-dimensional forces at the geckos’ feet were synchronously measured when
they climbed vertically upward, at a sample rate of 500 Hz in an isolated room that was under sufficient electromagnetic protection with a temperature corresponding to 22–25 °C and a humidity of 60%–70%. The aisle and probe surfaces were carefully cleaned with alcohol (analytically pure, Nanjing Chemical Reagent Co. Ltd.) before every trial. The complete test system and the animals were electrically neutralized with an ionizing air gun prior to the experiment to diminish the influence of the residual charge on the results. Additionally, anti-static clothing and gloves were utilized throughout the experiment.

Trials in which geckos moved discontinuously or exhibited unstable attachment were eliminated to diminish the uncertainty of the results.

The force signals and charge signals were processed in MATLAB (The Mathworks, Inc, Mass, USA) by using a method used in a previous study [16]. The recorded videos were decomposed into images such that the nominal contact areas $A$ were acquired through image processing.

### 3 Results

A top view of the charge probe was shown in Fig. 3(a) whereas a lateral view was shown in Fig. 3(b). The substrate corresponded to an insulator film with an average thickness of 12.4 μm. Due to surficial defects, such as scratches, the film surface was not very smooth and was full of micro pits and asperities (Figs. 3(a) and 3(c)). A calculation of the exact surface profiles led to obtaining a root mean square roughness (Rq) of 0.69 μm and a mean roughness (Ra) of 0.48 μm for the insulator substrate.

When a gecko treaded on a sensor unit (Fig. 4(a)) embedded in the aisle, its foot interacted with the polymer substrate to obtain forces that caused its locomotion, and this resulted in synchronous tribocharge and forces at the foot-polymer interfaces. The charge was generated at the moment when the foot of the gecko came into contact with the substrate, and it increased steadily before it reached a maximum at the moment of separation (Fig. 4(b)). Additionally,

<table>
<thead>
<tr>
<th>Sensor type</th>
<th>Full scale (FS)</th>
<th>Coupling (%FS)</th>
<th>Non-linearity (%FS)</th>
<th>Hysteresis (%FS)</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>1,500 mN</td>
<td>1.29</td>
<td>0.13</td>
<td>0.21</td>
<td>2 mN</td>
</tr>
<tr>
<td>Force</td>
<td>1,500 mN</td>
<td>1.74</td>
<td>0.14</td>
<td>0.11</td>
<td>2 mN</td>
</tr>
<tr>
<td>$z$</td>
<td>1,500 mN</td>
<td>2.10</td>
<td>0.13</td>
<td>0.21</td>
<td>3 mN</td>
</tr>
<tr>
<td>Charge</td>
<td>2 nC</td>
<td>—</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.1 PC</td>
</tr>
</tbody>
</table>

Table 1 Summary of the performance of the sensors.
the peaks of frictional adhesion at the geckos’ hind footpads ($F_y$) that resulted from the forces in fore-aft direction ($F_y$) and lateral direction ($F_x$) increased with increases in the nominal contact areas and exhibited a relationship corresponding to $F_y = 1.95 \times 10^{-3} A$ ($R^2 = 0.71$, $n = 24$). In contrast, the results indicated that the normal force peaks ($F_z$) are irrelevant with respect to the contact areas (Fig. 4(c)). Interestingly, when the geckos’ seta arrays interacted with the polymeric films on a vertical wall, the geckos’ hairy foot pads generally exhibited a positive charge with a magnitude ranging from a few tenths of a nano-Coulomb to a few nano-Coulombs. Furthermore, the charge increased with increases in the contact areas ($R^2 = 0.79$, $n = 24$) (Fig. 4(d)). As shown in Fig. 4(e), the contact charge was evidently related to the contact forces. An increase in the frictional adhesion led to a significant increase in the contact charge ($R^2 = 0.86$, $n = 24$).

4 Discussion

As shown in Fig. 3, the substrate that interacted with geckos’ footpads was not as smooth as mirrors and instead possessed a roughness of $Rq = 0.67$ that was caused by surficial defects such as scratches. It should be noted the roughness of the substrate significantly exceeded the roughness span ($Rq = 0.1–0.3$) in which the geckos could hardly adhere to the wall [29]. Thus, it was concluded that geckos could sufficiently contact the substrate without requiring any macroscopic sliding and generated reliable forces. When the geckos climbed vertically upward, their bodies moved sinusoidally (as shown in Fig. 4(a) insert), and this resulted in a wave-like locomotion [30] with a velocity of $0.5 \pm 0.1$ m/s ($n = 24$), and this was in agreement with observations in previous studies [31, 32].

When a gecko interacted with the insulator substrate through its feet to obtain forces that drove or retarded its locomotion, the tribocharge was generated at the moment when the hairy footpads were in contact with the substrate and began increasing before it plateaued at the moment of contact separation (Fig. 4(b)). This was in agreement with the property of tribo-electrification. Specifically, the charge increased steadily even when the dominant forces ($F_y$) already decreased (Fig. 4(b)), and this indicated that this part of the charge resulted from frictional contact. Moreover,

![Fig. 4](image)

**Fig. 4** Experiment results. (a) An illustration of the interaction between a gecko’s foot and the insulator film; insert, the trace of the middle point of the gecko’s front shoulders. (b) Typical test results. (c) The relationship between nominal contact areas and forces. (d) The relationship between a contact area and tribocharge. (e) The relationship between tribocharge and contact forces. The data are obtained from the left hind feet, and therefore the normal forces are positive.
the geckos’ attachment completely constituted a type of frictional contact since their foot hairs could only adhere to a target following relatively micro slips [7]. Therefore, it was concluded that the measured tribocharge mainly resulted from the friction between geckos’ foot hairs and the acrylic oligomer film. Extant studies on other materials indicated that the tribocharge accumulation rate strongly depended on the contact-separation speed [12], and this may also hold for the tribo-electrification of geckos. It should be noted that the geckos moved with a similar speed in the selected trials, and the influence of contact-separation speed was not revealed here.

Linear regressions shown in Fig. 4(c) illustrate that the frictional adhesion forces \( F_s \) are linearly related to the nominal contact areas \( A \) \( (R^2 = 0.71, n = 24) \) and result in a nominal strength of \( 0.17 \pm 0.07 \text{ N/cm}^2 \). Conversely, the normal forces \( F_z \) were irrelevant to the nominal contact areas. The geckos’ toes constitute a type of soft and anisotropic structure that is proficient at generating adhesion and at bearing tensile stresses. However, they are incapable of sustaining compressive stresses due to the small and irregular normal forces at the soft toes of the hind limbs of the geckos. Furthermore, this indicated that a large normal force was not indispensable for adhesive friction in geckos.

The study involved measuring the charge at the substrate instead of that at the moving target, and thus the negative charge at the insulator film suggested that an equal albeit positive charge existed at the geckos’ footpads. This was in agreement with the fact that human hairs are always positively charged when they are rubbed with polymers [13, 33] because the geckos’ foot hair possessed components similar to human hairs [34]. Despite the relatively low correlation \( (R^2 = 0.79, n = 24) \), the tribocharge appeared to increase with increases in the contact areas in the experiment (Fig. 4(d)). This was consistent with the results reported in materials used in previous studies [35, 36]. Generally, large contact areas allow increased setae-polymer interaction that contributes to both tribocharge and forces.

Interestingly, the study findings indicated that the tribocharge significantly increased with an increase in the frictional adhesion forces \( (R^2 = 0.86, n = 24) \) (Fig. 4(e)). Potentially, the large forces enhanced the electrification that occurred at the seta-solid interfaces. Obviously, the tribocharge divided by the contact areas gives a charge density \( \sigma \) of \( 4.05 \pm 1.03 \text{ pC/mm}^2 \) \( (n = 24) \). However, it should be noted that the charge density is definitely underestimated here for two reasons. For one thing, the measured charge (i.e., the charge induced by the real tribocharge) is generally a little inferior to the actual ones. For another, the nominal contact areas are considerably overnumber the actual ones and probably even larger than quintuple of the same [37]. Interestingly, the comparison between the nominal frictional adhesion \( (0.19 \text{ N/cm}^2) \) here and the actual frictional adhesion corresponding to a gecko’s foot on polymers (for example, 2.3 N/cm² on PTFE [14]) approximately estimated the actual contact areas as corresponding to about one tenth of the nominal areas.

Thus, the actual charge density should significantly exceed \( 4.05 \text{ pC/mm}^2 \). Nevertheless, the density of tribocharge that was determined as corresponding to the footpads of freely, vertically, and upward climbing geckos was still smaller than that reported by Izadi et al. [14]. This is because Izadi et al. used different materials (PTFE AF and PDMS) that correspond to the most negative position of the tribo-electric series [35, 36] as a substrate. Additionally, the hairy footpads were compelled to slide on the substrate for approximately 10 mm in their experiment, and the long sliding significantly benefited the tribo-electrification [38]. Moreover, high humidity may also weaken the tribo-electrification in the experiment performed in the present study.

A parallel plate capacitor model was used to approximately estimate the electric field force per unit area that results from the tribo-charge if the boundary effects are overlooked [27] as follows:

\[
F = -\frac{\sigma^2}{2\varepsilon}
\]

where \( \varepsilon \) denotes the effective permittivity. The above charge density was substituted into Eq. (1) to obtain a nominal electrostatic attraction strength that approximately corresponded to 1 N/m² if the charge dissipation was ignored. Considering the negative deviation of measured charge density, an exact electrostatic attraction with a magnitude of approximately dozens of to hundreds of newtons per square meter would
be generated if a more precise charge density was obtained.

Nonetheless, the electrostatic attraction is still significantly smaller than that obtained by Izadi et al. by pulling a gecko’s foot on a PTFE AF film for a few millimeters (approximately $4.7 \times 10^4 \text{ N/m}^2$) [14]. The disparity indicates that the electrostatic interaction at the setae-acrylic oligomer interface may not function as dominantly as the claim in the previous study when geckos move freely. Given that the function range of electrostatic interaction significantly exceeded that of van der Waals interaction, the electrostatic attraction was ideally expected to function in the next step from a distance in which the van der Waals forces were unable to function to reduce the setae-substrate gap such that the van der Waals force could come into effect. However, the assumption was still indeterminable and will be strictly verified in future experiments.

5 Conclusions

Conclusively, the present study involved using a self-developed apparatus to measure the tribocharge at geckos’ footpads when they climbed vertically upwards. The successful synchronous measurement provides a method of deeply investigating tribo-electrification in animal attachment. Interestingly, the results indicate that the tribocharge is related to the frictional forces and contact areas. The experiments illustrate a nominal charge density of $4.05 \text{ pC/mm}^2$ that generates a nominal electrostatic attraction corresponding to $1 \text{ N/m}^2$. An electrostatic attraction with a strength corresponding to approximately $25–100 \text{ N/m}^2$ is expected to be obtained given that the actual contact areas as opposed to the nominal areas are applied for the calculation. However, the electrostatic attraction continues to remain smaller than the van der Waals forces.

There are still a great many of issues unclear additional to this preliminary study on the tribo-electrification in animal attachment. As a continuation, the issues including the actual contact areas of freely moving animals, the influence of substrate materials on the charge, and the precise extent to which the tribocharge contributes to the adhesion and friction will be further investigated.

Acknowledgements

The authors extend their sincere appreciation to Wenbo Wang for raising geckos and aiding in performing the 3D microscopy and to the reviewers for their constructive advices. This study was supported by grants from the National Natural Science Foundation of China (Grants No. 51435008) and funding from the Jiangsu Innovation Program for Graduate Education (Grants No. KYLX16_0328).

All applicable institutional and/or national guidelines for the care and use of animals were followed in the study.

Open Access: The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References


Yi SONG. He received his bachelor and master degrees in engineering mechanics from Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2012 and 2015, respectively. He became a Ph.D. student in mechanical design and theory at NUAA in 2015 and his research interests include bio-tribology, biomechanics, and bio-inspired adhesion.

Zhendong DAI. He is a professor, a supervisor of PhD students, and the director of Institute of Bio-inspired Surface and Engineering in Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. He obtained his Ph.D. degree in 1999 from College of Mechanical and Electrical Engineering, NUAA. After completing his postdoctoral research in 2001 in Max Planck Institute for Developmental Biology, he joined NUAA as a professor. His research areas include bionics, light material, control of bionics, bio-robots, and biological robots.
Chemical effects on the sliding friction of Ag and Au(111)

H. E. KO, S. G. KWAN, H. W. PARK, A. CARON
KoreaTech–Korea University of Technology and Education, Chungnam Province 330-708, Republic of Korea
Received: 08 February 2017 / Revised: 22 April 2017 / Accepted: 10 May 2017
© The author(s) 2017. This article is published with open access at Springerlink.com

Abstract: We have investigated the sliding friction behavior of metallic couples with different enthalpy of mixing or reaction by friction force microscopy. Comparing the friction behavior of miscible and immiscible couples we find that in the first case friction is governed by adhesion while the shear strength is low ($\tau = 3-6$ MPa). In the latter case of immiscible couples, adhesion is found to be low and the shear strength is large ($\tau \approx 70$ MPa). Statistical analysis of atomic stick-slip images recorded on an Au(111) surface with tips of different affinities with gold allows for a deeper understanding of our results. The periodicity of atomic stick-slip images corresponds to the interatomic distance of gold for immiscible counter-bodies. In contrast, for a reactive couple the periodicity of atomic stick-slip significantly differs from the gold interatomic distance and may correspond to the structural length of an ordered intermediate phase at the tip-surface interface.

Keywords: metals; surface; friction; nanotribology; atomic force microscopy

1 Introduction

Understanding and controlling the friction between materials is crucial for the development of new devices. With the development of atomic force microscopy (AFM) and friction force microscopy (FFM), the sliding response of a nanometer-scaled single asperity onto flat surfaces has become experimentally accessible. This has allowed bridging the gap between macro-scale measurements of friction and wear and their underlying theories based on mechanisms taking place at the microscopic scales. Owing to their technical relevance, the studies of friction and wear have long been focused on metals. The friction and wear of metals has been described in terms of two main mechanisms (see for example Refs. [1, 2]):

- The formation and shearing of adhesive bonds between surface asperities leading ultimately to their rupture and material transfer;
- The ploughing of a softer surface by a harder asperity involving plastic deformation and debris formation.

The development of nano-/micro-fabrication methods has given raise to the production of nano-/micro-electromechanical devices (N/MEMS) [3]. Such devices are usually based on silicon. For such application the use of liquid lubricant is not possible. Oxidized silicon surfaces have been known to have poor tribological properties [4]. The growth of thin metallic films onto silicon devices for operation in reciproc motion appears to be a feasible alternative. Recently, the nanoscale friction and wear behavior of metals has been investigated [5–10]. In the low load regime where adhesive force prevails, friction has been reported to vanish and the sliding of a single asperity onto single crystalline metallic surfaces was characterized by atomic stick-slip [5]. Further the effect of surface modifications on the nanotribology of Au(111) has been investigated and compared to the nanotribological response of a clean Au(111) surface [7]. There the author demonstrated the role of a Au-neck formation at the AFM-tip/Au(111) interface on the nanotribology of gold. Further, the authors demonstrated the role of passivating self-assembled monolayer in suppressing the formation...
of such a neck and the switching of nanoscale friction on Au(111) by electrochemical control. Moreover, the nanotribology of thermally deposited gold on graphite was studied [8, 9]. In contrast to single-asperity friction measurements performed in ultra-high vacuum (UHV)-conditions in Refs. [5, 6], the authors reported on $F_n^{2/3}$-dependence of the friction force on gold nanoparticles measured in ambient conditions, where $F_n$ is the normal force [8]. In Ref. [9] the sliding of gold nano-island deposited on graphite was reported to yield similar friction values as in UHV-conditions, thus pointing at the importance of chemical inertness of the tribological couple and the absence of contamination at the sliding interface for superlubricity. The effect of tip chemistry has been inferred as the coefficients of friction measured on Pt(111) with a SiO$_x$ AFM-tip or a diamond AFM-tip differed by an order of magnitude [10]. At larger load wear was observed to be mediated by mechanisms of plastic deformation [10].

In this work we investigate the sliding friction behavior of different metallic couples with different enthalpies of mixing or reaction by means of FFM. We discuss our results on the basis of the adhesive friction theory of metals. Further, we perform statistical analysis of the observed atomic stick-slip on Au(111) with tips of different chemistries to provide new insights in the surface reactivity of metallic couples.

2 Samples preparation and characterization

A polycrystalline high purity Ag sample purchased from Alfa Aesar was metallographically prepared by mechanical grinding with SiC paper and polishing with diamond suspension with a particle size down to 1 µm. Subsequently, vibration polishing with an Al$_2$O$_3$ suspension with a particle of 25 nm was applied. A 200 nm thick polycrystalline Au(111) layer grown by physical vapor deposition on mica was purchased by Phasis, Switzerland.

Figure 1 shows (a, d, e) non-contact and (b, c) contact AFM topography images recorded on the as-prepared Ag sample and the as received Au sample. The non-contact AFM images were recorded with a stiff diamond coated cantilever (type: CDT-NCLR, manufactured by NanoSensors, Switzerland) with a first bending resonance frequency $f_0 = 192.54$ kHz and a bending stiffness $C_n = 50.14$ N/m. The contact AFM images were recorded with a soft PtSi coated cantilever (type: PtSi-CONTR, manufactured by NanoSensors, Switzerland) with a first bending resonance frequency $f_0 = 14.05$ kHz and a bending stiffness $C_n = 0.386$ N/m. The topography images of the metallographically prepared Ag surface reveal its polycrystalline microstructure. The polycrystalline nature of the Au thin film grown on mica is evidenced in the topography image shown in Fig. 1(d). Higher magnification images

Fig. 1 AFM-topography images on (a–c) polycrystalline Ag and (d, e) a polycrystalline Au(111) thin film grown on mica.
within a single grain show atomically smooth terraces with a width of the order of 100 nm and atomic steps with heights of 0.288 nm or multiples (see Fig. 1(e)). Also to be observed on the topography images recorded on Au(111) is the presence of adsorbates concentrated at steps.

3 Experimental methods

Sliding friction experiments were performed in ambient conditions ($T = 293$ K, RH = 40%) by FFM [11] using an AFM XE-100 manufactured by Park Instruments, Republik of Korea, with three different soft AFM-cantilevers of the types PtSi-CONTR, CONTSC-Au and CONTSC-Pt (manufactured by NanoSensors, Switzerland) for each samples. A single tip of each type and for each sample has been used for repeated measurements. The authors would like to stress out the representative character of the measurements presented here. The significance of AFM-measurements relies on the idealization of a tribological contact between a smooth surface and a single asperity, i.e., the AFM-tip. As demonstrated in Fig. 2 the geometry of the different tip apex could be treated as equivalent. The chemistry of the counter bodies was chosen on the basis of the respective phase diagrams between the samples and counter bodies, in order to probe the effect of miscibility between tip and sample materials. Owing to its chemical inertness and its high thermodynamic stability a PtSi counter body was used as a reference and in order to realize the case of complete immiscibility with Ag and Au(111). The thermodynamic stability of PtSi is demonstrated by its high melting point $T_m = 1,504.74$ K [12] and its highly negative standard enthalpy of formation $\Delta H_f = -118$ kJ/mol [13]. In contrast, a gold counter body was selected owing to its full miscibility with Ag so as to exemplify the case of a negative enthalpy of mixing ($\Delta H^\text{mix}_{\text{Ag-Pt}} = -4.5$ kJ/(g·atom) at 800 K) [14], on the one hand. On the other hand, the case of a Au counter body sliding on a Au(111) surface serves the purpose of realizing a tribological couple with a zero enthalpy of mixing. Moreover, a platinum counter body was selected to study the effect of immiscibility between materials involved in a tribological couple; the Ag–Pt phase diagram exhibits a peritectic decomposition of the liquid phase characterized by a peritectic temperature of 1,459 K and a peritectic concentration $X_{\text{Pt}} = 40.6$ at.%. At room temperature a biphased domain is usually accepted although the existence of intermetallic phases below 1,238 K is still under debate (see Ref. [14]). Further, the Au–Pt phase diagram exhibits a miscibility gap below 1,533 K at $X_{\text{Pt}} = 61$ at.%. At room temperature Pt is soluble into Au up to $X_{\text{Pt}} = 7$ at.%; for larger platinum compositions the Au–Pt phase diagram exhibits a biphased domain at room temperature [14].

