Thermodynamics of surface degradation, self-organization and self-healing for biomimetic surfaces

BY MICHAEL NOSONOVSKY 1 AND BHARAT BHUSHAN 2,*

1 Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030, USA
2 Nanoprobe Laboratory for Bio- & Nanotechnology and Biomimetics, Ohio State University, 201 West 19th Avenue, Columbus, OH 43210-1142, USA

Friction is a dissipative irreversible process; therefore, entropy is produced during frictional contact. The rate of entropy production can serve as a measure of degradation (e.g. wear). However, in many cases friction leads to self-organization at the surface. This is because the excess entropy is either driven away from the surface, or it is released at the nanoscale, while the mesoscale entropy decreases. As a result, the orderliness at the surface grows. Self-organization leads to surface secondary structures either due to the mutual adjustment of the contacting surfaces (e.g. by wear) or due to the formation of regular deformation patterns, such as friction-induced slip waves caused by dynamic instabilities. The effect has practical applications, since self-organization is usually beneficial because it leads to friction and wear reduction (minimum entropy production rate at the self-organized state). Self-organization is common in biological systems, including self-healing and self-cleaning surfaces. Therefore, designing a successful biomimetic surface requires an understanding of the thermodynamics of frictional self-organization. We suggest a multiscale decomposition of entropy and formulate a thermodynamic framework for irreversible degradation and for self-organization during friction. The criteria for self-organization due to dynamic instabilities are discussed, as well as the principles of biomimetic self-cleaning, self-lubricating and self-repairing surfaces by encapsulation and micro/nanopatterning.

Keywords: friction; tribology; self-organization; self-healing; biomimetic surfaces

1. Introduction

One of the most remarkable properties of biological surfaces is their ability for self-healing and repairing the damage caused by friction and wear. One purpose of biomimetic surface design is to mimic this ability. Self-healing is in apparent contradiction with the second law of thermodynamics, which states that dissipative processes, such as friction, are irreversible. However, in many nonlinear thermodynamically open systems that operate far from equilibrium,
including frictional systems, self-organization can occur (Fox-Rabinovich & Totten 2006). It is therefore important to investigate the thermodynamics of friction and self-organization during friction in order to be able to design biomimetic self-healing surfaces.

Friction and wear are complex phenomena that involve many physical and chemical processes, such as deformation, adhesion, abrasion, fracture, ploughing, chemical reactions, capillary condensation and many others. At the same time, friction and wear occur almost universally during the contact of solid surfaces. Furthermore, simple relationships, such as Amontons’ empirical law stating that the friction force $F$ is linearly proportional to the normal load force $W$,

$$F = \mu W,$$  \hspace{1cm} \text{(1.1)}

where $\mu$ is a coefficient of friction (constant for a given materials combination), are valid for a large range of material combinations, operation conditions and for forces ranging from nanonewtons to thousands of tons (Bhushan 2002).

Although empirical models of friction and wear currently prevail, a number of attempts have been made to explain friction using first-principles approaches. Historically, the first naive attempts to formulate and explain the linearity of Amontons’ friction law 1 were made by Leonardo da Vinci (1452–1519), Amontons (1663–1706) and Coulomb (1736–1806). The typical explanation was that asperities at a surface constitute the reason for friction and that the coefficient of friction is equal to the average absolute value of the asperity slope. Interestingly, similar mechanisms, known as the ratchet mechanism (at the macro and microscale), the cobblestone (Gao et al. 2004) and hard core repulsion (Sokoloff 2006) mechanisms (at the atomic scale), are still considered today when physical explanations of equation (1.1) are discussed. These are the slope-controlled theories of friction.

A different type of theory relates friction to the surface roughness or heterogeneity. According to the Greenwood & Williamson (1