Prior to friction experiments and for each cantilever used in this work the sensitivity of the AFM-photodiode $S$ was calibrated by recording a force distance curve on the respective sample surfaces and by calculating the slope of the repulsive part of the curve. Subsequently, the bending and torsion stiffness

![Fig. 2](https://mc03.manuscriptcentral.com/friction)
of each cantilever were determined from its thermal noise vibration [15]. The cantilevers’ stiffnesses are listed in Table 1. The friction experiments consisted in recording the lateral deflection signal of the AFM-cantilevers in both forward and backward direction of the x-scanner. Thereby the lateral deflection signal \( V_l \) was recorded in volt and was subsequently converted into units of force according to \( F = \frac{L^3 h}{2 C_l} S V_l \) [16], where \( h \), the tip height, was taken to be 12 \( \mu \)m, in agreement with the manufacturer’s data. In the case of Ag the images size was selected as 500 nm \( \times \) 500 nm and experiments were recorded with a normal load \( F_n = 0–10 \) nN. In the case of Au(111) the experiments were performed over an area of 10 nm \( \times \) 10 nm with a normal load \( F_n = 0–10 \) nN. On Au(111) friction measurements were repeated twice for each load value. In this work, two different scan areas were selected for Ag and Au(111). On the polished Ag surface a larger scan area was selected to average out the effects of topography resulting from the surface preparation. With such a preparation method of the Ag sample no atomically smooth surface can be achieved. The as-sputtered Au(111) thin film on mica displayed large atomically flat terraces that were suitable to perform friction measurements at the atomic scale. While the measurements on Ag aimed at a proof of concept, we expected the measurements on Au(111) to further provide insights into the mechanisms responsible for the chemical effects on metals friction at the atomic scale.

For each measurement the friction force was calculated according to \( F_i = \frac{F_{i,\text{fwd}} - F_{i,\text{bwd}}}{2} \), where \( F_{i,\text{fwd}} \) and \( F_{i,\text{bwd}} \) are the forward and backward images of the lateral force, respectively. Subsequently, the calculated friction force image was averaged line-by-line and a corresponding error was calculated as the standard deviation from the mean value using the Matlab software package (see script1.m in Electronic Supplementary Material (ESM)) [17]. Moreover the shear strength \( \tau \) and the adhesion force \( F_{ad} \) corresponding to each tribological couple were calculated by fitting the \( F(F_n) \)-plot with the function \( F_i = \tau A_c(F_n) \), where we consider \( \tau \) to be a constant and \( A_c(F_n) \) is the normal force dependent real area of contact between surface and tip (see Refs. [18]). In the frame of the Hertzian theory of elastic contact between a spherical body and a flat surface the real contact area is expressed as \( A_c = \pi \left( \frac{3 F_{ad} R}{2 E} \right)^{2/3} \), where \( R \) is the radius of the spherical body and \( E' = \left( \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \right)^{-1} \) is the reduced modulus of elasticity, and \( E_i \) and \( \nu_i \) are the Young’s moduli and the Poisson’s radii of the two elastic bodies involved in the contact [19]. The Hertzian contact theory however do not consider the effect of adhesion forces. Based on the Johnson-Kendall-Roberts (JKR) theory the real area of an adhesive contact between a spherical elastic body and the flat surface of an elastic body can be expressed as \( A_c = \pi \left( \frac{R}{E} \right)^{2/3} \left( F_n - F_{ad} \right)^2 + 2 F_{ad} + \sqrt{4 F_{ad} (F_n - F_{ad})^2 + (2 F_{ad})^2} \right)^{2/3} \), where \( F_{ad} \) is the adhesion force between the two elastic bodies [19]. Inserting this expression of \( A_c \) into \( F_i = \tau A_c(F_n) \) results in a \( F_n^{2/3} \)-dependence of the friction force that has for example been experimentally observed in Refs. [8, 18, 20]. Following values were used for the Young’s modulus and the Poisson’s ratio: \( E_{Ag} = 83 \) GPa and \( \nu_{Ag} = 0.37 \), \( E_{Au} = 75 \) GPa and \( \nu_{Au} = 0.44 \), \( E_{Pt} = 168 \) GPA and \( \nu_{Pt} = 0.38 \), and \( E_{PtSi} = 238.5 \) GPa

<table>
<thead>
<tr>
<th>Sample</th>
<th>Cantilever type</th>
<th>( C_n ) (N/m)</th>
<th>( C_l ) (N/m)</th>
<th>( L^{*} ) (mm)</th>
<th>( R^{**} ) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag and Au(111)</td>
<td>PtSi-CONTR</td>
<td>0.396</td>
<td>307.403</td>
<td>450</td>
<td>25</td>
</tr>
<tr>
<td>Ag</td>
<td>CONTSC-Au</td>
<td>0.650</td>
<td>129.21</td>
<td>225</td>
<td>25</td>
</tr>
<tr>
<td>Ag</td>
<td>CONTSC-Pt</td>
<td>0.783</td>
<td>155.71</td>
<td>225</td>
<td>25</td>
</tr>
<tr>
<td>Au(111)</td>
<td>CONTSC-Au</td>
<td>0.685</td>
<td>136.24</td>
<td>225</td>
<td>25</td>
</tr>
<tr>
<td>Au(111)</td>
<td>CONTSC-Pt</td>
<td>0.951</td>
<td>189.81</td>
<td>225</td>
<td>25</td>
</tr>
</tbody>
</table>
and $V_{\text{PtSi}} = 0.316$ [21, 22]. In this work the radius of curvature $R$ was determined for each tip from scanning electron microscopy (SEM) images recorded after the friction experiments (see Fig. 2 and Table 1) using a Helios 600i dual-beam FIB-SEM manufactured by FEI, the Netherlands. For all tips a similar value $R \approx 25$ nm was found and used to fit the experimental $F_f(F_n)$-plots. In Fig. 2(a) circle with a radius of 25 nm is overlaid on each of the tips to demonstrate the validity of this value. Further, elemental chemical analysis by energy dissipative X-ray spectroscopy (EDS) was performed in SEM at the apex of the Pt-tip shown in Fig. 2(e) after friction measurements on Au(111) by using a XFlash Detector 6 manufactured by Bruker, Germany.

In the case of Au(111) atomic-scale stick-slip was observed with all tips and statistically analyzed. A first analysis consisted in calculating the slip-length and its frequency as a function of the load. Thereby, the peak force positions in each line of recorded $F_{\text{L,fwd}}$-images were determined with the findpeaks-function of the Matlab software package [17]. Beforehand an average-baseline had been calculated and subtracted from each line of the image. To account for the instrumental noise, we also fitted our experimental data line-by-line with an interpolant function (spline-function). After determination of the peak force positions, the slip-distances were calculated as the distance between two successive force peaks for each line of the analyzed image and plotted in a single histogram (see script2.m in ESM). A second analysis consisted in line-by-line calculation of the power spectral density (PSD)-function of each recorded $F_{\text{L,fwd}}$-images using the pburg-function of the Matlab software package (see script3.m in ESM) [17]. The calculated PSD-functions corresponding to each line were averaged to provide a single PSD-function out of one $F_{\text{L,fwd}}$-image. This statistical analysis transforms a signal in real-space into a one-dimensional reciprocal space ($k$-space) signal, from which characteristic wavelengths $\lambda = 2\pi/k$ can be identified.

4 Results

Figure 3 shows the load dependence of friction on polycrystalline Ag and on Au(111) with three different counter-bodies: PtSi, Au, and Pt; the corresponding shear strength and adhesion force values are listed in Table 2. In the case of Ag, the load dependent friction force values were found lowest with a PtSi-coated tip. Fitting the load dependence of the friction for the Ag/PtSi tribological couple yields a shear strength value $\tau = 74.97$ MPa and an adhesion force value $F_{\text{ad}} = 5.26$ nN. For the same sample, the largest load dependent friction force values were obtained with a Au-coated tip and corresponded to a shear strength

![Fig. 3](https://mc03.manuscriptcentral.com/friction)
Table 2  Shear strength $\tau$ and adhesion force $F_{ad}$ of the different tribological couples investigated in this work.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Counter body</th>
<th>$\tau$ (MPa)</th>
<th>$F_{ad}$ (nN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>PtSi</td>
<td>74.97</td>
<td>5.26</td>
</tr>
<tr>
<td></td>
<td>Au</td>
<td>3.7</td>
<td>2805</td>
</tr>
<tr>
<td></td>
<td>Pt</td>
<td>24.41</td>
<td>19.07</td>
</tr>
<tr>
<td></td>
<td>PtSi</td>
<td>68.21</td>
<td>0</td>
</tr>
<tr>
<td>Au(111)</td>
<td>Au</td>
<td>24.21</td>
<td>25.8</td>
</tr>
<tr>
<td></td>
<td>Pt</td>
<td>6.6</td>
<td>443.5</td>
</tr>
</tbody>
</table>

$\tau = 3.7$ MPa and an adhesion force $F_{ad} = 2,805$ nN. Both cases represent the extreme of fully-immiscible tribological partners in the case of Ag/PtSi and fully miscible partners in the case of Ag/Au. For the Ag/Pt tribological couple the load dependent friction force values were intermediate but closer to the values obtained with a Pt-coated tip than with an Au-coated tip. For the Ag/Pt tribological couple, we found $\tau = 24.41$ MPa and $F_{ad} = 19.07$ nN. The Ag–Pt phase diagram exhibits a peritectic decomposition with a bi-phased domain over all compositions at room temperature, where Ag and Pt are thus immiscible.

In the case of Au(111), the load dependent friction force values was found to be highest with an Au-coated tip. For this tribological couple a shear strength value $\tau = 24.21$ MPa and an adhesion force value $F_{ad} = 25.8$ nN were evaluated. The lowest value of the shear strength $\tau = 6.6$ MPa and the largest adhesion force value $F_{ad} = 443.5$ nN were found for the Au/Pt tribological couple, whose load dependent friction force values are slightly lower than for the Au/Au tribological couple. The lowest load dependent friction force values were measured for the Au/PtSi tribological partner, for which $\tau = 68.21$ MPa and $F_{ad} = 0$ nN were found. The lowest adhesion force and highest shear strength values correspond to the case of full immiscibility between PtSi and Au(111). On the opposite the largest adhesion force and lowest shear strength values correspond to the case of partial miscibility between Pt and Au(111). Pt has a large solubility in Au up to ~8 at.% at room temperature. For larger Pt-concentrations the Au-Pt phase diagram exhibits a miscibility gap. At small Pt-concentration the free Gibbs energy of mixing with Au is thus negative. For the Au/Au tribological couple, in which case the free Gibbs energy of mixing can be taken as zero, we found the load dependent friction force values to correspond to a shear strength $\tau = 24.21$ MPa and an adhesion force $F_{ad} = 25.8$ nN.

Figure 4 shows three friction images and corresponding friction loops recorded on Au(111) with a PtSi-coated tip, a Au-coated tip and a Pt-coated tip. In all cases, atomic scale stick-slip can be observed.
From these measurements one can however recognize that the periodicity of the friction signal significantly depends on the tip chemistry, with the friction signals recorded with a PtSi-coated tip and an Au-coated tip appearing to be the more periodic than the signal recorded with a Pt-coated tip. In the case of the friction signal recorded with an Au-coated tip we further observe that the sticking force was almost constant. With a PtSi-coated tip sliding on Au(111), we observe a modulation of the sticking force. In the case of the Pt-coated tip the friction signal is less ordered. To compare the degree of order/disorder of the friction measurements performed with tips of different chemistries, we statistically analyzed the slip length by counting the number of slip events as a function of the sleep length. For all tips a maximum in the frequency of slip events was observed for a range of slip length $l_{\text{slip}} = 0.256-0.3$ nm (see Figs. 5(a)–5(c)); for comparison the interatomic distance of Au in its [110] crystallographic direction is $a_{[110]} = 288$ pm.

The width of the slip distance distribution however significantly changed depending on the chemistry of the counter-bodies. In agreement with above observations the width of the slip-length distribution is narrowest for the Au-coated tip and increases in the order of the PtSi-coated and Pt-coated tips. In these later cases, the slip length distribution can be rationalized as a sharp maximum for the $a_{[110]}$ interatomic distance overlaid by a wider distribution of slip lengths. As shown in Fig. 6 the slip length distributions for all three counter-bodies were not significantly affected by the normal load, except at the highest applied load in this work $F_n = 10$ nN, at which wear was observed to set on and atomic

---

**Fig. 5**  (a–c) Slip-length distribution and (d–f) average one-dimensional power spectrum density functions calculated from FFM-images recorded on Au(111) with (a, d) a PtSi-coated tip, (b, e) a Au-coated tip, and (c, f) a Pt-coated tip.

**Fig. 6**  Slip-length distribution on Au(111) as a function of the normal load with (a) a PtSi-coated tip, (b) a Au-coated tip, and (c) a Pt-coated tip.
stick-slip partially disappeared. In particular, at $F_n = 10$ nN the slip-length distribution corresponding to the Au(111)/Pt tribological couple exhibited a less prominent maximum for $l_{slip}$ and a wider distribution of slip lengths on the overall.

In order to further characterize the degree of order at the interface between Au(111) and the different counter-bodies used in our friction experiments, we further approximated the averaged power spectrum density (PSD)-functions of the friction signals recorded at different loads. Figures 5(d)–5(f) show typical PSD-functions for the three different tips used on Au(111). The PSD-function corresponding to a typical friction measurement on Au(111) with a Au-coated tip shows a single peak at a wavenumber $k = 21.36$ rad/nm (see Fig. 5(e)). Neither the position nor the amplitude of this peak were found to change upon increasing load; except for $F_n = 10$ nN, in which case two slightly less prominent peaks were observed at $k = 20.11$ rad/nm and $k = 22.62$ rad/nm (see Fig. 7(b)).

For the PtSi-coated tip a typical PSD-function shown in Fig. 5(d) exhibits two maxima at $k = 20.11$ rad/nm and $k = 23.25$ rad/nm. These two peaks were observed for a load ranging from 0 to 8 nN. At higher load values up to $F_n = 10$ nN, a single peak was observed at $k = 21.36$ rad/nm (see Fig. 7(a)). A power density spectrum corresponding to the Pt-coated tip sliding on Au(111) is shown in Fig. 5(f). There, three most prominent peaks are observed at $k = 6.28$ rad/nm, $k = 16.96$ rad/nm and $k = 23.88$ rad/nm. Further peaks can be further recognized in Fig. 7(c) at $k = 10.05$ rad/nm, $k = 13.82$ rad/nm, $k = 34.56$ rad/nm and $k = 50.27$ rad/nm, the intensities of which increase with the normal load.

The characteristic wavelength $\lambda$-values corresponding to the above wavenumbers are listed in Table 3. In the case of the PtSi- and Au-coated tips the characteristic wavelength $\lambda_2 = 0.294$ nm corresponds to the interatomic distance of Au in the [110] direction ($a_{[110]} = 288$ pm). The small discrepancy arises from the numerical approximation of the PSD function. In Figs. 4(d), 6(a) and 6(b) the maximum in the PSD-functions are also found to split into two equidistant peaks with corresponding wavelength-values $\lambda_3 = 0.277$ nm and $\lambda_1 = 0.312$ nm, respectively. We attribute these peaks to the herringbone reconstruction of the Au(111) surface and the resulting different tilt angles of the fcc and hcp domains with respect to the unreconstructed surface [23]. For the Pt-coated tip three main characteristic wavelengths can be determined from Figs. 5(f) and 7(c): $\lambda_1 = 1$ nm, $\lambda_2 = 0.37$ nm, and $\lambda_3 = 0.263$ nm. The smaller of these values may correspond to the Pt–Pt interatomic distance. The metallic radius of Pt is given in the literature as 0.138 nm [24], which corresponds to the interatomic distance of pure Pt in the [110] crystallographic direction $a_{[110]} = 0.276$ nm. The difference between this value and the one obtained from the PSD function may arise from the numerical approximation. Further the $\lambda_2$-value found for the

![Fig. 7](image)

**Fig. 7** Load dependence of the power spectrum density functions calculated from FFM measurements on Au(111) with (a) a PtSi-coated tip, (b) a Au-coated tip, and (c) a Pt-coated tip.

<table>
<thead>
<tr>
<th>Counter body</th>
<th>$k_1$ (rad/nm)</th>
<th>$k_2$ (rad/nm)</th>
<th>$k_3$ (rad/nm)</th>
<th>$\lambda_1$ (nm)</th>
<th>$\lambda_2$ (nm)</th>
<th>$\lambda_3$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PtSi</td>
<td>20.11</td>
<td>21.36</td>
<td>23.25</td>
<td>0.312</td>
<td>0.294</td>
<td>0.277</td>
</tr>
<tr>
<td>Au</td>
<td>20.11</td>
<td>21.36</td>
<td>22.62</td>
<td>0.312</td>
<td>0.294</td>
<td>0.277</td>
</tr>
<tr>
<td>Pt</td>
<td>6.28</td>
<td>16.96</td>
<td>23.88</td>
<td>1</td>
<td>0.37</td>
<td>0.263</td>
</tr>
</tbody>
</table>
Pt-coated tip matches literature values for the lattice parameters of Au-Pt intermediate phases such as Au₃Pt (cubic lattice; \(a = 0.392\) nm), AuPt (tetragonal lattice; \(a = 0.392\) nm and \(c = 0.39\)) or AuPt₃ (cubic lattice; \(a = 0.392\) nm) [14].

To support these results we analyzed the Pt-coated tip used Au(111) with regard to its elemental chemistry by energy dissipative X-ray spectroscopy (EDS) inside a SEM. Figure 8(a) shows the energy spectrum recorded with an electron acceleration voltage of 10 kV on an area equivalent to the overlaid circle at the tip apex in Fig. 2(c). From this spectrum five elements were detected: C, O, Si, Pt and Au. Figures 8(b)–8(d) are magnifications of the spectrum in Fig. 8(a). As shown in Fig. 8(b) the most prominent EDS-peak was found at an energy value 0.277 keV that corresponds to the Kα-spectral line of carbon and is a result of contamination. Although Pt and Au have N-spectral lines at 0.251 keV and 0.258 keV, respectively, fitting this peak with the sum of three distinct Gaussian distributions did not yield a satisfactory deconvolution. This can be explained by the prominence of the carbon signal compared to the ones for Pt and Au. A second peak was detected at an energy value 0.525 keV that corresponds to the Kα-spectral line of oxygen (see Fig. 8(c)). In Fig. 8(d) three peaks can be distinguished. There, a first peak at an energy value 1.74 keV is associated to the Kα-spectral line of silicon, which is the basis material of the AFM-tip. A second peak is found to consist in the convolution of two Gaussian distributions: one associated to the Mα-spectral line of Pt at an energy value 2.05 keV that corresponds to the coating material of the tip and the other to the Mα-spectral line of Au at an energy value 2.123 keV, that is associated to transferred material from the Au(111) surface to the tip. The elemental peak areas were further used to evaluate the chemical concentration at the tip apex. Thereby, only the elemental peaks for Si, Pt, and Au were considered. We found \(X_{Si} = 71.3\) at.%, \(X_{Pt} = 25.35\) at.% and \(X_{Au} = 3.35\) at.%. The volume probed by EDS reaches far below the tip surface, as indicated by the detection of Si. The Pt-coating thickness is given by the manufacturer to be \(\sim 100\) nm. These two factors explain the larger concentrations of Si and Pt compared to the Au-concentration, since Au is expected to be localized at the tip surface. It is therefore expected that the Au/Pt-surface concentration ratio is significantly larger than 0.13 as calculated from the above concentration values. To obtain an accurate value of the chemical surface composition and structure of the tip surface, small area x-ray photoelectron spectroscopy and atom probe tomography should be used. The application of both methods remains however very challenging.

Fig. 8  Chemical EDS-analysis of a Pt-coated tip used for friction measurements on Au(111).
at this moment. To the best of the authors knowledge, none of these methods has been applied to determine the chemical composition or structure at the surface of an AFM tip. Alternatively, a combination of transmission electron microscopy (TEM) and electron energy loss spectroscopy could be applied to provide a lateral resolution than provided by EDS. However, the collected electron signal would in this case too have a contribution from the tip volume through which the primary electron would transmit.

5 Discussion

According to Bowden and Tabor, the friction force during relative sliding between metallic surfaces results from the formation of junctions at surface asperities and their shearing [1]. The formation of junctions can proceed by severe plastic deformation and cold welding. In such cases, the shear strength of a junction relates to its resistance to plastic deformation and shearing off the junction usually results in materials transfer from the softer to the harder asperity. The macroscopic sliding friction of metals has been extensively investigated by Rigney [25]. It has been reported that material transfer occurs for both miscible and immiscible tribological partners. In both cases, smearing of the softer metals onto the harder one has been observed. Also, Rabinowicz has suggested that during adhesive wear between metals the friction coefficient depends on their respective phase diagrams [2], i.e., metallurgical compatibility, with compatible metals having a high friction coefficient due to their work of adhesion. There, the work of adhesion \( W_{ad} = c_m (\gamma_a + \gamma_b) \), where \( c_m \) is a compatibility factor between the metals \( a \) and \( b \) with the respective surface energies \( \gamma_a \) and \( \gamma_b \). Based on the experimental determination of the work of adhesion between different metals and their respective surface energies, the values of \( c_m \) have been calculated to be 1 for identical metals and to decrease to 0.5 for compatible metals and to 0.2 for incompatible metals. Under unlubricated conditions, Rabinowicz reported the friction coefficient between metals subjected to adhesive wear to be 0.8 for identical metals and to decrease to 0.63 for compatible metals and to 0.35 for incompatible metals [2]. Following a similar scheme, Ando et al. measured the friction coefficient of various metallic couples and discussed their results based on the difference in atomic sizes of the involved elements [26]. The authors found that friction was largest for small atomic size differences between the tribological partners. This corresponds to the situation where the formation of substituitional solid solution is not impeded by strain energy.

Table 4 lists the values of surface energies for the investigated materials, for which the adhesion energy values were calculated using the \( c_m \)-values suggested by Rabinowicz. Thereby, Rabinowicz considered both Ag and Pt or Au and Pt to be compatible, while their respective equilibrium phase diagrams indicate almost no solubility at room temperature [14]. The Ag-Pt equilibrium phase diagram shows a peritectic decomposition of the liquid phase with a peritectic temperature of 2,139 K and a peritectic concentration

<table>
<thead>
<tr>
<th></th>
<th>Ag</th>
<th>Au</th>
<th>Pt</th>
<th>PtSi</th>
<th>Ag-Au</th>
<th>Ag-Pt</th>
<th>Ag-PtSi</th>
<th>Au-Au</th>
<th>Au-Pt</th>
<th>Au-PtSi</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma (J/m^2) )</td>
<td>0.92</td>
<td>1.12</td>
<td>1.8</td>
<td>1.4(^{\dagger} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( c_m )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.8(^{\dagger} )</td>
<td>0.8(^{\dagger} )</td>
<td>0.2</td>
<td>1</td>
<td>0.8(^{\dagger} )</td>
<td>0.2</td>
</tr>
<tr>
<td>( W_{ad} (J/m^2) )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.63</td>
<td>2.17</td>
<td>0.46</td>
<td>2.24</td>
<td>2.33</td>
<td>0.5</td>
</tr>
<tr>
<td>( r (pm) )</td>
<td>144</td>
<td>144</td>
<td>138</td>
<td>103(^{\dagger} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(</td>
<td>\Delta r</td>
<td>/r ) (%)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>3.82</td>
<td>28.47</td>
<td>0</td>
</tr>
<tr>
<td>( \Delta H_{mix} (kJ/mol) )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-4.5</td>
<td>6(^{\ddagger} )</td>
<td>( \infty^+ )</td>
<td>0</td>
<td>6(^{\ddagger} )</td>
<td>( \infty^+ )</td>
</tr>
</tbody>
</table>

Table 4 Surface energy \( \gamma \) [2], calculated adhesion work per unit surface \( W_{ad} \), atomic radius \( r \) [24], relative atomic difference \( |\Delta r|/r \), and enthalpy of mixing. \(^{\dagger}\) Calculated surface energy value for PtSi from Ref. [29]; \(^{\ddagger}\) the \( c_m \)-values correspond to the values for the studied metallic pairs suggested in Ref. [2]. \(^{\ast}\) the \( r \)-value given for PtSi corresponds to the half bond length between Pt and Si as reported in Ref. [30]; \(^{\circ}\) this value corresponds to the excess enthalpy of the equimolar Ag-Pt solid solution [14]; \(^{\&}\) the \( \Delta H_{mix} \)-value between PtSi and Au or Ag is here assumed to be largely positive; \(^{\pi}\) this value corresponds to the calculated excess free Gibbs energy of formation of solid Au–Pt alloys at 1,423 K reported in Ref. [14].
Friction $X_{\text{Pt}} = 40.6 \text{ at.}\%$ and a phase mixture of Ag and Pt at room temperature. The preparation of Ag–Pt metastable solid solution has been reported in Ref. [27], where the authors used splat cooling to quench the homogeneous Ag–Pt liquid phase. In Ref. [28], Peng et al. have reported on the organometallic synthesis and stability of Ag–Pt nanoalloys, i.e., nanoparticles consisting in solid solutions within the miscibility gap of the Ag–Pt system. This may indicate a larger contribution of the entropy of mixing at surfaces than in the bulk. According to Ref. [14], the excess enthalpy of the equimolar Ag-Pt solid solution is 6 kJ/mol.

It is interesting to compare the values of the adhesion force and friction coefficient listed in Table 2 for the same metals combination as in Table 4. We observe a better correlation between the measured adhesion force and enthalpy of mixing or the relative atomic size difference than with the calculated energy of mixing based on the compatibility parameter. However, our results are insofar in contrast to earlier works by Rabinowicz or Ando et al. that in the present work metallic couples having a high adhesion exhibit a low friction coefficient and vice versa.

In Refs. [2, 26] the friction was measured at the macroscale, in which case the contact between metallic surfaces is known to consist of many asperities that undergo plastic deformation leading to the cold welding of junctions. In this configuration friction corresponds to the resistance of these junctions to be sheared. Ultimately, the shearing of junctions leads to their fracture and results in material transfer of the softer metal onto the harder. In our experiments the response of a single nanometer-scale metallic asperity sliding on metal was recorded in the wear-less regime as indicated by the occurrence of atomic stick-slip on the Au(111) surface (see Fig. 4). In Ref. [5], the authors investigated the sliding friction behavior of single crystalline metallic surfaces by AFM under UHV-conditions and by using a SiO$_2$ tip as a single-asperity counter-body. There, the friction force was observed to vanish in the adhesive regime. Although in this regime almost no energy dissipation was observed, the lateral force signals were characterized by a clear atomic stick-slip modulation. At higher loads, i.e., in the regime of repulsive forces, wear was observed to set on: the periodicity of stick-slip was disrupted and significant friction was measured. In Ref. [6], Gosvami et al. investigated the ageing of single-asperity sliding contacts between an AFM-tip and an Au(111) surface at low temperature and under UHV conditions. There the authors also observed stable stick-slip with atomic periodicity. An interesting observation was that at low temperature friction increased with time and resulted in a distortion of the stick-slip behavior. Reinitializing the contact after pulling out the tip away from the surface resulted in a decrease of friction and the restoration of atomic stick-slip periodicity. These experimental results were further supported by molecular dynamics simulation of the growing process of gold junctions. The authors concluded that periodic stick-slip observed at room temperature is the result of a dynamic equilibrium of the contact that involves surface diffusional self-healing of microscopic wear damage. The results by Gosvami et al. are further in line with previous results by Merkle and Marks demonstrating the wetting of a sharp tungsten tip by a gold surface and subsequent formation and elongation of a neck upon pulling the tip away from the surface [31]. More recently, Cihan et al. pointed at the importance of chemical inertness of the tribological couple and the absence of contamination at the sliding interface for superlubricity. In ambient conditions, the authors reported a $F^2/3_n$-dependence of the friction force on gold [8]. However, the friction force arising from sliding a gold nano-island deposited on graphite yielded similar friction force values as in UHV conditions [9].

Our experimental results indicate that a high adhesion force corresponds to high friction force values and to low shear strength values. For low adhesive contacts, the friction forces are found to significantly increase with the normal load and correspond to a larger shear strength value. Our measurements were all performed under same ambient conditions ($T = 293 \text{ K}$ and RH = 40%). For such conditions, it has been known that an adsorbed water layer at the interface between tip and sample surface can affect the adhesion force to a degree dependent on the hydrophilic character of the involved surfaces in the contact. Our results do not allow to exclude the role of water. However, all involved metal surface materials are hydrophilic in nature as indicated by the similar
adsorption energy of water onto them (see Ref. [32]). It thus appears reasonable to consider the contribution of adsorbed water to the adhesion force as a constant off-set value for all tribological couples investigated in this work that do not affect the relative differences reported here.

The relation between the friction forces, the shear strength and the adhesion force can be explained by the formation of a stable and load independent junction geometry in the case of miscible systems such as Ag-Au and Au(111)-Pt that is most probably assisted by wetting effects. For immiscible systems such as Ag-PtSi, Au(111)-PtSi and Ag-Pt the formation of the tip-sample junction is supposed to be mediated by plastic deformation, as inferred in Ref. [5]. The formation of a junction in the case of the Au(111)-Au system appears to be both affected by wetting and plastic deformation, giving raise to both high adhesion and shear strength.

Figures 4–7 show the differences in stick-slip characteristics during sliding friction measurements on Au(111) with different tips. With a PtSi-tip and an Au-tip, clear periodic stick-slip events were observed. There, the periodicity coincided with the interatomic distance between Au atoms in the [110]-direction. Also, in the case of the PtSi-tip a further modulation of the friction signal was observed that has been attributed in Ref. [23] to the Herringbone reconstruction of the Au(111) surface. In our statistical analysis, the presence of the Herringbone reconstruction resulted in a splitting of the PSD-peak, into two symmetrical peaks around the value corresponding to the interatomic distance of gold for immiscible counter-bodies. In contrast, for a reactive couple the periodicity of atomic stick-slip significantly differs from the gold interatomic distance and may correspond to structural length of an ordered phase at the tip-surface interface. These results provide new insights in the formation of interfacial alloys and their effects on metals friction. Furthermore, our findings shall serve as new guidelines for the selection of materials couples for micromechanical devices involving sliding contacts.

6 Conclusions

Comparing the friction behavior of miscible and immiscible couples we find that in the first case friction is governed by adhesion and the shear strength has a low value. In the latter case of immiscible couples, adhesion is found to be low and the shear strength exhibits a large value. Statistical analysis of atomic stick-slip images recorded on an Au(111) surface with tips of different affinities with gold allows for a deeper understanding of our results. Expectedly, the periodicity of atomic stick-slip images corresponds to the interatomic distance of gold for immiscible counter-bodies. In contrast, for a reactive couple the periodicity of atomic stick-slip significantly differs from the gold interatomic distance and may correspond to structural length of an ordered phase at the tip-surface interface. These results provide new insights in the formation of interfacial alloys and their effects on metals friction. Furthermore, our findings shall serve as new guidelines for the selection of materials couples for micromechanical devices involving sliding contacts.

Acknowledgement

This work was supported by KoreaTech.

Electronic Supplementary Material: The Matlab software packages (script1.m, script2.m, and script3.m) are available in the online version of this article at https://doi.org/10.1007/s40544-017-0167-5.

Open Access: The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In this work we used the version R2015b. Matlab is manufactured and distributed by MathWorks, U.S.A.

References

[17] In this work we used the version R2015b. Matlab is manufactured and distributed by MathWorks, U.S.A.
**Arnaud CARON.** He is a materials scientist with expertise in the multi-scale mechanical behavior of materials, surfaces and micro-components. Since 2015 Arnaud Caron is assistant professor in the School of Energy, Materials and Chemical Engineering at KoreaTech–Korea University of Technology and Education, Republic of Korea. Arnaud Caron obtained his engineering degree in materials science in 2004 from the University of Saarland, Germany and was awarded with the Schiebold Medal. In 2009 he earned his doctoral degree in materials science from the University of Saarland, Germany. From 2007 to 2015 Arnaud Caron worked as a research associate at the Institute of Micro- and Nanomaterials of the University of Ulm, Germany, the WPI-Advanced Institute of Materials Research, Japan and the Leibniz–Institute for New Materials, Germany.

**Hui Eun KO.** She received her bachelor degree in material science and engineering in 2016 from KoreaTech—Korea University of Technology and Education, Cheonan, Republic of Korea. Since then, she has been studying for her master degree in material science and engineering at KoreaTech. Her research interests include nano-tribology and mechanical properties of metals and metallic glasses.
Transfer of electrons on scratched iron surfaces: Photoelectron emission and X-ray photoelectron spectroscopy studies

Yoshihiro MOMOSE\(^1\)*, Daisuke SUZUKI\(^1\), Keika TSURUYA\(^1\), Takao SAKURAI\(^2\), Keiji NAKAYAMA\(^3\)

\(^1\) Department of Materials Science, Ibaraki University, 4-12-1 Nakamarusawa, Hitachi 316-8511, Japan
\(^2\) Ashikaga Institute of Technology, Ashikaga 326-8558, Japan
\(^3\) Institute of Mesotechnology, 2-2-1-201 Nakahara, Kashiwa 277-0085, Japan

Received: 22 March 2017 / Revised: 27 April 2017 / Accepted: 22 May 2017
© The author(s) 2017. This article is published with open access at Springerlink.com

Abstract: We report the activation energy, \(\Delta E_a\), for the quantum yield in thermally assisted photoelectron emission (TAPE) under 210-nm-wavelength light irradiation, and the associated X-ray photoelectron spectroscopy (XPS) results. Samples were cleaned only in acetone and scratched in air, water, methanol, ethanol, acetone, benzene, and cyclohexane. Glow curves, describing the temperature dependence of photoelectron emission (PE) quantum yield (emitted electrons/photon), \(Y\), were obtained. A simple method of determining \(\Delta E_a\) using \(Y\), called \(Y\)GC, at seven temperatures up to 353 °C, for the same \(Y\) glow curve, was proposed. The \(\Delta E_a\) obtained using this method was almost the same as that obtained from \(Y\) for seven stationary temperatures \((Y_{st})\). For scratched samples, the TAPE was measured over two cycles of temperature increase and subsequent decrease (Up1, Down1 and Up2, Down2 scans) in the 25–339 °C range, and \(\Delta E_a\) was obtained from \(Y\)GC. The Arrhenius plot was approximated by a straight line, although a convex swelling peak appeared in the Up1 scan. \(\Delta E_{a\text{Up1}}\) was in the 0.212–0.035 eV range, depending on the environment in which scratching was performed; \(\Delta E_{a\text{Up1}}\) for water was much higher than that for acetone. This was explained in terms of the mode of the acid–base interaction between the liquid molecules and the hydroxyl group of Fe–OH. The values of \(\Delta E_{a\text{Down1}}, \Delta E_{a\text{Up2}},\) and \(\Delta E_{a\text{Down2}}\) were in the 0.038–0.012 eV range. The total count of electrons emitted during the Up1 and Up2 scans was found to decrease with increasing \(\Delta E_{a\text{Up1}}\) and \(\Delta E_{a\text{Up2}}\), respectively. \(\Delta E_{a\text{Up2}}\) was found to increase with increasing presence of the FeO component in the analyzed Fe oxides. The convex swelling peak was attributed to the removal of carbon materials from the scratched surface and the effect of the increased electron density of the surface hydroxyl group of FeOH under the light irradiation.

Keywords: thermally assisted photoelectron emission; XPS; real iron; scratch-inducing environment; Arrhenius activation energy; environment molecule-surface hydroxyl group interaction

1 Introduction

Much attention has been given to studying the role of charge transfer and electronic excitation in the overlayers of real metallic surfaces in various problems in tribology, adhesion, and corrosion. However, little is known about the behavior of electrons in these superficial layers, owing to the difficulties associated with experimental measurements of such electrons. Photoelectron emission (PE) occurs when photons are absorbed by electrons, leading to the release of these photon-absorbing electrons from the surface. For real iron surfaces, however, little electron emission is observed at temperatures close to room temperature, even in response to irradiation by photons with the energy higher than the metal’s work function. This is caused by the effect of the surface overlayer. The surface overlayer consists of adsorbed species such as

* Corresponding author: Yoshihiro MOMOSE, E-mail: y.momose@cpost.plala.or.jp
oxides, hydroxyl groups, carbon materials, and reaction products, owing to the metal’s interaction with its environment. To understand the trapping or transport of electrons excited from the metal base on the surface overlayer, we have studied the temperature dependence of PE from real iron surfaces, using a non-vacuum-based method. The theoretical basis of PE is well-captured by the Spicer’s three-step model [1], which prescribes that the emission process occurs in three successive steps: (a) optical excitation, (b) transport of the excited electrons to the surface, and (c) release of the excited electrons into vacuum. We previously obtained the activation energies of quantum yields and electron PE probabilities, for PE from real iron surfaces cleaned only in acetone. These quantum yields and probabilities were obtained for different temperatures and wavelengths, using the PE analysis of Fowler [2] and DuBridge [3, 4] and XPS analysis [5, 6]. The observed reduction in the PE intensity at low temperatures was attributed to the ability of the surface overlayer to transmit the incident light, which is the rate-limiting step in the PE process. Therefore, we believe that the surface overlayer assumes two roles in PE: (a) controlling the penetration or absorption of the incident light and (b) controlling the transfer of excited metallic electrons to the overlayer and their subsequent transport to the outermost surface.

We have also previously reported PE and XPS results for scratched real iron surfaces [7], which we briefly summarize here. The metal surfaces were scratched using a diamond cutter, in seven environments: air, water, methanol, ethanol, acetone, benzene, and cyclohexane. The PE intensity was then measured as a function of temperature in response to the 210 nm wavelength irradiation, using a Geiger counter. Temperature sweeps (in the 25–339 °C range) were performed in two cycles of temperature increase and subsequent reduction. The first cycle consisted of Up1 and Down1 scans, while the second cycle consisted of Up2 and Down2 scans. PE glow curves (PE intensity vs. temperature) were obtained, with four glow curves for each scratched sample, as shown in Fig. 1 in Ref. [7]. The PE glow curves that were obtained in the Up1 scans were found to be strongly affected by the environment in which scratching was performed, while those that were obtained in the other scans exhibited a similar behavior and were almost environment-independent. The Up1 scan PE glow curves for temperatures in the 40–339 °C range, and the XPS results that were obtained at 25, 200, and 339 °C, revealed the following: (a) The PE intensity gradually increased with temperature, with a broad convex swelling peak in the 150–250 °C range, depending on the environment. (b) At 40 °C, the PE intensities in the different environments satisfied: air < water ≈ methanol ≈ cyclohexane < ethanol < benzene < acetone. (c) For all environments except acetone and benzene, the PE intensity in the Up1 scan significantly increased with increasing the presence of two oxygen components, OH and O²⁻, in the overlayer, \( Z_\text{O} = O^2^-/(OH + O^2^-) \), up to 200 °C. Above that temperature, the PE intensity increased more gently until 339 °C. (d) The Arrhenius plot of the \( Z_\text{O} \) values for samples scratched in methanol and cyclohexane yielded a straight line, corresponding to the \( \Delta E_{Z_\text{O}} \) activation energies of 0.023 eV (methanol) and 0.043 eV (cyclohexane). These values were much lower than the value of 0.113 eV that was obtained for real iron surfaces cleaned only in acetone [6]. These \( \Delta E_{Z_\text{O}} \) values will be compared with the \( \Delta E_a \) values obtained in the present study. (e) The PE intensity increased with the decreasing presence of Fe as one of Fe components in the overlayer, for all environments except benzene in the 200–339 °C range. (f) With the increasing FeOOH content, the PE intensity increased in the 25–200 °C range, but in the 200–339 °C range the PE intensity for air, water, methanol, ethanol, and cyclohexane environments increased with almost the same content of FeOOH, while for acetone and benzene environments the PE intensity remained almost constant.

The goals of the present study were: (a) to establish a simple method of estimating \( \Delta E_a \) from the PE glow curve; (b) to determine \( \Delta E_a \) during the increase in the PE quantum yield with increasing temperature, for all temperature scans; (c) to clarify the relationship between \( \Delta E_a \) and the total count of emitted electrons in the Up1 scan, for different environments; (d) to clarify the relationship between \( \Delta E_a \) and the PE intensity in the Up1 scan and the acid–base interaction of the environment’s liquid molecules with surface hydroxyl groups (Fe–OH) and with the \( \Delta E_{Z_\text{O}} \) of \( Z_\text{O} \); (e) to clarify the relationship...
between the total count of emitted electrons in the Up2 scan and both $\Delta E_a$ and the intensity of iron oxide in the Fe3p XPS spectra; (f) to determine the mechanism of convex swelling in the PE glow curves.

2 Experimental

2.1 Materials

The metal samples were 99.5% purity 0.1-mm-thick commercial rolled iron sheets (Nilaco, Tokyo, Japan). The samples' dimensions were 20 mm x 30 mm. Prior to use, the samples were ultrasonically cleaned in 30 mL acetone (reagent grade) for 15 min twice (i.e., total cleaning time of 30 min), dried in vacuum for 15 min, and then kept in a desiccator. An as-cleaned sample, i.e., only cleaned in acetone, was used as a reference for $\Delta E_a$ values obtained from both $Y_{ST}$ and $Y_{GC}$. The cleaned samples were scratched as follows. The sample to be scratched was placed in ambient air, distilled water, or an organic liquid (reagent grade) in a glass Petri dish, and the entire surface of the sample was then scratched uniformly using a manually operated diamond cutter; the scratching was performed at room temperature and lasted 5 min. The organic liquids used were methanol (CH$_3$OH), ethanol (C$_2$H$_5$OH), acetone ((CH$_3$)$_2$CO), benzene (C$_6$H$_6$), and cyclohexane (C$_6$H$_{12}$). Finally, the scratched surfaces were dried in vacuum for 15 min. In addition to measuring the PE glow curves for the scratched surfaces, we also measured the PE glow curve for the unscratched sample.

2.2 Thermally assisted PE and XPS

Figure 1 schematically shows the experimental setup for measuring TAPE [5]. The system consisted of an electron measuring chamber filled with a flowing Q gas, an electron counting system, a light irradiation system, and a heating system. The measuring chamber was basically a gas-flow type Geiger–Müller counter, consisting of an earthed cylindrical brass cathode (diameter, 40 mm) with a brass grid (mesh gauze, 30), and a tungsten wire ring anode (wire diameter, 100 μm). A stainless sample holder with a heater and a chromel-alumel thermocouple was connected to a temperature controller (Rigaku, Akishima, Tokyo). The sample holder was vertically installed 20 mm below the grid. A quartz glass window was inserted into the side wall of the apparatus for light irradiation. The light irradiation system consisted of a grating monochromator and a wavelength drive unit connected to a UV light source. The light source was a deuterium lamp (Hamamatsu Photonics). In all experiments, the direction of the light irradiation was approximately normal to the plane of the analyzed sample. The sample's area exposed to the incident light was 0.5 mm x 3 mm. In the electron counting the voltage of 1,400 V was applied to the anode. The electron counting process can be explained as follows. First,
the electrons that are released from the sample surface in response to light irradiation are attracted toward the earthed grid, because the sample is negatively biased relative to the grid. Next, the electrons pass through the grid and reach the anode. Finally, gas multiplication by the Q gas yields more electrons, resulting in the production of a measurable pulse. Thus, the number of the emitted electrons can be counted by a radiation counter and a linear count rate meter.

In the TAPE measurement, as shown in Fig. 1, the intensity of TAPE can be measured in two ways: (1) during a temperature scan under light irradiation at a fixed wavelength and (2) during a wavelength scan at a fixed temperature. This study reports the results for temperature scans at fixed wavelengths. After a sample had been mounted on the sample holder, the Q gas was allowed to continuously flow into the measuring chamber at a flow rate of 100 sccm at atmospheric pressure, with the sample’s surface directly exposed to the Q gas. After that, the anode voltage was applied from the radiation counter (Ohyo koken Kogyo, Tokyo) and the sample was negatively biased at 94 V relative to the earthed grid, using a battery. The wavelengths of the incident light of interest in the present experiment, \( \lambda \), were 200 and 210 nm.

The power of the light was measured using a power meter. For the samples cleaned only in acetone that were used to obtain \( Y_{ST} \) and \( Y_{GC} \), two light wavelengths were used: 200 nm (photon energy = 6.199 eV, light intensity = \( 1.07 \times 10^{11} \) photons/s) and 210 nm (5.904 eV, \( 1.28 \times 10^{11} \) photons/s). For the scratched and unscratched samples that were used to obtain only \( Y_{GC} \), the 210 nm (1.16 \( \times 10^{11} \) photons/s) irradiation was used.

Below, we describe the procedure of TAPE measurements in the temperature scans at fixed wavelengths. The measurements were conducted as follows: (a) for the samples cleaned only in acetone the glow curves were measured up to seven maximal temperatures of \( T_{\text{max}} = 25, 108, 152, 203, 254, 303, \) and 353 °C. Then, the analyzed samples were maintained at the maximal temperature for ~5 min. The heating rate in the glow curves measurements was 20 °C/min. The maximal temperatures were chosen by taking into consideration the fact that desorption and chemical reactions on metal surfaces are well observed in the temperature range that includes these chosen values and the operational temperature of the Geiger–Müller counter; (b) for both the scratched and unscratched samples the glow curves were measured during the first cycle of temperature increase from 25 to 339 °C (Up1 scan, required time: 950 s) and subsequent decrease from 339 to 40 °C (Down1, 1,525 s), and then during a second cycle of increase from 40 to 339 °C (Up2, 900 s) and subsequent decrease from 339 to 25 °C (Down2, 1,548 s). The heating rate in the Up1 and Up2 scans was 20 °C/min, and the cooling rate in the Down1 and Down2 scans was 20 °C/min from 339 to 100 °C, after which the samples were spontaneously cooled to room temperature. The second cycle scans were performed to examine the effect of the previous heating to 339 °C on the glow curves, because the surface overlayer is considered to attain a thermally stabilized structure after the first cycle of heating.

The results of the standard XPS measurements were previously reported in detail only for samples cleaned in acetone [5] and scratched in the different environments [7]. Here, we briefly describe the XPS analysis of the scratched samples. XPS was performed at 25 °C after the PE wavelength scan from 300 nm to 200 nm at 25 °C, and after cooling to 25 °C following the wavelength scans from 300 nm to 200 nm at 200 °C and 339 °C [7]. The XPS measurements were conducted once for each experimental condition. Fe2p, O1s, C1s, and Fe3p core spectra were measured using a Shimadzu ESCA 750 spectrometer (Shimadzu, Kyoto, Japan) with an Mg K\( \alpha \) X-ray source (8 kV and 30 mA). The energy range and sensitivity factors were the same as described previously [5], except that the energy for Fe3p ranged from 65 eV to 45 eV. The O1s spectra were resolved by curve-fitting using the Gaussian function, into two main components that were attributed to the hydroxyl (OH\(^-\)) and oxide (O\(^{2-}\)) groups. The Fe3p and Fe2p spectra were resolved into four components of FeOOH, Fe\(_2\)O\(_3\) (Fe\(^{3+}\)), FeO (Fe\(^{2+}\)), and Fe metal using the Gaussian and Lorentzian functions and the peak shape analysis method of Hesse et al. [8]. From the temperature dependence of the Fe3p and Fe2p spectra, the resolution of the Fe3p spectra was found to be better than that of the Fe2p spectra [7]. Therefore, the intensities of the four components in the Fe3p spectra were used in the following analysis. It is noted that the main peak in the C1s spectra appeared at the binding energy of 285 eV.
3 Results

3.1 Method to determine the activation energy from the PE quantum yield glow curve

Conventionally, the PE quantum yield (emitted electrons/photon), \( Y \), called \( Y_{\text{ST}} \), is obtained from the PE intensity measured when the final temperature (\( T \)) of the PE glow curve is held stationary, and then the logarithm of \( Y_{\text{ST}} \) is plotted against \( 1/T \) (\( T \) in kelvin) to obtain \( \Delta E_a \) \[5\]. However, this method becomes time-consuming when many glow curves need to be measured. In contrast, in the simple method proposed here, the PE intensities are obtained by acquiring only one glow curve up to a maximal temperature, and then by setting the values of \( Y \), called \( Y_{\text{GC}} \), at selected temperatures (\( T \)). \( \Delta E_a \) is subsequently obtained by plotting the logarithm of \( Y_{\text{GC}} \) against \( 1/T \). It is shown below that the \( \Delta E_a \) values obtained by the conventional and simple methods are in a good agreement, and therefore our simple method can be used to obtain \( \Delta E_a \).

Figure 2 shows electron emission intensities measured under irradiation by light with the wavelengths of 200 nm and 210 nm, as a function of time during temperature scanning, for a sample cleaned only in acetone. The emission intensity gradually increased with time and then saturated above the different times denoted by arrows that correspond to different \( T_{\text{max}} \). Both curves (a) to \( T_{\text{max}} = 353 \, ^{\circ} \text{C} \) shown in Figs. 2(a) and 2(b) also clearly exhibited a broad small convex swelling at ~600 s. The saturated emission intensity increased with \( T_{\text{max}} \) and became constant at \( T_{\text{max}} \). The saturated emission intensity was much higher at 200 nm than at 210 nm. \( T_{\text{max}} \) was a stationary temperature, so the PE quantum yield in this case was denoted by \( Y_{\text{ST}} \). The Arrhenius-type equation for the sets of \( Y_{\text{ST}} \) and \( T_{\text{max}} \) under the irradiation by light with a given \( \lambda \) is

\[
Y_{\text{ST}} = Y_0 \exp(-\Delta E_a/k_b T)
\]

where \( Y_0 \) is the pre-exponential factor, \( \Delta E_a \) is the activation energy, \( k_b \) is Boltzmann’s constant, and \( T \) is the temperature of \( T_{\text{max}} \) in kelvins. Figure 3 shows Arrhenius plots of ln(\( Y_{\text{ST}} \)) against the reciprocal of \( T_{\text{max}} \) for both \( \lambda \). The \( Y_{\text{ST}} \) values were determined from the saturated emission intensity and the incident photon intensity (units: photons/s), in the units of the number of emitted electrons per photon, using emitted electrons instead of counts for the saturated emission intensity. The Arrhenius plots were nearly linear over the entire range of temperatures, from 25 °C to 353 °C,
for both $\lambda$, indicating that the PE quantum yield, $Y_{ST}$, obeyed the Arrhenius-type equation. The $\Delta E_a$ and $Y_o$ values obtained from $Y_{ST}$ are listed in Table 1 (Table 1 in Ref. [5]).

We next describe the proposed simple method for determining the activation energy. Figure 4 shows the curves of the PE quantum yield, $Y$, as a function of temperature during temperature scanning to 353 °C, for the irradiation light wavelengths of 200 nm and 210 nm. These $Y$ glow curves correspond to the curves (a) in Figs. 2(a) and 2(b). The Arrhenius plots for the set of $Y_{GC}$ values obtained at 40, 100, 150, 200, 250, 300, and 339 °C, selected from the $Y$ glow curves at 200 nm and 210 nm, are shown in Fig. 5. The data points clearly fall on a straight line for both wavelengths. In Table 1, the activation energy and the pre-exponential factor, $\Delta E_a$ and $Y_o$, respectively, obtained from $Y_{ST}$ and $Y_{GC}$ are listed. The $\Delta E_a$ values obtained from $Y_{GC}$ were 14% at 200 nm and 6% at 210 nm, higher than those that were obtained from $Y_{ST}$, and the $Y_o$ values obtained from $Y_{GC}$ were also somewhat higher than those obtained from $Y_{ST}$ for both wavelengths. Considering the deviation of the $\Delta E_a$ values at 210 nm, we believe that $Y_{GC}$ can be used for determining the value of $\Delta E_a$ for the PE quantum yield. In addition, it is interesting that the light wavelength affects $Y_{GC}$; the $\Delta E_a$ values were higher for the wavelength of 200 nm than for the wavelength of 210 nm, in the same way as those

<table>
<thead>
<tr>
<th>Type of the quantum yield</th>
<th>Wavelength ($\lambda$/nm)</th>
<th>Photon energy ($h/\nu$/eV)</th>
<th>Light intensity ($I_h/10^{11}$photons/s)</th>
<th>Activation energy ($\Delta E_a$/eV)</th>
<th>Pre-exponential factor ($Y_o/10^{-8}$emitted electrons/photon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{ST}$</td>
<td>200</td>
<td>6.199</td>
<td>1.07</td>
<td>0.112</td>
<td>1.47</td>
</tr>
<tr>
<td>$Y_{GC}$</td>
<td>200</td>
<td>6.199</td>
<td>1.07</td>
<td>0.128</td>
<td>1.88</td>
</tr>
<tr>
<td>$Y_{ST}$</td>
<td>210</td>
<td>5.904</td>
<td>1.28</td>
<td>0.100</td>
<td>0.57</td>
</tr>
<tr>
<td>$Y_{GC}$</td>
<td>210</td>
<td>5.904</td>
<td>1.28</td>
<td>0.106</td>
<td>0.69</td>
</tr>
</tbody>
</table>

$Y_{ST}$ and $Y_{GC}$ mean the quantum yields at stationary temperatures and at temperatures selected in one glow curve up to 353 °C, respectively.

Fig. 4 PE quantum yield, $Y$, as a function of temperature during the temperature scan to 353 °C, at a constant heating rate under light irradiation with $\lambda = 200$ nm (a) and 210 nm (b). Samples cleaned only in acetone without scratching were used.

Fig. 5 Arrhenius plots of $\ln$(PE quantum yield, $Y_{GC}$) at selected temperatures in the glow curves for samples cleaned only in acetone without scratching. The $Y_{GC}$ values were determined from the electron emission intensities at 40, 100, 150, 200, 250, 300, and 339 °C in the glow curves shown in Fig. 4 and the photon intensity at $\lambda = 200$ nm (a) and 210 nm (b). The equations and $R^2$ of the fitted straight lines are given in the figure.
obtained from \( Y_{ST} \) [5], although the reason for this remains unclear.

### 3.2 Determination of activation energy for scratched real iron surfaces

Figure 6 shows typical PE quantum yield glow curves obtained for the irradiation wavelength of 210 nm, for a sample scratched in air. In the figure, the quantum yield, \( Y \), in units of emitted electrons/photon is shown vs. the scan temperature, for the Up1, Down1, Up2, and Down2 scans. The glow curves for the Up 1 and Up2 scans, and those for the Down1 and Down2 scans, are shown together in the 40–339 °C range. The PE intensity glow curves for all scans of samples scratched in all environments and unscratched samples have been reported previously [7], in units of counts/min. Figures 7(a) to 7(h) show Arrhenius plots for the data obtained from \( Y_{GC} \) at seven selected temperatures in the Up1 and Up2 scans, for all scratched and unscratched samples. In these plots, \( \ln(Y_{GC}) \) is plotted against \( 1/T \) on the same scale. In the case of the Up1 scan for environments such as water, methanol, and cyclohexane, the data points considerably deviated from a straight line owing to the broad convex swelling peak, while in the case of the Up2 scans, the data points fell on a straight line. We note that in the present study, a straight line approximation was needed as a first step toward examining the primary dependence of \( \ln(Y_{GC}) \) on temperature. Therefore, as shown in Fig. 7, the data obtained in all scans were approximated by a straight line, and the values of \( \Delta E_a \) and \( Y_0 \) were obtained from the slope and the intercept of the straight line, respectively. The broad convex swelling appearing in some of the PE glow curves is addressed later.

Table 2 summarizes the \( \Delta E_a \) values obtained from the Up1, Down1, Up2, and Down2 scans and the \( Y_0 \) values obtained from the Up1 scan, for the scratched and unscratched samples. The activation energies for the oxygen component ratio, \( Z_O = \frac{O^2}{OH^- + O^2} \), \( \Delta E_{ZO} \) for methanol, ethanol, and cyclohexane are also given. The \( \Delta E_{ZO} \) values will be discussed later.

\[ \Delta E_{aUp1} \] was in the 0.212–0.035 eV range, while \( \Delta E_{aDown1}, \Delta E_{aUp2}, \) and \( \Delta E_{aDown2} \) were in the 0.038–0.012 eV range, and \( Y_{0Up1} \) was in the \((5.77 \pm 0.22) \times 10^{-8}\) range, with all units corresponding to the number of emitted electrons per photon. The following behavior was observed:

(a) \( \Delta E_{aUp1} \) was strongly environment-dependent and was considerably higher than \( \Delta E_{aDown1}, \Delta E_{aUp2}, \) and \( \Delta E_{aDown2} \), except in the case of acetone and benzene, for which all of the \( \Delta E_a \) values were nearly the same.

(b) \( \Delta E_{aUp1} \) for the different environments satisfied the following: acetone < benzene < ethanol < cyclohexane < water < methanol < air. The value of \( \Delta E_{aUp1} \) for the unscratched sample was situated between those for the samples scratched in benzene and ethanol.

(c) \( \Delta E_{aDown1}, \Delta E_{aUp2}, \) and \( \Delta E_{aDown2} \) were almost the same across all environments, but the values for water, benzene, and cyclohexane were somewhat higher than the values for air, methanol, ethanol, and acetone.

(d) \( Y_{0Up1} \) was considerably high for air, methanol, and water, but extremely low for acetone and benzene. Furthermore, strikingly, for all environments the plots of \( Y_{Up1} \) vs. \( \Delta E_{aUp1} \) exhibited a parabolic dependence on \( \Delta E_{aUp1} \). Here, it should be noted that the values of \( Y_{GC} \) shown in Fig. 7 and accordingly those of the Arrhenius activation energy and the pre-exponential factor listed in Table 2 were estimated from the experimental data, except for the background count rate. The count rates of the wavelength scans measured from 289.8 nm (4.278 eV) to 260 nm (4.768 eV) at 25, 200,
Fig. 7  Plots of ln(PE quantum yield, $Y_{QC}$) against the reciprocal of temperature in kelvins chosen from the glow curves during the Up1 and Up2 scans for the scratched (a–g) and unscratched (h) samples. The environments were (a) air, (b) water, (c) methanol, (d) ethanol, (e) acetone, (f) benzene, and (g) cyclohexane. The equations and $R^2$ of the fitted straight lines are given in the figures. The chosen temperatures were 40, 100, 150, 200, 250, 300, and 339 °C.
and 339 °C were used to determine the background count rates for the scratched samples in the 25–339 °C range of temperatures. In this range of wavelengths, the count rate was constant, which confirmed no PE emission from the samples. The background rate for the scratched samples was in the range of 150–408 counts/min. For the unscratched sample, the wavelength scan data in Ref. [6] were used and the background rate was in the range of 182–248 counts/min.

Considering that the $\Delta E_a$ obtained from the Up1 scan was significantly different from the values obtained from the other scans, as listed in Table 2, the effect of the interaction of the real iron surface with the different environments during scratching was reflected only in the PE of the Up1 scan up to 339 °C. The surface overlayer, which contains oxides, hydroxyl groups, and other reaction products, is considered to be environment-dependent. This environment dependence of $\Delta E_{a,\text{Up1}}$ will be explained later, based on the acid–base interaction. In contrast, during the Down1, Up2, and Down2 scans the surface overlayer was considered to preferentially contain surface oxides grown during the heating to 339 °C in the Up1 scan. The finding that $\Delta E_{a,\text{Down1}}$, $\Delta E_{a,\text{Up2}}$, and $\Delta E_{a,\text{Down2}}$ were nearly constant across the different environments may suggest that a thermally stabilized structure emerged when heating to the final temperature during the Up1 scan, and this structure persisted. The relation between $\Delta E_{a,\text{Up2}}$ and the intensities of the FeO and Fe$_2$O$_3$ components of the Fe3p spectra observed after heating at 339 °C will be described later. We believe that the values of $\Delta E_{a,\text{Down1}}$, $\Delta E_{a,\text{Up2}}$, and $\Delta E_{a,\text{Down2}}$ may be associated with the transport of electrons through the oxide layer. However, for the Down1, Up2, and Down2 scans, $\Delta E_a$ was very close to the difference between the thermal energy values represented by $k_B T$ at 25 °C and 339 °C, 0.025 eV and 0.053 eV, respectively, which equals 0.028 eV. Therefore, the activation energy obtained from the Down1, Up2, and Down2 scans may also originate from thermal excitation of electrons in the proximity of the Fermi level of the metal itself.

### 3.3 Temperature dependence of XPS results for scratched real iron surfaces

Figures 8 and 9 show the surface O1s and C1s composition and the intensity of the four Fe3p components, respectively, at 25, 200, and 339 °C. The surface composition was determined from the O1s, C1s, and Fe3p spectra, and the Fe3p components corresponded to the Fe metal, FeO, Fe$_2$O$_3$, and FeOOH. The following conclusions can be made by inspecting Fig. 8(a) regarding the environment dependence of the O1s and C1s compositions at 25 and 339 °C, at 25 °C the O1s composition decreased from 36.8% (water) to 23.7% (cyclohexane), while the C1s composition gradually increased from 0.025 eV and 0.053 eV, respectively, which equals 0.028 eV. Therefore, the activation energy obtained from the Down1, Up2, and Down2 scans may also originate from thermal excitation of electrons in the proximity of the Fermi level of the metal itself.
effect of scratching in the liquid environments on the surface compositions of the samples was much stronger at the lower temperature than that at the higher temperature. (b) At 25 °C and 200 °C, there was little change in the O1s and C1s compositions for the samples scratched in air and water, while in the case of the other environments, the O1s composition increased and the C1s composition decreased more clearly. This finding implies that, compared with the other environments, in the case of air and water oxygen species and carbon materials were stronger bound to the metal surface. (c) As the temperature increased...
from 200 °C to 339 °C, in the case of air, water, ethanol, and acetone, the O1s composition increased and the C1s composition decreased, while the opposite trend was observed for methanol, benzene, and cyclohexane. This finding suggests that the ability of oxygen species and carbon materials to remain on the sample surface for temperatures in the 200–339 °C range was different in the different environments. The effect of co-adsorbed oxygen species and carbon materials on the convex swelling peak observed in the Up1 scan PE glow curves will be described later.

In Fig. 9, we focused on analyzing the intensity of the FeO and Fe$_2$O$_3$ components at 339 °C. The intensity of both oxides formed at 339 °C was strikingly environment-dependent. The FeO intensities for the different environments satisfied the following relation: methanol < benzene < acetone < ethanol < air < water < cyclohexane, while the Fe$_2$O$_3$ intensities satisfied the following order: benzene < cyclohexane < acetone < water < ethanol < air < methanol. The relationship between these orders and both $\Delta E_{a\text{Up}1}$ and the total count of electrons emitted during the Up2 scan will be discussed later.

### 4 Discussion

#### 4.1 Activation energy of PE in the Up1 scan

It is believed that $\Delta E_{a\text{Up}1}$ was determined by the surface interactions on the surface overlayer formed during scratching in the selected environments and its temperature dependence. We have examined the dependence of $\Delta E_{a\text{Up}1}$ on the total count of electrons emitted during the Up1 scan and the acid–base interactions in liquid environments. Figure 10 shows the relationship between the total count of electrons emitted during the Up1 scan and $\Delta E_{a\text{Up}1}$ for the different environments. Interestingly, the total count of emitted electrons decreased approximately linearly with increasing $\Delta E_{a\text{Up}1}$. This finding reveals that the thermal activation process importantly shapes the PE intensity.

We reconsidered the two roles of the surface overlayer in the PE mentioned above, based on the values of $\Delta E_{a\text{Up}1}$ and $\Delta E_{ZO}$ for methanol, ethanol, and cyclohexane listed in Table 2. The oxygen component ratio $Z_O$ is given by $Z_O = O^2-/(OH^- + O^2-)$. The $\Delta E_{ZO}$ values were obtained from the straight line in the Arrhenius plot of ln($Z_O$) against $1/T$. The $\Delta E_{ZO}$ values for methanol and cyclohexane were reported previously [7], while that for ethanol was obtained in the present study. For these environments, $\Delta E_{a\text{Up}1}$ was much higher than the $\Delta E_{ZO}$. In a previous paper [6], it was concluded, based on the similarity between the $\Delta E_a$ and $\Delta E_O$ values for real iron surfaces cleaned only in acetone, that the penetration of the incident light through the surface overlayer is the first step in the PE process. The present case, however, suggests that $\Delta E_{a\text{Up}1}$ cannot be attributed only to $Z_O$. Therefore, we think that $\Delta E_{a\text{Up}1}$ may be related to the transfer of electrons excited in the metal to the surface overlayer and subsequent transport through the overlayer, rather than to the penetration of incident light.

As suggested by Table 2, for all environments the plots of $Y_{a\text{Up}1}$ vs. $\Delta E_{a\text{al}p}$ exhibited a parabolic increase of $Y_{a\text{Up}1}$ with $\Delta E_{a\text{al}p}$. This behavior may be explained by the ability of the overlayer formed during scratching in the different environments to hold electrons. The TAPE process is considered as follows: first, excited electrons are formed in the light-irradiated area on the metal surface; next, the transfer of the excited electrons to the overlayer and their subsequent transport through the overlayer occur. Therefore, it is presumed that, in the case of air (which yields the highest $\Delta E_{a\text{Up}1}$), the excited electrons may be strongly bound in the light-irradiated area of the overlayer, so that the transfer and transport of excited electrons.
becomes difficult, while in the case of acetone (which yields the lowest $\Delta E_{\text{aUp1}}$), the excited electrons may be loosely bound in the irradiated area, allowing them to move easily. Thus, in the case of air, the values of both $Y_{\text{aUp1}}$ and $\Delta E_{\text{aUp1}}$ are higher than those in the case of acetone.

Regarding the thickness effect of the overlayer on $\Delta E_{\text{aUp1}}$, we considered that the amount of the percentage reduction in the Fe-metal (%) from 25 °C to 339 °C (that is, $\Delta \text{Fe-metal} (%) = \text{Fe-metal at 25 °C} - \text{Fe-metal at 339 °C}$) may be related to $\Delta E_{\text{aUp1}}$, because $\Delta \text{Fe-metal}$ is equal to the increase in the sum of FeO (%), Fe$_2$O$_3$ (%), and FeOOH (%) from 25 °C to 339 °C. We examined the relationship between $\Delta E_{\text{aUp1}}$ and $\Delta \text{Fe-metal} (%)$ using the Fe-metal values shown in Fig. 9(a). In the case of water (with $\Delta E_{\text{aUp1}} = 0.145$ eV), Fe-metal (%) reduced from 25.8 to 9.5, becoming $\Delta \text{Fe-metal} (%) = 16.3$, while in the case of C$_6$H$_{12}$ (with $\Delta E_{\text{aUp1}} = 0.111$ eV), Fe-metal (%) reduced from 58.2 to 11.0, becoming $\Delta \text{Fe-metal} (%) = 47.2$. This finding suggests that the relation between $\Delta \text{Fe-metal}$ and $\Delta E_{\text{aUp1}}$ is weak. We posit that the chemical and electrical properties of the overlayer more strongly affect $\Delta E_{\text{aUp1}}$ than the overlayer’s thickness.

### 4.2 Effect of acid–base properties of liquids molecules on PE in the Up1 scan

Figures 11(a) and 11(b) show the dependence of $\Delta E_{\text{aUp1}}$ on the acceptor number [9] and proton affinity [10] of the liquid environment, respectively. Interestingly, $\Delta E_{\text{aUp1}}$ increased with increasing acceptor number, but tended to decrease with increasing proton affinity, although the data points were rather scattered. This finding indicates that molecules with higher acceptor number suppressed the PE intensity, while molecules with higher proton affinity promoted the PE intensity.

Based on the acid–base properties of molecules, we propose that the effect of electric charges formed in the overlayer is among the most important surface properties affecting the PE. In our previous study [7], based on the work of Bolger and Michaels [11], we explained the acid–base interaction of surface hydroxyl groups (–MOH) with the environment molecules. The interaction mode can be changed by varying the acid–base properties of the molecules, but as described later, the electron density of the oxygen atom of –MOH can also play an important role in the acid–base interaction. In the present experiment, MOH was represented by FeOH from the FeO(OH) component. We examined the surface interactions between the FeOH and the different liquid environments based on their acceptor number, which represents the Lewis acidity or electrophilicity of a solvent, and proton affinity, which represents a measure of the tendency of a sample molecule to accept a proton or the gas-phase basicity.

Figure 12 shows examples of orientations of electric dipoles formed by the acid–base interaction modes between the molecules of (CH$_3$)$_2$CO, C$_2$H$_5$OH, and H$_2$O, and the surface hydroxyl groups, based on Ref. [11]. In Fig. 12, as the molecules are considered from left to right, the acceptor number increases (Fig. 12(a)) and the proton affinity increases (Fig. 12(b)). We explain the interaction modes of H$_2$O, which has the highest acceptor number, and (CH$_3$)$_2$CO, which has the highest proton affinity, with FeOH. In the case of
H₂O (Fig. 12(a)), the H atom of H₂O is attracted to the unshared electron pair of the O atom of \( \text{FeOH} \), producing an electric dipole with its negatively charged end oriented toward the outside, represented by \( \text{FeOH} \cdots \text{H}^+ \text{OH}^- \). In the case of \((\text{CH}_3)_2\text{CO}\) (Fig. 12(b)), the H atom of FeOH is attracted to the unshared electron pair of the O atom of \((\text{CH}_3)_2\text{CO}\), producing an electric dipole with its positively charged end oriented toward the outside, represented by \( \text{FeO}^\cdots \\text{H}^\cdots \text{O} = \text{C}(\text{CH}_3)_2 \). Here, the symbol \( \cdots \) denotes the hydrogen bond. Thus, these orientations of the electric charges contribute to the increase or reduction in \( \Delta E_{\text{aUp1}} \) respectively, which significantly affects the PE intensity. In addition, the finding that the total count of electrons emitted in the Up1 scan is low in the case of water, but high in the case of acetone (Fig. 10) can be reasonably explained in terms of these electric charge orientation modes. In addition, we examined the relationship between \( \Delta E_{\text{aUp1}} \) and the dielectric constants \( (\varepsilon/\varepsilon_0) \) [12] of the liquids: H₂O \( (\Delta E_{\text{aUp1}} = 0.145 \text{ eV}, \varepsilon/\varepsilon_0 = 81.7) \), CH₃OH \( (0.159 \text{ eV}, 32.6) \), C₂H₅OH \( (0.079 \text{ eV}, 24.3) \), \((\text{CH}_3)_2\text{CO}\) \( (0.035 \text{ eV}, 20.7) \), and C₆H₆ \( (0.039 \text{ eV}, 2.3) \). The plots of \( \Delta E_{\text{aUp1}} \) vs. \( \varepsilon/\varepsilon_0 \) for the different liquids revealed the tendency of \( \Delta E_{\text{aUp1}} \) to increase with \( \varepsilon/\varepsilon_0 \). This suggests that in the case of water and methanol the value of \( \Delta E_{\text{aUp1}} \) during the Up1 scan may be associated with the liquid’s electrical polarization, although the mechanism remains unclear.

### 4.3 Convex swelling peak in the PE glow curves in the Up1 scan

The PE glow curves observed in the Up1 scan for all environments (Fig. 1 in Ref. [7]) are shown on the same plot in Fig. 13. Convex swelling peaks with different intensities are apparent in all glow curves. The appearance of these environment-dependent convex swelling peaks proves that the effects of liquid

---

**Fig. 12** Examples of orientations of electric dipoles formed by the acid–base interaction modes between the molecules of \((\text{CH}_3)_2\text{CO}\), C₂H₅OH, and H₂O and the surface hydroxyl groups (Fe–OH) based on: (a) the acceptor number, which represents the strength of the acidity, and (b) the proton affinity, which quantifies the tendency of a sample molecule to accept a proton or the gas-phase basicity. The former mode increases \( \Delta E_{\text{aUp1}} \) and the latter decreases \( \Delta E_{\text{aUp1}} \). The dotted lines in the figure indicate the hydrogen bond, although the interaction mode for acetone as an acid is unclear.

**Fig. 13** Photoelectron emission glow curves observed during the Up1 scans, for the different environments.
environments on the scratched surfaces persist up to the temperatures of ~300 °C, even if the adsorbed liquids are partially evaporated. The presence of these convex swelling peaks may have been caused by the following two emission modes: (a) the direct emission of the electrons excited from the metal base without being trapped in the overlayer and (b) the emission of electrons via electron trapping sites formed in the overlayer after their excitation by the incident light. In fact, the glow curves are considered to include the electrons emitted via both emission routes, although the mechanism remains unclear [7].

We examined more deeply the mechanisms underlying the convex swelling peaks, emphasizing the effects of surface-bound carbon materials and electron density on the oxygen atom of the Fe–OH group under the irradiation by light. First, we considered the effect of the surface composition (Fig. 8) on the PE. The magnitude of the swelling of the glow curve, MSG, was defined, for comparing the environment dependence of the MSG to that of the C1s composition shown in Fig. 8(b). For each Up1 scan PE glow curve, a straight line was drawn to connect the electron emission intensities at 40 °C and 339 °C, and the emission intensities at 40, 100, 150, 200, 250, 300, and 339 °C were estimated from this straight line by interpolation. Each of these emission intensities was then subtracted from the actual emission intensity observed at the same temperature. The difference between the two emission intensities yielded the MSG. The idea of drawing a straight line between the electron emission intensities at 40 °C and 339 °C in the PE glow curve came from the results obtained for the wavelength scans from 300 nm to 200 nm and measured at 25, 200, and 339 °C. These results were as follows: (a) the total count of electrons emitted during the wavelength scans increased approximately linearly with temperature. (b) The plots of the total count of emitted electrons vs. temperature were qualitatively similar for all environments. Thus, the PE intensity measured in the wavelength scans was assumed to be unaffected by the environment and the internal photoelectric effect, which can be speculated to have occurred under the continual light irradiation at 210 nm during the PE glow curve measurements.

Figure 14 shows the plots of MSG vs. temperature. Except in the case of air, all plots feature a broad peak in the temperature range of 100–300 °C. In the case of air, the MSG values are negative over the entire range of temperatures. In Fig. 14, at 200 °C the MSG for the different environment increases in the following order: air < water < benzene ≤ cyclohexane ≤ methanol ≤ acetone < ethanol, while in Fig. 8(b), the C1s intensities for the different environments, measured at 200 °C, in the order of increasing MSG, are as follows: air (C1s intensity = 57.0%) > water (51.1%) > benzene (45.2%) ≈ cyclohexane (42.1%) ≈ methanol (41.9%) ≈ acetone (47.8%) ≈ ethanol (45.7%). This order indicates that the C1s intensities for air and water were considerably higher than those for the organic environments, which were almost the same. This finding suggests that the thermal removal of carbon materials occurred more easily from samples scratched in organic environments, resulting in the exposure of the inner layer, while the carbon materials were more strongly bound to the surfaces of samples that were scratched in air, and acted to suppress the convex swelling peak.

In general, when solids are irradiated with light, electrical effects such as the external photoelectric effect (called photoemission or PE) and internal effects (such as photoconduction) can be observed. The temperature dependence of the PE intensity under irradiation with the wavelength of 210 nm was significantly different across the different environments (Fig. 13), while the
total count of electrons emitted during the wavelength scan was approximately the same. Therefore, we posit that, as an internal photoelectric effect, the electron density on the O atom of Fe–OH may be increased during irradiation with light, enhancing the acid–base interaction of Fe–OH with the molecules of the liquid environment. Momose et al. [13] reported the effect of light on optically stimulated exoelectron emission (OSEE) from commercial aluminum surfaces, which was measured using a Geiger counter with a counter gas consisting of Ar and a small amount of C₂H₅OH vapor. They found that the plot of the OSEE intensity against the power of the stimulating light exhibited a maximum, and explained that this behavior could be attributed to the alteration of an electric dipole formed by ethanol molecules adsorbed on Al–OH of the hydroxylated oxide layer on the metal surface. As the electron density of the O atom in Al–OH increases owing to the irradiation, the ability of the O atom in Al–OH to attract the H atom in the hydroxyl group of C₂H₅OH (that is, the basicity of the O atom of Al–OH) increases. Therefore, the acid–base interaction mode of C₂H₅OH with Al–OH changes from Al–O⁻–H⁺–HOC₂H₅ to Al–O⁻–H⁺–HOC₂H₅ in which C₂H₅OH acts as a base, to Al–OH–H⁻–O⁻C₂H₅ in which C₂H₅OH acts as an acid, producing a maximum in the OSEE intensity.

In the present study, we believe that the acidity or basicity of not only the liquid molecules but also the Fe–OH group played an important role in surface interactions. Let us consider the adsorption of water and acetone as examples. Owing to the increase in the electron density of the O atom in Fe–OH under the irradiation, the acidic interaction with water, which has a high acceptor number, is enhanced, leading to a decreased PE intensity, while the basic interaction with acetone, which has a high proton acceptor, becomes much stronger, enhancing the PE intensity, as shown in Figs. 10(a) and 10(c) in Ref. [7]. We previously reported that the oxygen component ratio, Zo = O²⁻/(OH⁻ + O²⁻), strongly affects the PE intensity depending on the temperature (Fig. 6 in Ref. [7]). In addition, it should be noted that the OH⁻ component is considered to be related mainly to the lattice OH⁻ (which originates from FeOH₂ and FeOOH) and adsorbed OH and H₂O. The O²⁻ component is attributed to the lattice O²⁻, which stems from FeO, Fe₂O₃ and FeOOH [7, 14]. Therefore, among these surface oxygen species FeOH plays a more important role in the interaction with the molecules of liquid environments.

We consider the effect of the surface overlayer on the PE intensity observed in the Up1 scan for C₂H₅OH. The following steps are considered: (a) The species such as adsorbed water, surface hydroxyl groups, and carbon materials weakly bound at 25 °C have little effect on the electron emission intensity. (b) These species gradually desorb with increasing temperature. (c) FeOH then predominantly forms and its H atom is attracted to the O atom from C₂H₅OH as a base, which creates an electric dipole with its positively charged end oriented outwards, causing an increase in the electron emission intensity. (d) Finally, a gradual decrease in the amount of FeOH occurs as a result of the dehydration of the FeOH groups, which leads to an increase in the amount of O²⁻ groups. This creates a negative surface charge, and FeOH having an O atom with increased electron density remains and acts more strongly on C₂H₅OH as an acid. These chemical changes reduce the electron emission intensity. This produces the convex swelling peak observed in the PE glow curves (Fig. 13). It remains unclear whether this model can be related to the formation of electron trapping sites in the overlayer, as assumed above.

Furthermore, it should be noted that the glow curves of the Up2 scan exhibited a nearly monotonic increase in emission with increasing temperature, although that for benzene was almost the same as the Up1 scan [7]. It remains unclear why the swelling peak is present even in the Up2 scan in the case of benzene.

### 4.4 Effect of oxide components and activation energy on PE in the Up2 scan

Figure 15 shows the relationship between the total count of electrons emitted during the Up2 scan and the intensities of the FeO and Fe₂O₃ components at 339 °C, for the different environments. Interestingly, as shown in Fig. 15(a), the total count of emitted electrons decreased nearly linearly with increasing FeO intensity, for all environments except benzene. In contrast, as shown in Fig. 15(b), the total count of emitted electrons tended to increase with increasing Fe₂O₃ intensity, for all environments except acetone.
Fig. 15 Plots of the total count of electrons emitted during the Up2 scan, vs. the intensities of the (a) FeO and (b) Fe$_2$O$_3$ components at 339 °C, for the different environments. The approximate straight lines given by the equation and $R^2$ are shown except for those of benzene in the top figure and acetone in the bottom figure.

In Fig. 7(b) in Ref. [7], we reported that the total count of electrons emitted in the second cycle (i.e., Up2 and Down2 scans) tended to decrease with increasing O$_{1s}$/Fe$_{3p}$ ratio at 339 °C, for all environments except cyclohexane. Here, we would like to correct the subscript in the “Total count” label given in that figure. That is, “Total count$_{Up2}$” should be replaced by “Total count$_{Second cycle}$”. Comparing the results in Fig. 7(b) in Ref. [7] with the effect of FeO on the total count of emitted electrons (Fig. 15(a)), it is presumed that the former results may have originated from the effect of the FeO component at 339 °C, although the order of environments for increasing O$_{1s}$/Fe$_{3p}$ did not completely agree with that for increasing FeO.

Next, we ask why the increase in the FeO component reduces the total count of emitted electrons. With respect to the electrical conductivity of metal oxides, according to Conder [15], FeO is a semiconductor and its conductivity increases with temperature from $5 \times 10^3$ S·m$^{-1}$ (330 K) to $10^4$ S·m$^{-1}$ (500 K). Meanwhile, the conductivity of Fe$_2$O$_3$ is very low, but increases significantly with temperature, from $10^{-36}$ S·m$^{-1}$ (330 K) to $10^{-6}$ S·m$^{-1}$ (500 K). Therefore, an increase in the amount of FeO in the overlayer is expected to increase its electrical conductivity, leading to an increased total count of emitted electrons. However, this explanation does not agree with the present experimental results. Roosendaal et al. [16] reported that pre-annealing at 200 °C under vacuum caused the rate of oxygen incorporation during the room temperature oxidation of Fe(100) to increase by a factor of ~500. This was explained as follows: during annealing at 200 °C, electrons and cations are transported to the surface of the metal; the Fe$^{3+}$ ions present near the surface are reduced to Fe$^{2+}$. In addition, Xue et al. [17] reported that Fe$_2$O$_3$ can be transformed to Fe$_3$O$_4$ after annealing at 800 K without ambient oxygen. Therefore, regarding the electronic properties of the surface overlayer after the Up1 scan to 339 °C, the outermost region contained accumulated electrons, which did not lead to emission. This negatively charged region acted to suppress the transport of electrons excited from the metal to the outermost surface. Judging from the decrease in the total count of emitted electrons with increasing FeO intensity (Fig. 15(a)), this electron accumulation seemed to increase with increasing FeO intensity.

We next consider the values of $\Delta E_{aUp2}$ obtained in the Up2 scan (Table 2). Except for benzene, the $\Delta E_{aUp2}$ values for the different environments satisfied the following: methanol ($\Delta E_{aUp2} = 0.020$ eV) < acetone (0.021 eV) < ethanol (0.026 eV) < air (0.028 eV) < cyclohexane (0.033 eV) < water (0.038 eV). Interestingly, this order was almost the same as that of the reduction in the total count of electrons emitted in the Up2 scan and that of the increase in the FeO intensity, as shown in Fig. 15(a). Therefore, these findings indicate that the total count of electrons emitted in the Up2 scan was closely related to both $\Delta E_{aUp2}$ and FeO intensity, for all environments except benzene. It is surprising that the FeO component, which is adjacent to the metal itself, can contribute to the increase in the PE activation energy. Here, we emphasize that not only the FeO intensity but also the electronic properties such as electron accumulation presumed for the FeO become important factors that retard the PE, although the detailed mechanism remains unclear. On the other hand, in Fig. 15(b), an increased Fe$_2$O$_3$ intensity seemed to increase the total count of electrons emitted in the Up2 scan. This behavior may be related to the reduction...
in $\Delta E_{\text{aUp2}}$. In this case it seems that the action of Fe$^{2+}$ ions in Fe$_3$O$_4$, which is considered to be formed from Fe$_2$O$_3$ at the outer surface at 339 °C, contributed to the reduction in $\Delta E_{\text{aUp2}}$ for all environments except acetone. Finally, regarding the deviations of the benzene and acetone data from the straight line in Figs. 15(a) and 15(b), respectively, the following should be noted: In Fig. 9(d), the FeOOH intensities for benzene and acetone at 339 °C are considerably higher than those for the other environments (benzene 42.5%, acetone 42.4%). Therefore, the increase in the FeOOH intensity led to the lowering of the FeO and Fe$_2$O$_3$ intensities. In addition, in our previous paper [7] we reported that for benzene and acetone the electron emission intensity measured at 200 °C and 339 °C in the Up1 scan was almost the same, in spite of the increase in the FeOOH intensity from 200 °C to 339 °C. The FeOOH intensity at 339 °C seemed to have little effect on the emission.

In the future, based on the PE results obtained from wavelength scans from 300 nm to 200 nm in the same way as in a previous paper [6], we will report the temperature dependence of the electron photoemission probability, $\alpha A$, where $A$ is identical to the Richardson constant, and of the photothreshold, $\phi$, and their relationship with the activation energies obtained from the Up1 scans described above.

5 Conclusions

We have reported the effect of surface interactions on photoelectron emission (PE) as a function of temperature for real iron surfaces scratched in air, water, and organic environments. The use of several environments and repeated temperature scans revealed a new aspect of the PE activation energy and its relationship to the surface chemical structure analyzed by XPS. The PE was measured using a non-vacuum-based analysis method. First, we demonstrated how the Arrhenius activation energy may be obtained from one PE quantum yield glow curve using samples cleaned only in acetone. Next, PE quantum glow curves were measured for scratched samples for temperatures in the 25–339 °C range, during two cycles of temperature increase and decrease (Up1, Down1, Up2, and Down2 scans), and a simple method of obtaining the activation energies was introduced and used. The activation energy obtained from the Up1 scan was strongly environment-dependent, while those obtained from the other scans were low and almost the same. The total count of electrons emitted during the Up1 scan decreased with increasing activation energy. In addition, the activation energy obtained from the Up1 scan was explained in terms of the orientation of electric dipoles formed by the acid–base interaction between the molecules of the liquid environment and the surface hydroxyl groups. The total count of electrons emitted in the Up2 scan decreased with increasing activation energy and the increasing intensity of the FeO component of the Fe3p spectrum at 339 °C. Furthermore, the convex swelling peak in the PE glow curves in the Up1 scan was explained in terms of the thermal removal of carbon materials weakly bound to the sample surface and the change in the basicity of the FeOH group under light irradiation.

Acknowledgements

The authors would like to thank the Ministry of Education, Culture, Sports, Science and Technology of Japan for supporting this work through a grant in aid.

Open Access: The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Yoshihiro MOMOSE. He is a professor emeritus of Ibaraki University, Japan. He received his bachelor and master degrees in chemistry in 1963 and 1965, respectively, from Tohoku University, Japan. He joined the faculty of Chemical Research Institute of Non-aqueous Solutions, Tohoku University. He earned his PhD degree in 1970 from Tohoku University. After that he worked in the Department of Engineering Chemistry and Materials Science at Ibaraki University as an associate professor and then a professor from 1970 to 2006. His main research interest focuses on the transfer of charge and electrons on the real surfaces in the fields such as tribochemistry, adhesion, and corrosion.

Takao SAKURAI. He received his M.S. and PhD degrees in applied physics from Tokyo Institute of Technology, Japan, in 1969 and 1973, respectively. His current position is an emeritus professor of Ashikaga Institute of Technology. His research interests include curve analyses in the field of thermally stimulated relaxation processes.

Keiji NAKAYAMA. He is the president of the Institute of Meso-technology, Japan. He received his master and PhD degrees in chemical engineering from Tokyo Institute of Technology, Japan, in 1971 and 1974, respectively. He joined to the Mechanical Engineering Laboratory (MEL), AIST, MITI, Japan in 1974. After studying in the field of tribology as a researcher, a senior researcher, and the head of division in MEL, he worked as a professor in the Advanced Research Institute, Chiba Institute of Technology, Japan from 2008 to 2014. He is in the current position since 2014. His main research interest focuses on the basic and application studies of the triboelectromagnetic phenomena, which embrace triboplasma generation, triboemission of charged particles and photons, and tribocharging.
Abstract: In recent years, a new 2D-layered material—black phosphorus (BP)—has been a rising star after the era of graphene owing to its high charge carrier mobility, tunable direct bandgap and unique in-plane anisotropic structure. With the development of the synthesis and modification methods of BP, its extensive applications, e.g., transistors, batteries and optoelectronics have emerged. In order to explore its full potential, research into the tribological properties of BP 2D-layered materials such as lubrication additives and fillers in self-lubricating composite materials would be not only of high scientific value but also of practical significance. In this work, recent advances on the friction and lubrication properties of BP nanosheets made by our group, including the micro-friction properties, the lubrication properties of BP nanosheets as water-based and oil-based lubrication additives, and the friction and wear of BP/PVDF composites will be presented. Finally, the future challenges and opportunities in the use of BP materials as lubricants will be discussed.

Keywords: black phosphorus; two-dimensional (2D) material; lubricant additive; self-lubricating composite materials; friction

1 Introduction

Friction and wear are primary factors in the energy loss and failure of all types of mechanical systems such as engines [1]. Statistics show that friction consumes a third of the world’s primary energy, wear causes approximately 60% damage of the machine parts and more than 50% accidents involving machinery equipment originate from lubrication failure and excessive wear [2]. Thus, effective controls and reduction of the friction and wear of moving machineries, can improve the efficiency and lifetime of mechanical parts, and have already become the greatest challenges for today’s moving mechanical assemblies.

One of the main purposes of lubrication is to reduce friction and wear. The lubricant medium can be either liquid or solid, or even gas on certain occasions [3, 4]. Solid lubricants are usually prepared via physical or chemical vapor depositions on the surface of the friction pair. A solid lubricant usually fails within a certain period owing to the finite lubricant thickness, and it can be easily affected by the environment. Liquid lubricant can function efficiently after the formation of a liquid film between the friction pairs [5]. However, larger relative velocity of friction pairs or relatively high viscosity of the liquid is an important prerequisite for the formation of the liquid lubricant film. Antiwear additives and friction modifying additives are frequently blended with lubricants in order to improve the wear resistance and reduce the fuel consumption. In order to minimize wear and friction, various friction-reducing additives are widely used as extreme pressure additives to lubricant oils. The lubrication film formed by tribo-chemical reaction is the key to providing a tool surface with high resistance to seizure. Otherwise, insufficient lubrication and severe wear will occur for example, at the start and stop stages of a machine. It is apparent that the development of new lubricant materials and lubrication methods for achieving high reliability and long service life of machines is still

* Corresponding authors: Jianbin LUO, E-mail: luojb@tsinghua.edu.cn; Guoxin XIE, E-mail: xgx2014@tsinghua.edu.cn
very essential, owing to the importance of lubrication and the limitations of existing lubrication strategies [6].

Since the first synthesis of graphene in 2004, two-dimensional (2D) nanomaterials represented by graphene, molybdenum disulfide (MoS2) and hexagonal boron nitride (h-BN) have attracted immense interest in various research areas owing to the combination of remarkable electrical, thermal, optical and mechanical properties and their special atomic structures [7–11]. In recent years, many experimental and theoretical works have been conducted on the mechanical and tribological properties of 2D nanomaterials [12–16]. For instance, many interesting nanotribological properties of 2D materials, e.g., the layer-number-dependent effect of film/substrate binding strength, have been revealed. These works paved the way for the potential application of ultrathin lubricant films in micro/nano-devices. In contrast, the research on the macrotribological properties of these materials has been less extensive owing to the numerous influencing factors. Furthermore, extensive research on these 2D materials as lubricant nano-additives [17–20] and nano-fillers [21–23] in composite materials has been conducted.

Among all the attractive 2D materials, black phosphorus (BP) has recently received considerable attention [24–26]. As a new kind of 2D material, it is still unknown whether it can be potentially used in tribological areas. Based on this consideration, relevant systematic research has been conducted including the synthesis and characterization of BP nanomaterials and their tribological properties, demonstrating that such a kind of 2D material is promising as a future high-performance lubricant.

2 Brief introduction of black phosphorus

Phosphorus belongs to the group V of the periodic table and has three major allotropes—white phosphorus (WP), red phosphorus (RP), and black phosphorus (BP). BP is nontoxic and the most thermodynamically stable material among all the allotropes of phosphorus at room temperature [27]. BP was first synthesized in 1914, and has aroused research interest as a promising 2D material since 2014. Among the emerging post-graphene 2D-layered nanomaterials, BP is a rising star and has been intensively investigated as the functional material in electronic devices such as field-effect transistors owing to its unique properties [28, 29]. A detailed description of the basic characteristics of BP and its potential applications will be presented in the subsequent sections.

2.1 Structure of BP

Similar to other 2D materials such as graphite, BP has a layered structure with the single element P, and the layers weakly interact through van der waals forces. However, BP has different structure characteristics as compared to graphite; the phosphorus atom in BP has five valence shell electrons available for bonding, with a valence shell configuration of 3s23p3. Each phosphorus atom in layered BP has three bonding nearby atoms through sp3 hybridized orbitals and one lone pair, forming a puckered honeycomb lattice [30]. The schematic diagram of the crystal structure of BP is shown in Fig. 1. The single-layer BP consists of two atomic layers and two kinds of P–P bonds. The bond length between the nearest P atoms in the same atomic layer is 0.2224 nm, and the bond length between P atoms of different atomic layers is 0.2244 nm [31, 32]. Furthermore, there are two inequivalent directions at the edge of BP: zigzag (along the puckered direction) and armchair (perpendicular to the puckered direction). These anisotropic characteristics of BP will be discussed in the later part of this article. The bond angles of the hexagonal structure are configured as 102.09° and 96.34° from the top view of BP. The individual layers of BP are assembled by weak van der waals forces and the interlayer distance is 5.3 Å [33, 34].

Fig. 1 Schematic diagram of the crystal structure of BP.


2.2 Syntheses of BP nanomaterials

Various kinds of methods have been developed to synthesize BP since Bridgman first successfully converted WP to BP at 200 °C under 1.2 GPa in 1914 [35]. In contrast to other kinds of allotropes of phosphorus, BP shows good stability at normal temperatures and pressures. Several Japanese groups made important contributions to optimize this method in the 1980s, which significantly improved the quality and size of BP single crystals [36–38]. Phase transformation from RP to BP has also occurred under the high-energy mechanical milling (HEMM) condition, which also provides high pressures and temperatures. Park et al. reported the synthesis of BP by HEMM using steel balls and milling vessel for 54 h [39]. Krebs et al. first reported a “low-pressure” synthesis of BP using the catalytic action of Hg on WP at high temperatures in 1955 [40], and this “low-pressure” method was tremendously promoted by Nilges et al. in 2008 [41]. This low-pressure route method to produce high-quality BP included chemical vapor transport reaction method using Au, Sn, RP, and SnI4 as reactant sources [27]. Since rediscovery of BP in 2014, it has attracted immense interest, and the synthesis of BP has achieved progress in that the efficiency of synthesis has been greatly enhanced and the quality of BP materials has been improved dramatically.

Similar to graphene and other 2D materials, the preparation methods of phosphorene or few-layer BP have been comprehensively developed. The first method reported in 2014 was micromechanical cleavage followed by transfer onto SiO2/Si substrate [42]. As mentioned above, BP is a kind of layered material and the interlayer interaction is dominated by weak van der waals forces. Such a structural feature satisfies the applicable range of the top-down procedure. However, single-layer phosphorus was hard to obtain by the mechanical cleavage owing to the strong binding forces among the layers. Lu et al. reported a method using Ar plasma to thin few-layer BP to monolayer phosphorene [43], and the combination of mechanical cleavage and plasma ablation was promising for controlling the layer number of BP.

The aforementioned methods are effective for producing few- and monolayer BP, but not suitable for mass production. In order to avoid this limitation, liquid phase exfoliation was reported as a reliable method to produce low-size flakes of few- and single-layer BP in greater quantities [44]. The stable dispersions of BP were prepared via ultra-sonication, wherein collapsing cavitation bubbles yielded intense tensile and shear stress fields that exfoliated and fragmented bulk crystals. In order to achieve thinner nano-flakes, aprotic and polar solvents such as N, N-dimethylformamide (DMF), and dimethyl sulphoxide were used to form uniform and stable dispersions. As compared with top-down methods, the bottom-up methods are infrequently used. The chemical synthesis of phosphorene has many advantages such as good crystallinity and high purity. Although chemical bottom-up methods have not made a breakthrough, the rapid development of chemical vapor deposition substrate creates a promising future for phosphorene chemical growth [45]. The requirement for high quality, strong controllability and efficiency will further increase with the development of the field of nanoelectronics.

2.3 Basic physic-chemical properties of BP

2.3.1 Physical properties

Contrary to graphene, BP is a semiconductor with a thickness-dependent bandgap. As a result of the interlayer interaction, different thicknesses of BP film correspond to different bandgap values ranging from 0.3 eV in the bulk to 2.0 eV in a single layer [46, 47]. It indicates that any particular value of bandgap within the range can be obtained by controlling the thickness. The tunable bandgap of BP supplies the gap between graphene (0 eV) and transition metal chalcogenides (1.0–2.0 eV) [48–50]. Notably, the energy range of BP covers the near-IR (near-infrared) and mid-IR parts of the electromagnetic wave spectrum, suggesting a promising future in the fields of photo catalysis, photovoltaics, thermoelectric and thermal imaging [51, 52].

Another typical characteristic of BP is the anisotropic properties on the basis of its special crystal structure. Compared with other 2D materials such as Mo- and W-based transition metal dichalcogenides, the anisotropies of BP are much more apparent. The mechanical properties show strong anisotropy along the armchair and zigzag directions. Jiang et al. studied the Young's modulus values of BP via first-principles calculation.
and determined the apparent difference between armchair and zigzag directions [53]. In the case of monolayer phosphorene, the Young’s moduli for the armchair and zigzag directions were 21.9 N·m⁻¹ and 56.3 N·m⁻¹, respectively, and thus, the value along the zigzag direction was approximately twice that along the armchair direction. Considering phonon dispersion, Qin et al. [54] computed the thermal conductivity of BP. At 300 K, the thermal conductivity along the zigzag direction was 30.15 W·m⁻¹·K⁻¹ whereas that along the armchair direction was only 13.65 W·m⁻¹·K⁻¹. Using non-equilibrium molecular dynamics, Zhang et al. [55] observed similar anisotropy in that the thermal conductivities of infinitely large monolayer BP along the zigzag and armchair directions were suggested to be 42.553 W·m⁻¹·K⁻¹ and 9.891 W·m⁻¹·K⁻¹, respectively. They also evaluated the correlation between the thickness and external strain, observed the anisotropic behaviors, and it was determined that the thermal conductivity of BP was layer-independent and the applied strain affected the thermal conductivity in a unique way.

2.3.2 Chemical properties

Influenced by the lone pair of phosphorus atoms, BP is very reactive to air, demonstrating limited ambient stability. Light-induced ambient degradation process of BP forms PₓOᵧ on the surface, which further turns into phosphoric acid. Generally, the degradation of newly exfoliated few-layer BP can be observed after approximately 2 h under normal conditions. The combined effect of visible light and oxygen (O₂) contributes to the degradation of BP, and thus, the progress can be hindered by controlling light or O₂. A recent study showed that BP nanoparticles were stable in deaerated aqueous dispersions for weeks [56], suggesting that water interacted weakly with BP. Furthermore, the oxidized surface could effectively prevent the BP underneath from degrading, and therefore, the bulk BP appears to be stable under normal conditions [57].

2.4 BP applications

Owing to the aforementioned attractive properties of BP, the exploration of relevant applications has made rapid progress in the fields of electronics, field-effect transistors [58, 59], sensors [60], energy storage [60–62], batteries [63], photocatalysis [64, 65], and optoelectronics [66], and other applications.

3 Potential of BP as lubricants

Bulk BP is a layered material in which individual atomic layers are stacked together by van der waals interactions, similar to bulk graphite, MoS₂, and h-BN. BP possesses a unique combination of electronic and optical properties, and thus, it has been intensely investigated for discovering potential applications in electronics and optoelectronics. A solid lubricant is a solid material with low shear strength and high wear resistance. In general, solid lubricants can be used in the forms of powders or coating on the surfaces of contacting pairs, and they can also be dispersed into oils or greases in order to improve the friction and wear performances. Some of the typical features of common solid lubricants are summarized in Table 1.

Few-layer BP which has low shear strength and good thermal stability, is a potential materials using as lubricant additives or composite filler materials. First, BP nanosheets have higher mechanical strength. The ideal elastic moduli of zigzag and armchair BP nanoribbons are approximately 65 and 27 GPa, respectively [26]. Morenomoreno et al. [67] investigated the effects of environmental conditions on the mechanical properties of few-layer BP, and the results indicated that the elastic modulus and the breaking strength of BP nanosheets were 46 ± 10 GPa and 2.1 ± 1 GPa, respectively. The in-plane Young’s modulus of single-layer BP was 41.3 GPa in the direction perpendicular to the pucker and 106.4 GPa in the parallel direction, and the ideal strains were 0.48 and 0.11 in the perpendicular and parallel directions.

Table 1 The thermal stability and the allowed temperature range of the solid lubricant materials used in dry tribo-contacts.

<table>
<thead>
<tr>
<th>Materials</th>
<th>Temperature range (°C)</th>
<th>Temperature of thermal stability (°C)</th>
<th>Coefficient of friction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MoS₂</td>
<td>–184–400</td>
<td>350</td>
<td>0.05–0.25</td>
</tr>
<tr>
<td>WS₂</td>
<td>–454</td>
<td>425</td>
<td>0.05–0.25</td>
</tr>
<tr>
<td>graphite</td>
<td>–184–650</td>
<td>500</td>
<td>0.1–0.3</td>
</tr>
<tr>
<td>PTFE</td>
<td>–184–260</td>
<td>260</td>
<td>0.02</td>
</tr>
<tr>
<td>h-BN</td>
<td>–184–538</td>
<td>700</td>
<td>0.1–0.2</td>
</tr>
</tbody>
</table>
respectively [68]. Second, advancements have been made to enhance the stability of few-layer BP in ambient air [56, 63, 64]. All the properties mentioned above render BP very attractive for achieving low-friction and low-wear regimes. In the following sections, the investigations of the tribological properties of BP including its friction in the microscale, nanocomposite fillers, and lubrication additives will be discussed.

4 BP synthesis

BP, the bulk counterpart of phosphorene, was first synthesized from WP at 200 °C and under 1.2 GPa in 1914 [35]. In recent years, BP was obtained from WP and RP at high temperatures and high pressures. The synthesis of BP has been widely reported; however, in most of the procedures, extreme operating conditions such as high temperature (>200 °C), high pressure (>1.2 GPa), toxic catalyst were employed.

The layered structure of BP is similar to that of graphite, MoS$_2$ and h-BN. In order to expand the application of BP in the tribological field, the cost of preparation should be considered. The high energy mechanical milling (HEMM) technique at room temperature and normal pressure is a simple method for transforming RP into orthorhombic BP [69]. In the HEMM process, the temperature can rise above 200 °C and the local pressure can be of the order of 6 GPa in steel vial by the violent collisions of balls. Under these conditions, the transformation from RP to BP is possible. Thus, in this work the HEMM technique was employed for preparing orthorhombic BP.

4.1 Preparation of BP

BP powders were prepared using the HEMM technique in a Fritsch P7 premium line planetary ball mill at room temperature and normal pressure. RP powder (purity>99.999%) was purchased from Aladdin Ltd. (Shanghai, China). The materials were weighed initially, and placed in a hardened steel vial (80 mL) with 10 mm and 5 mm milling balls (hardened steel). The weight ratio of 10 mm ball to 5 mm ball was 3:1. The weight ratio of milling balls to reagent powders was 30:1. The milling vial was filled with argon gas and set in the ball miller. The milling process was performed for different times. The revolution speed was set to 800 rpm. Milling was performed in a sequence of 30 min milling and 10 min pause. The characterization of the active powder was carried out using powder X-ray diffractometers (XRD; Shimadzu XRD-6000, CuK$_\alpha$ source), scanning electron microscopes (SEM) and transmission electron microscope (TEM, JEM-2100HT and JEM-2100FEF). Approximately 10 mg of BP powder was weighed, and subjected to thermogravimetric analysis (Thermo plus EV02 TG8121; Rigaku Corp) in Ar atmosphere with a heating rate of 10 °C/min. A Raman spectrophotometer (Lab-Ram HR-800; Horiba Jobin Yvon, Inc.) was used to identify the structural units in the BP powders. X-ray photoelectron spectroscopy (XPS) measurement was performed (K-Alpha; Thermo Fisher Scientific Inc.). Monochromatic Al-K$_\alpha$ radiation (1486.6 eV) was used as the X-ray source. The spectra were calibrated by setting the measured binding energy of the C1s peak of adventitious carbon accumulated in the analysis chamber of the spectrometer to 284.7 eV. Samples were transferred to the XPS analysis chamber by using a vessel in which the samples were packed in dry Ar gas to avoid structural changes of the sulfides after exposure to air.

4.2 Results and discussion

Figure 2(a) shows the XRD results of RP in the XRD image. Initially, the starting RP powder has two broad peaks at 2θ=30° and 55°, suggesting the amorphous structure of RP, which is consistent with previously reported amorphous RP [60, 69, 70]. The TEM electron diffraction pattern shows a diffuse ring of RP (Fig. 3(a)), confirming the amorphous nature of RP. The phase transformation from crystalline RP to BP in the ball milling process depends sensitively on the milling time. After ball milling for only 2 h (as shown in Fig. 2(a)), the diffraction peaks of RP disappear and instead distinct diffraction peaks appear at 2θ = 25, 35 and 56°, which is consistent with standard orthorhombic BP (JCPDS No. 76-1957), suggesting a phase transformation of RP to orthorhombic BP. The high-resolution TEM (HRTEM) images (Fig. 3(b)) and combining with the XRD analysis results demonstrate the orthorhombic BP in the black powder.
The XRD patterns and Raman spectra of RP and BP powders at different ball-milling times are also shown in Fig. 2. It can be observed that, as the ball milling time increases, the XRD diffraction peaks become similar. Only the (021) and (111) diffraction peaks widen and shift to a low diffraction angle, indicating that the BP powders become increasingly finer with the increase of the ball-milling time (Fig. 2(a)). Furthermore, the difference in the Raman spectra also demonstrates the transformation of RP into BP powders (Fig. 2(b)). All Raman peaks of BP are much sharper than those of RP, indicating that BP nanoparticles are better crystallized than RP nanoparticles. Furthermore, the broad peak around 348 cm$^{-1}$ in RP disappears after ball milling, and a new peak at 357.6 cm$^{-1}$ appears in BP. Three sharp peaks can be attributed to an out-of-plane phonon mode ($A_g^1$) at 357.6 cm$^{-1}$, and two in-plane modes ($B_2g$ and $A_g^2$) at 431.1 and 459.3 cm$^{-1}$, respectively [28, 30, 39, 41].

A typical SEM image in Fig. 4 shows that the as-prepared BP powder (milling for 30 h, 800 r/min) is composed of micro particles with a size of approximately 500 nm. The energy dispersive X-ray spectroscopy (EDS) analysis indicates that the content of P is 91.17% (wt.%). The TG curve in Fig. 5(a) shows that the sublimation occurred at approximately 440 °C in Ar atmosphere. The resultant BP nanoparticles were further characterized using XPS (Fig. 5(a)). The absence of other elements apart from C, O, and P indicates the high purity of BP nanoparticles. The two distinct peaks at 129.3 eV and 130.2 eV in the P2p spectrum are assigned to the 2p3/2 and 2p1/2 orbitals of zero-valent phosphorous (P$^0$), respectively. The broad peak at 133.9 eV is attributed to oxidized phosphorus (i.e., P$^{5+}$), indicating the partial oxidization of P$^0$ during sample preparation and purification, as the surface of BP nanostructures is sensitive to oxygen and moisture [70].

5 BP as a solid lubricant in the microscale

5.1 Experimental section

5.1.1 Fabrication of BP flakes

In this section, the influence of oxidation effect on the atomic structure and friction properties of BP will be discussed. The bulk BP was fixed on an iron sheet with adhesion tape. Prior to the experiments, the bulk BP was exfoliated using a Scotch tape to expose the surface of pure BP. Bulk BP (99.99% purity) was purchased from XFNANO.
5.1.2 Influence of oxidation on the atomic structure of BP: Experiment

A Cypher S (Asylum Research) AFM system was used, and it operated under contact mode to obtain topography and friction data simultaneously. A TR400BP (Asylum Research, AR) cantilever was used for these experiments. The following parameters were employed: the scan rate was 19.53 Hz, integral gain was 65 and setpoint was 1 V (0.17 nN). The scan sizes of 10 nm × 10 nm and 20 nm × 20 nm were used to reduce the data drift and improve the atomic images.

Some of the prepared BP flakes were selected for Raman experiments. The Raman spectra analyses were conducted by using LabRAM HR800 from Horiba. The samples were excited with a 514 nm green laser, and the lateral resolution was 1 μm.

5.1.3 Influence of oxidation effect on the friction properties of BP: Experiment

A Cypher S (Asylum Research) AFM system was used, and it operated under contact mode to obtain topography and friction data simultaneously. An HQ: XSC11/AIBS cantilever with a spring constant of 0.178 N/m and a torsional constant of 660 nN/V, calibrated using an improved wedge method, was used. The setpoint was changed to vary the applied load, and the tip slid on different flakes back and forth with a sliding distance of 100 nm and a frequency of 4.88 Hz. The friction data was acquired during loading. BP flakes were fabricated on the silicon substrates.

5.2 Influence of oxidation effect on the atomic structure of BP

The atomic image of BP obtained using AFM is shown in Fig. 6. The theoretical atomic structure includes two “layers”: the upper and lower layers. However, owing to the honeycomb structure, only the phosphorus atoms of the upper layer on the sample surface were scanned. Thus, each unit shown in the figure was observed to have four atoms rather than six ones, confirming the special “puckered” and layered structure of BP.

The Raman spectra of fabricated BP flakes at various air exposure times are shown in Fig. 7. Three prominent peaks in the spectra can be observed at 362, 439 and 467 cm⁻¹, corresponding to \( A_{1g} \), \( B_{1g} \) and \( A_{2g} \), respectively [71]. Figure 7 shows that the intensities of the three peaks depend on air exposure time. The newly-exfoliated BP flake presents the strongest Raman intensities, which are much larger than those of the flakes after exposure to air. A decrease in the peak intensity can be observed over time monotonously, which disappears gradually after 80 min of exposure in ambient conditions. Specifically, the change in the Raman intensities can be divided into three steps: from 0 to 50 min, the peak intensities decrease gradually with the reduction in the rate of decrease; from 50 to 80 min, the peak intensities show a rise of 30%; after 80 min, the Raman intensities remain steady.

The oxidation reaction under ambient conditions occurs on the surface of BP owing to the existence of oxygen molecules, water molecules in the air and visible light, modifying the properties of the sample. The visible light generated by excitation stimulates charge carriers between oxygen molecules and the BP surface. During the initial period (0–50 min), the volume of the BP sample increases under ambient conditions over time, owing to the aforementioned oxidation process, reducing the Raman intensities of BP and preventing oxygen molecules from contacting the inner layers of BP, which minimizes the oxidation process. However, long-term exposure to ambient conditions leads to a layer-by-layer etching process; consequently, the swollen sample shrinks and even reduces to the single-layer thickness [72]. Thus, when
the oxidation rate slows down and is less than the etching rate (from 50 to 80 min), the volume of the oxidation layer on the BP sample surface starts to decrease, and subsequently, the Raman intensity increases. The Raman intensity maintains remains steady after 80 min of exposure to air, both by oxidation reaction and the etching process.

5.3 Influence of oxidation effect on the friction properties of BP

To investigate the effect of oxidation on the friction properties of BP, the friction at various air exposure times were studied. As shown in Fig. 8, the friction of BP tends to be larger over the measurement period. Specifically, the increase of the friction is not obvious, which is considered to be stable for the initial 30 min from the newly-fabricated form of BP. However, the friction increases considerably after 30 min and a friction increase of over 200% at 120 min has been observed. Initially, sliding occurs between the contact region of fresh BP surface and the AFM tip. Owing to their two-dimensional structure, newly-exfoliated BP
samples exhibit relatively low friction. As mentioned above, the pure BP layer tends to react with oxygen molecules, water molecules and visible light, forming an oxide layer that consists of $\text{P}_2\text{O}_5$ and $\text{p-P}_2\text{O}_4$ [72, 73]. In this case, sliding occurs between the oxide layer and the AFM tip, resulting in the increase of friction.

6 BP nanosheets as oil-based lubricant additive

6.1 Preparation of ultrathin BP nanosheets in water

In the previous subsection regarding the friction properties of BP in the microscale, the BP nanosheets were prepared via mechanical exfoliation with a Scotch tape. This method has low efficiency and the prepared BP flakes are susceptible to ambient exposure, leading to an irreversible degradation into oxidized phosphorus compounds. In order to harvest the predominant properties of the material, optimized synthesis strategies are required to produce large quantities of exfoliated nanosheets. Liquid exfoliation routes, however, are potentially appropriate for the facile large-scale production of 2D nanomaterials, wherein sonication-induced cavitation and shear are often used together to separate layers from a solvated bulk material.

At present, the most popular and common way of liquid exfoliation to produce mono- or few-layer BP from bulk BP adopts an organic phase as the solvent. These solvents such as N-methyl-2-pyrrolidone, DMF, dimethyl sulfoxide and N-cyclohexyl-2-pyrrolidone have high boiling points, and are difficult to dismantle in post-processing to aqueous solutions. Herein, we established a scalable, high-yield and environmentally benign method for preparing BP nanosheets via ultraso- nication in anhydrous ethyl alcohol.

BP powders were prepared using HEMM technique and stored in an Ar glove box. Anhydrous ethyl alcohol was purged with ultrahigh purity grade Ar for at least 1 h to remove the dissolved oxygen. Ultrapure water (18.25 M$\Omega$·cm, 25°C) was used to prepare all the solutions. The BP nanosheets were prepared using a simple liquid exfoliation technique involving ultrasound probe sonication followed by bath sonication of BP powders. Further, 20 mg of BP powder was added to 40 mL of anhydrous ethyl alcohol in a 50 mL sealed conical tube and sonicated with a sonic tip for 3 h at the power of 1,200 W. The ultrasonic frequency was varied from 19 to 25 kHz and the ultrasound probe worked for 3 s with an interval of 3 s. The dispersion was subsequently sonicated in an ultrasonic bath continuously for another 10 h at the power of 300 W. The temperature of the sample solution was maintained below 277 K using an ice bath. The resulting dispersion was centrifuged for 20 min at 3,000 rpm, and the supernatant containing BP nanosheets was decanted gently. Subsequently, the BP nanosheets solution was centrifuged for 20 min at 7,000 rpm, and the precipitate was repeatedly rinsed with water and re-suspended in the aqueous solution.

Figure 9(a) shows the AFM images of liquid-exfoliated BP samples, and they exhibit a triangular structure. From the SEM of liquid-exfoliated BP samples (Fig. 9(b)), it can be observed that the size of BP nanosheets is approximately 350 ± 13 nm. Such a result is consistent with the size distribution of BP dispersion. The TEM and HRTEM analyses were carried out to investigate the morphology and crystalline quality of the as-synthesized BP nanosheets and nanoparticle samples. Figures 10(a) and 10(b) reveal the morphology of BP nanosheets collected as precipitate
at 3,000 rpm. The TEM images show that the obtained samples consist of nanosheets with lengths in the range of a few micrometers. Figured 10(c) show the HRTEM images of the same BP nanosheets. The selected area electron diffraction (SAED) pattern, which is consistent with the reported data in the literature [74], confirms that the obtained nanosheets are crystalline in nature and are orthorhombic in crystal structure. The lattice fringe is 0.316 nm, which could be ascribed to the (110) plane of the BP crystal.

6.2 Preparation and principle of BP nanosheets as oil-dispersible additives

In this subsection, the synthesis process of the suspensions of BP nanosheets in hexadecane is presented. BP nanosheets were first dispersed in de-ionized water, and this step was completed as described in Section 6.1. The result was a homogeneous dark brown solution of dispersed BP nanosheets. Subsequently, oleylamine was added to the colloidal suspension. Owing to the strong interaction between oleylamine and BP nanosheets, the oleylamine was absorbed on the surface of BP nanosheets, and BP nanosheets became hydrophobic BP nanosheets-oleylamine, floating in water. Subsequently, hexadecane was added to the mixture. The phase transfer process occurred spontaneously, and there was a distinct phase interface between the aqueous phase and hexadecane in 1 day. After removing the aqueous phase using a pipette, the stable suspensions of BP nanosheets in hexadecane were obtained.

6.3 Lubrication characteristics of BP as oil-dispersible additive at lower loads

All the lubrication tests were performed using a commercial tribotester (Optimal SRV4) and the mode of the test is reciprocating ball-on-disk. The disk was fixed and the ball was pressed onto the disk, and reciprocating motion was employed with a stroke of 2 mm and a frequency of 50 Hz. The normal load was increased from 10 N to 70 N. For each test, a new ball and an un-rubbed position of the disk were used. All the tribological tests were performed at least three times to ensure the repeatability of the experimental results. Both the ball and disk were made of bearing steel (AISI 52100); the diameter of the ball was 10 mm and the surface roughness (Ra) was approximately 18.5 nm. The surface of the disk was polished and the roughness (Ra) was approximately 20.4 nm.

Further, molybdenum disulfide and graphene oxide (GO) were dispersed in hexadecane in the same way as the dispersion of BP nanosheets. From Fig. 11, it can be observed that the coefficient of friction (COF) of hexadecane and three dispersions present three features: (1) BP-16C has similar tribological properties as GO-16C and MoS₂-16C; (2) The COFs of the three dispersions decreased as the load was increased from 10 to 70 N; (3) From Fig. 11(d), the COF of hexadecane is apparently different and is higher in the early stages (about 200 s). As the load was increased from 10 to 70 N, the maximum Hertzian type pressure also increased from 978 to 1,870 MPa. This indicates that the maximum Hertzian pressure that hexadecane could sustain was 1,870 MPa. In order to further clarify the difference among hexadecane and the three dispersions, the wear properties are shown in Fig. 12. In general, as the load was increased from 10 to 70 N, the wear scar diameter of the balls increased. The wear scar diameters of the three dispersions were in the range of 175–400 μm, whereas that of hexadecane was in the range of 525–900 μm. The wear scar diameters of hexadecane were three times as large as those of the three dispersions, especially at 70 N. For hexadecane,
The coefficient of friction of three dispersions at lower load (16C is the hexadecane, BP-16C: BP nanosheets were dispersed in hexadecane, GO-16C: graphene oxide were dispersed in hexadecane, MoS2-16C: MoS2 nanosheets were dispersed in hexadecane).

Wear properties of upper samples (ball): (a) wear scar diameter of balls; (b) OM.

Fig. 11 The coefficient of friction of three dispersion at lower load (16C is the hexadecane, BP-16C: BP nanosheets were dispersed in hexadecane, GO-16C: graphene oxide were dispersed in hexadecane, MoS2-16C: MoS2 nanosheets were dispersed in hexadecane).

Fig. 12 Wear properties of upper samples (ball): (a) wear scar diameter of balls; (b) OM.

In this section, the load-climbing test was used where the normal load was first maintained at 50 N for 30 s during the running-in process. Subsequently, the load was increased to 100 N and maintained for 15 min. Subsequently, the load was increased by 100 N every 2 min and the test would stop when the COF increased abruptly over 0.3, indicating that lubrication failure occurred and direct contact (even seizure) between the ball and the disk occurred. The load before lubrication failure was referred to as the highest load without seizure.

In order to investigate the lubrication properties of the four dispersions, the load-climbing test was adopted to fully evaluate the ability of the three kinds of nanosheets to enter the contact area under high loads. MoS2-16C, GO-16C, BP-16C and BP-LP were tested for comparison, and the test temperature was set as 50 °C. BP-LP denotes that BP nanosheets were dispersed in liquid paraffin. As compared with hexadecane, liquid paraffin has higher viscosity. The COF and load over time are displayed in Fig. 14. Notably, the tribological test with BP-LP did not stop until the load reached 2,000 N, which is the highest load that can be provided by the tribotester, which indicates that the highest load with no seizure of the lubricant was no less than 2,000 N. The COF of the lubricant containing BP-16C rose abruptly when the load reached 1,300 N. The highest load of the lubricant containing GO-16C was only 800 N. The performance of the lubricant containing MoS2-16C was the worst and 600 N was its upper limit of the sustainable load. It is evident that BP-16C has the optimal extreme pressure resistance.
Fig. 13  Wear properties of lower samples (disc) at 70 N: (a) hexadecane; (b) BP diapered in hexadecane; (c) GO diapered in hexadecane; (d) MoS$_2$ diapered in hexadecane.

Fig. 14  The results of the load-climbing tribological tests. The tested samples include four lubricants: (a) GO-16C: graphene oxide were dispersed in hexadecane; (b) MoS$_2$-16C: MoS$_2$ nanosheets were dispersed in hexadecane; (c) BP-16C: BP nanosheets were dispersed in hexadecane; (d) BP-LP: BP nanosheets were dispersed in liquid paraffin; (e) Highest load with no seizure of these lubricants; (f) COF of different samples at same time.
6.5 Lubrication characteristics of BP as oil-dispersible additives at high loads

The extreme pressure properties of the three dispersions were evaluated using the load-climbing test, and subsequently, the lubrication properties of the three dispersions were evaluated using SRV4 at different loads for 30 min. Figure 15(a) shows the friction properties of BP-16C at various loads, and it can be observed that, when the load was increased to 400 N (the corresponding maximum Hertzian type pressure was 3343 MPa), the COF increased abruptly for the first 60 s and remained approximately constant thereafter. From Fig. 15(b), it can be observed that the steady COF decreased as the load was increased from 50 to 300 N. The COF increased as the load was increased from 350 to 400 N, and only an incremental rate is less than the decrease rate. From Fig. 15(c), the wear width and wear height also increased as the load was increased, and the wear rate increased abruptly when the load was above 350 N. In order to clarify the reason why the variation trends of friction and wear were different when the load was above 300 N, the surface morphologies and distributions of elements on the worn surfaces are presented in Fig. 16. It can be observed that the surface morphology at 350 N is different from those below 350 N. In this case, the wear track along the sliding direction is examined. In the case of the element distribution, the concentration of Fe and C increased whereas the

![Image of Figure 15](image1)

**Fig. 15** Friction properties of BP-16C at various loads, (a) COF of BP-16C at various loads; (b) the change trend of steady friction coefficient; (c) wear width and wear height of lower samples (disc).

![Image of Figure 16](image2)

**Fig. 16** Surface morphology and distribution of element in worn surface: (a) 50 N; (b) 100 N; (c) 150 N; (d) 200 N; (e) 250 N; (f) 300 N; (g) 350 N; (h) 400 N; (i) distribution of element.
concentration of O decreased as the load was increased above 300 N. This suggests that the iron and carbon oxides may be the reason for the decrease in friction. The mechanism of friction will be further discussed in the following section.

The same test methods were adopted for evaluating the lubricating properties of GO-16C. Figure 17 shows the friction properties of GO-16C at various loads, and the steady COF decreased as the load was increased from 50 to 200 N, whereas the COF increased as the load was increased to 250 N (the corresponding maximum Hertzian type pressure was 2858 MPa). Two reasons could be proposed for such a phenomenon: on the one hand, the COF increased abruptly for the first 80 s and remained approximately constant thereafter; on the other hand, when the load reached 250 N, the contact area increased and the Hertzian contact pressure decreased, and subsequently, the steady COF at 250 N decreased. Figure 17(b) summarizes the wear widths and wear heights of the lower samples (disk); the wear width and wear height also increased as the load was increased, and the wear rate increased abruptly when the load was achieved to 250 N. According to the definition stated previously, the highest sustainable load was 250 N. Figure 18 shows the friction properties of MoS2-16C at various loads. The COF for load above 150 N increased abruptly for the first 80 s and remained approximately constant thereafter. The highest sustainable load was 150 N (the corresponding maximum Hertzian type pressure was 2,411 MPa).

6.6 Comparison of three oil-based lubrication additives at ultra-low concentration

It is known that a lubrication additive can enhance the properties of base oil in different aspects. Usually, lubrication additives contain sulfur, phosphorous, etc. However, sulfur compounds in engine exhaust gases are known to poison the catalysts used in exhaust systems to reduce NOx emissions, and phosphorous is known to poison the hydrocarbon conversion catalysts used in those systems. Simply lowering the
amount is not a practical solution to the problem owing to the concomitant reductions of the antiwear properties and oxidation inhibition properties. Thus, the objective is to provide an engine oil for use in conjunction with low sulfur and phosphorous fuels (< 350 ppm) for enhanced emission control. Accordingly, the lubrication properties of BP, GO, and MoS₂ as additives of ultra-low concentrations will be investigated.

The lubrication properties of the oil dispersions with BP, GO and MoS₂ as additives were evaluated using SRV4 at 100 N for 30 min. The other experimental conditions were the same as the previous part, except that the concentration was varied. Figure 19 is the COFs of three dispersions as additives of various concentrations. The steady COF of BP-16C in Fig. 19(a) decreases as the concentration of the additive concentration was increased from 1 to 10 mg/L. However, when the concentration of BP-16C was decreased to 0.5 mg/L, the COF is unstable and fluctuates between 0.12 and 0.15. The COF curve of 0.1 mg/L BP-16C was different from those with higher concentrations, and the initial COF increased to 0.20 at the first 20 s and then fluctuated between 0.13 and 0.15. As the concentration continued to decrease, the COF increased abruptly during the initial period. Thus, on the basis of the definition of the lubrication failure in this work, the lowest sustainable concentration was 0.1 mg/L (0.1 ppm). Few layered graphene and MoS₂ are also the layered materials which stacked along the c-axis by van der waals forces. In recent years, extensive research about the graphene oxide and MoS₂ as lubricant additive has been conducted. To further evaluate the comprehensive performances of BP as oil based lubricant additive, the compare study of graphene oxide and MoS₂ at ultra-low concentration as oil based lubricant additive are introduced. The COFs of GO-16C at various concentrations were shown in Fig. 19(b). When the concentration decreased to 5 mg/L, the COF is instability and the initial COF increased to 0.6 at the first twenty seconds and then tended to be stable. The COF at 1 mg/L is entirely different from that of the higher concentration. The initial COF increased to 0.70 at the first twenty seconds and then fluctuated between 0.70 and 0.18 until 280 s to the steady state. In general, in the low concentration (< 25 mg/L) as the concentration increased the steady state COF decreased. Thus, the ultra-low concentration of GO lubricant additives is different from that of BP. In same situation, the lowest bearing concentration of GO lubricant additives is 5 mg/L (5 ppm). The COFs of MoS₂-16C at various

![Fig. 19](https://mc03.manuscriptcentral.com/friction)  
**Fig. 19** Coefficients of friction (COF) of BP-16C at various concentrations: (a) COF of BP-16C below 10 mg/L; (b) COF of GO-16C below 10 mg/L; (c) COF of MoS₂-16C below 10 mg/L; (d) Comparison of three oil-based lubrication additives at ultra-low concentration.
concentrations were shown in Fig. 19(c). The similar method was used for evaluating the ultra-low concentration of MoS$_2$-16C. In the concentration from 1 to 25 mg/L, the variety regulation of COF was not obvious. However, when the concentration decreased to 0.5 mg/L, the COF was displayed strikingly different behavior. At this concentration, the COF was instability and the initial COF was increased to 0.6 at the first twenty seconds and then tended to be stability. Thus, the lowest bearing concentration of MoS$_2$ lubricant additives is 0.5 mg/L (0.5 ppm). Comparison with three oil-based lubrication additives at ultra-low concentration (Fig. 19(d)), BP-16C has best lubrication properties at lowest concentration, 50 times more lubrication properties than that of GO-16C, and 5 times more lubrication properties than that of MoS$_2$-16C.

6.7 Comparison analysis of three 2D material lubrication additives at 1 mg/L

The tests in this section were also performed using a commercial tribotester (Optimal SRV4), and the mode of the test was reciprocating ball-on-disk. The disk was fixed and the ball was pressed onto the disk, and the reciprocating motion was employed with a stroke of 2 mm and a frequency of 50 Hz at a temperature of 50 °C. The normal load was 100 N (the corresponding maximum Hertzian type pressure was 2,106 MPa). For each test, a new ball and an un-rubbed position of the disk were used. All the tribological tests were performed at least three times to ensure the repeatability of the experimental results. Both the ball and disk were made of bearing steel (AISI 52100); the diameter of the ball was 10 mm and the surface roughness (Ra) was approximately 18.5 nm. The surface of the disk was polished and the roughness (Ra) was approximately 20.4 nm.

In the previous section, the lubricating properties of three 2D materials—BP, MoS$_2$ and GOs oil-based lubrication additives at ultralow concentration were investigated. The results showed that, under these conditions (100 N, 50 Hz, 50 °C), the lowest bearing concentrations of BP-16C, GO-16C and MoS$_2$-16C were 0.1 mg/L (0.1 ppm), 5 mg/L (5 ppm) and 0.5 mg/L (0.5 ppm), respectively. However, there is no direct comparative study among BP, GO and MoS$_2$ as oil-based lubricating additives at the same concentration.

In order to further explain the difference among the three 2D materials as lubricant additives at the same concentration, the friction and wear tests were conducted at 1 mg/L. Figure 20 shows the COF of the three 2D materials as lubrication additives. It can be observed that the COF of hexadecane increased to 0.8 in the first 160 s, and thereafter entered the steady COF (0.18), whereas the COF of the GO at the initial phase was 0.14 and lasted for 55 s, suddenly increased to 0.7 and lasted for 145 s, and finally entered the steady COF (0.167). However, the COFs of BP-16C and MoS$_2$-16C were basically stable and remained at 0.140. This result indicates that BP-16C and MoS$_2$-16C have similar properties of wear resistance.

6.8 Lubrication mechanism of BP-16C

In order to clarify the reason why the oil dispersion with BP-16C as the additive of ultra-low concentrations exhibits good lubrication abilities, the worn surfaces of the upper samples (balls) were analyzed by SEM and EDS (Fig. 21(a)). The chemical compositions of the friction surface mainly contained C, O, P, Cr and Fe. C originated mainly from the base oil (hexadecane), P originated from the BP nanosheet, and Cr and Fe mainly originated from the substrate. The content of BP in the black wear debris after frictional reaction was 1.51 wt.%, whereas less BP was observed in other regions of shallow color. Thus, BP was rich in wear debris on the surface of frictional balls, and the reaction film on the frictional surface mainly contained C, Fe and ferrous oxide. Figure 21(b) shows the sampling place of the worn surface using focused ion beam (FIB), and a Cr film and Pt layer were deposited on the sample surface before FIB cutting to protect...
the tribofilm. Figure 21(c) shows the TEM result of the specimen after fabricating using FIB. It was composed of a protective coating, frictional layer and substrate. The frictional layer had a polycrystalline structure (Figs. 21(d) and 21(e)), and it was mainly composed of black nanosheet, tiny wear debris and white substrate. In order to further explain the chemical compositions of the frictional layer, the distribution of the chemical elements in the frictional layer on the worn surfaces of balls is presented in Fig. 22. The white part in the frictional layer consists largely of C, and the black nanosheets are composed of iron oxide. The content of phosphorus is very difficult to detect in this situation. Thus, the frictional layer on the upper samples (balls) mainly consists of C, Fe and ferrous oxide.

The SEM and TEM images of the worn surfaces on the disks lubricated with 75 mg/L BP-16C are shown in Fig. 23. Local enrichment of debris in the wear track can be observed, and the thin nanosheets containing 3.09 wt.% phosphorus have higher content of phosphorus. The elements Si, Mn, and Cr originated from the substrate. Figure 23(b) shows the sampling place of the wear surface using FIB. A Cr layer was first deposited on the sample surface, and subsequently, a SiO$_2$ layer was deposited on the sample surface before the FIB fabrication to protect the tribofilm. Thus, the effect of impurity element Pt on P can be significantly alleviated because their characteristic peaks in EDS are close. Figure 23(c) shows the TEM result of the specimen prepared using FIB. It can be observed from Fig. 23(d) that the sample is composed of the SiO$_2$ layer, the friction layer and the substrate. The frictional layer is in homogeneously distributed, and the layer thickness was 20–200 nm. Dark and white areas were present in the frictional layer, whereas the chemical composition will be discussed in the following section. After the frictional reactions, the substrate exhibited various characteristics. Large plastic deformation was observed in the zone near the frictional layer, and the nanocrystal line was present in the red circle 4 of Fig. 23(d). From the surface of the frictional film to its interior, the microstructures transformed from a nanocrystal line to single crystals. During the friction process, there was plastic displacement of the metal on the substrate surface and the ball surface. Severe plastic deformation on the surface can improve the surface hardness, and thus, it is helpful to improve the wear resistance (Fig. 23(f)). Figure 24 shows the
distribution of the chemical elements in the frictional layer of the substrate (disks). It can be observed from the red box that the rim of the friction layer is composed of Cr, Si and O. The protective coating protects the friction layer from damage during the FIB processing. The friction layer consisted of Fe, O, C and P. Phosphorus was evenly distributed in the friction layer, and this result is different from the previous friction layer on the ball surface. The black nanosheets in the friction layer were composed of high-concentration Fe and O, whereas the white area was composed of high-concentration O and C. These results indicate that the friction layer was mainly composed of iron oxide, phosphide, and oxides of carbon.

In order to understand the lubrication mechanism of BP-16C, wear particles were collected on the wear scar of the steel ball after the friction test and they were investigated using TEM and HRTEM, as shown in Fig. 25. A variety of debris with different shapes and sizes can be observed and they are flat and elongated particles with a typical length of 20 nm. The Debye-Scherrer electron diffraction pattern obtained for this particle presents several continuous rings centered at 2.127, 3.58, 3.892, 4.073, and 4.558 nm$^{-1}$ (Fig. 25(b)). Inter-reticular distances calculated for these new rings (0.4701, 0.2793, 0.2569, 0.2455 and 0.2194 nm, respectively) fit well with the distances in the Fe$_3$O$_4$ structure (orthorhombic iron oxide γ-Fe$_3$O$_4$, JCPDS data).
52-1499 with cell = 0.5095 nm × 0.8789 nm × 0.9437 nm): 0.4718 nm for \{002\}, 0.2798 nm for \{311\}, 0.2560 nm for \{113\}, 0.2452 nm for \{131\} and 0.2197 nm for \{040\}. The BP nanosheets and amorphous carbon were also discovered in this microstructure using HRTEM and FFT pattern. The chemical composition of the debris is shown in Fig. 26. The chemical analysis shows that the debris mainly contained four elements: Fe, O, P and C. The size of the BP nanosheets was varied. If the larger BP nanosheets could not enter the contact area, the distribution of the chemical elements in Fig. 27 explains the lubrication mechanism of BP-16C. Very small BP nanosheets (approximately 20 nm) are shown in this figure and evenly distributed in the wear debris. This result indicates that the BP nanosheets entered into the contacted area, and reacted with the hexadecane and steel substrate.

In order to further explain the lubrication mechanism, Raman and XPS analysis were performed, as shown in Fig. 28. In the Raman spectrums (Fig. 28(a)), it can be seen that as the concentration of BP in hexadecane decreased, the strength of the characteristic peak also decreased. For the wear scar of the disk lubricated with the base oil, the typical peaks were detected at 656, 497, 403, 218 and 288 cm\(^{-1}\). Previous studies \[75, 76\] used Raman spectra for corrosion investigation of stainless steel and showed the characteristic peaks of iron oxides at ~700 cm\(^{-1}\) and below. Maslar et al. \[77\] investigated the \textit{in-situ} Raman spectroscopic of aqueous iron corrosion at elevated temperatures and pressures, and found that the characteristic peak at 656 cm\(^{-1}\) was detected, 497 cm\(^{-1}\) was the mixture of \(\alpha\)- and \(\gamma\)-Fe\(_2\)O\(_3\), and the characteristic peaks among 403, 288 and 220 are all the \(\alpha\)-Fe\(_2\)O\(_3\). Thus, the Raman signatures in our research of the wear tracks with BP-16C show the signal for iron oxides (below 700 cm\(^{-1}\)). The peak of 1,316 cm\(^{-1}\) presented C-C stretch skeletal vibrations \[78\], that’s means a protective film may have been formed by hexadecane in friction process. This lubricating film also reduced the friction and wear. These findings indicate that the deposition, self-repair, and formation of the chemical reaction film may be the main causes of the friction-reduction and anti-wear properties of BP-16C.

XPS analysis was performed to explore the lubricating mechanism of this lubrication additive. Figs. 28(b)–28(f) illustrate the full spectrum and curve-fitted XPS spectra of C1s, Fe2p, O1s, and P2p on the wear scar of the disk lubricated with 10 mg/L
BP-dispersed oil. The Fe2p peak around 710.8 eV is attributed to iron oxide and iron phosphate [79]. The Fe2p peaks around 706.7 eV is attributed to the iron substrate. P2p peaks at 132.4 eV is identified as PO\(_4^{3-}\). The N1s peak around 398.1 eV is attributed to organic amine that comes from oleylamine using as the modifiers of dispersion. The Fe2p3/2 at 710.8 eV and the O1s peak at 529.1 eV both fall in the binding energy ranges covered by Fe\(_2\)O\(_3\) [80]. Iron carbonate may still be present, but the Fe2p peaks associated with Fe\(_2\)O\(_3\) completely cover the traces. The indicator that FeCO\(_3\) was initially present is the small O1s peak at 530.9 eV and the C1s peak at 288.1 eV. In the carbon region the peak at 285.8 eV attributed to the C-O bond and the peak at 284.4 eV attributed to the C-C/C=C bond were identified. This was reported to be from the tribo-chemical reaction of hexadecane [81, 82]. Therefore, it can be concluded that a tribofilm made of BP and iron oxides was formed on the wear scar. These oxides play an important role in the friction-reducing ability of BP-16C.

The novel kind of oil-soluble BP nanosheets were prepared by phase transfer method. The tribology properties were evaluated by SRV-4 with ball-on-disc model. Schematic diagram of lubrication mechanism of BP nanosheets as lubrication additive was presented in Fig. 29. The aforementioned experimental results revealed that BP-16C has excellent lubrication properties under boundary lubrication. The lubrication mechanism with BP nanosheets as lubrication additives was concluded as follows. Firstly, when the micro peaks are moving each other, the ultrathin BP nanosheets will enter the contact area instead of being pushed away. When the contact pressure was applied to the touched micro peaks, because of the interlaminar shear between the few-layer the ultrathin BP nanosheets, they will not contact with each other directly. Secondly, shearing force in the hexadecane lubricant was another important impact factors. Finally, the plastic deformation of substrate was benefited for wear resistance. This study demonstrates that very low concentration of nicely dispersed BP nanosheets can generate the BP tribological thin film, which could...
efficiently reduce the friction and protect the sliding interfaces. This promises the potential of BP-based lubricants for conserving energy by lowering the friction and reduction of material losses (wear-preventive property) for tribological applications.

7 Friction properties of BP/PVDF composites materials

7.1 Preparation of BP/PVDF composites

PVDF powders were purchased from J&K Scientific Ltd., and BP powders were prepared using the HEMM method for 48 h. The weight fractions of BP in the composites were 0, 1%, 5%, 10%, 15% and 20%. The BP/PVDF composites were prepared according to the following steps (Fig. 30). First, a certain weight of BP and PVDF powders were weighed according to the target compositions, and subsequently wet ball-milled in anhydrous alcohol under argon atmosphere for 2 h. The milling time and speed were 2 h and 300 rpm, respectively. After being mixed and dried, the mixtures were sintered via spark plasma sintering (SPS) at 250 °C under a pressure of 10 MPa for 10 min in pure Ar atmosphere protection. A cylindrical graphite mold with an inner diameter of 20 mm was used.

7.2 Characterization and friction properties of BP/PVDF composites

The phase analysis of BP/PVDF composites was characterized using XRD with Cu-Kα radiation within the 2θ range of 10–90°, and the molecular structure of the composites was analyzed using Raman spectra. The wavelength of laser light was 532 nm, and the range of measured Raman shift was 100–1,100 cm⁻¹.

Fig. 30 Schematic diagram of preparation process of BP/PVDF composites.

The morphologies of BP/PVDF composites were characterized using SEM. The friction test was conducted on a universal micro-tribotester (UMT-5), and the reciprocating speed was 0.03 m/s. Stainless steel balls of diameter 4.4 mm were used. The sliding experiments were performed over a period of 30 min and at the applied loads of 1, 3 and 5 N.

7.3 Results and discussion

The XRD patterns of BP, PVDF and BP/PVDF composites with different BP contents are shown in Fig. 31(a). The XRD peaks at 2θ of 16.8°, 26.6°, 34.9° and 55.5° correspond to the (020), (021), (111) and (132) crystal planes of BP, respectively. The XRD peaks at 2θ of 18.4°, 19.9° and 26.4° correspond to the (020), (110) and (021) crystal planes of PVDF, respectively. The XRD patterns of BP/PVDF composites prepared by SPS are also presented in Fig. 31(a). The three strongest peaks of BP are marked in the graph, corresponding to the (111), (112), and (132) crystal planes. All the patterns of BP/PVDF composites show (111) peaks similar to that of pure BP, and the peak intensity increases with the increase in BP content. However, the (112) and (132) peaks cannot be clearly observed in the composites. The Raman spectroscopies of BP, PVDF, and BP/PVDF composites with different BP contents are presented in Fig. 31(b). In BP, the peaks of 359 cm⁻¹, 436 cm⁻¹ and 462 cm⁻¹, corresponding to A1g, B2g and A2g phonon modes, respectively, can be observed. The strongest peak of PVDF was observed at 796 cm⁻¹. The Raman spectra of BP/PVDF composites with different BP contents are also shown in Fig. 31(b). In all the composites, the characteristic peaks of BP were also presented, and moreover, as the content of BP in the composites increased,

Fig. 31 (a) XRD patterns of BP, PVDF and BP/PVDF composites with different BP contents; (b) Raman spectra of BP, PVDF, and BP/PVDF composites with different BP contents.
the intensities of peaks increased correspondingly. Figure 32 shows the SEM images of the composites with BP contents of 15 wt.% and 20 wt.%. The BP particles were properly dispersed in the PVDF matrix, and the size of the BP particles increased with the increase in the content of BP. However, the BP particles in the composites tended to aggregate with the increase in content of BP.

The mean COF value of BP/PVDF composites as a function of BP content and applied load are shown in Fig. 33. Under a constant applied load, the COF increased with the increase in the BP content, and reached the highest value when the BP content was approximately 10 wt.%. However, the COF value decreased rapidly as the content of BP was increased further. It is shown in Fig. 34 that, when the BP content in the BP/PVDF composite is lower than 10 wt.%, the addition of BP filler destroys the semi-crystalline structure of PVDF, resulting in the decrease of the yield strength, and correspondingly the COF of the composite decreases. In the case of the composite with BP content larger than 10 wt.%, the BP filler plays a more important role in the lubricating process. In order to demonstrate the results of this analysis, the wear scar region was analyzed using the Raman spectrometer. As shown in Fig. 35, the Raman spectra in the wear scar and on the surface of the composite show similar peaks at 362 cm\(^{-1}\), 432 cm\(^{-1}\) and 461 cm\(^{-1}\) owing to the BP filler, and at 794 cm\(^{-1}\) owing to the PVDF. This indicates that chemical reactions did not occur during the sliding process. Therefore, it can be inferred that the improvement of friction properties is ascribed to the increasing amount of BP filler and its lubricating properties.

However, the COFs of BP/PVDF composites with certain BP contents decreased with the increase in applied load. Considering the composite with 10 wt.% BP content as an example, the COF value decreased from 0.32 to 0.24 when the load increased from 1 to 5 N. A possible explanation is that the friction surface experienced not only plastic deformation but also elastic deformation [5]. In this case, the COF is inversely proportional to applied load, and shows a decreasing trend with the increase in load. Figure 36 shows the wear rate of BP/PVDF composites at different applied loads as a function of BP content. It can be observed that, under a certain applied load, the wear rate decreases rapidly with the increase in the content of BP filler. For instance, when the load was 5 N,
the wear rate reduced from $8.6 \times 10^{-4}$ mm$^3$//(N·m) to $2.5 \times 10^{-5}$ mm$^3$//(N·m) with the BP filler increasing from 0 to 20 wt.%. This indicates that the addition of BP filler is beneficial to the reduction in the wear loss of the composites.

8 Conclusion and outlook

In this work, the synthesis of BP nano-powders using the HEEM technique, modification of BP nanosheets as water-based and oil-based lubricant additives and the tribological characteristics of few-layer BP nanosheets have been systematically studied. Moreover, the tribological characteristics of MoS$_2$ and GO nanosheets as oil-based lubricant additives were also investigated for comparison. Furthermore, the tribological properties of polymeric composites with BP as the filler were also investigated. It was observed that, owing to the oxidation process in ambient environment, the Raman intensities at different air exposure times decreased initially and subsequently remained steady owing to the formation of oxidation layers, resulting in the increase in friction in the microscale. For the use of BP nanosheets as oil-based lubricant additives, the BP dispersions showed similar lubrication properties at low loads, whereas the BP dispersions exhibited outstanding resistance to extreme pressure and bearing capacity at high loads. For BP/PVDF composites, when the BP content was larger than 10%, the improvement of friction properties was more apparent, and the addition of BP could enhance the wear resistance of BP/PVDF composites.

In spite of the current progresses of the tribological properties of BP, several challenges still remain. First, the methods for cost-effective mass production of BP nanomaterials are still required. Second, while the small lateral size ($<1 \mu$m$^2$) of most nanosheet dispersions is sufficient for many technologies in tribological applications, larger flake size is more desirable for lubrication additives applications. Third, further work on the ultra-low friction properties of BP-based lubricant or composite is necessary.

Acknowledgements

The authors would like to acknowledge Xudong Chen, Ziyi Cui, Jinjin Li and Dan Guo for some of experiments present in the manuscript and their helpful discussions. The authors also would like to acknowledge the support of the National Natural Science Foundation of China (Grant Nos. 51527901, 51335005, 51475256, and 51605249).

Open Access: The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
References


Wei WANG. He received the Ph.D. degree in Materials Processing Engineering from Northwestern Polytechnical University, Xi’an, China, in 2015. After a two-year post-doctoral work at State Key Laboratory of Tribology of Tsinghua University, he is currently an associate professor from Xi’an University of Architecture and Technology. Dr. Wang has published more than 10 papers indexed by SCI as the first author. His major research areas include lubrication in material processing and nano-lubrication.
Guoxin XIE. He received his Ph.D. degree at Tsinghua University, China, in 2010, majoring in Mechanical Engineering. After that, he spent two years at State Key Laboratory of Tribology, Tsinghua University, China for postdoctoral research. From 2012 to 2014, he worked at Royal Institute of Technology, Sweden, for another two years’ post-doctoral research. Since 2014, he has worked at Tsinghua University as an associate Professor. His research interests include intelligent self-lubrication, electric contact lubrication, etc. He has published more than 50 referred papers in international journals. He won several important academic awards, such as Chinese Thousands of Young Talents, the Excellent Doctoral Dissertation Award of China, and Ragnar Holm Plaque from KTH, Sweden.

Jianbin LUO. He received his BEng degree from Northeastern University in 1982, and got his MEng degree from Xi’an University of Architecture and Technology in 1988. In 1994, he received his PhD degree from Tsinghua University and then joined the faculty of Tsinghua University. Prof. Jianbin Luo is an academician of the Chinese Academy of Sciences and a Yangtze River Scholar Distinguished Professor of Tsinghua University, Beijing, China. He was awarded the STLE International Award (2013), the Chinese National Technology Progress Prize (2008), the Chinese National Natural Science Prize (2001), and the Chinese National Invention Prize (1996). Prof. Luo has been engaged in the research of thin film lubrication and tribology in nanomanufacturing. He was invited as a keynote or plenary speaker for 20 times on the international conferences.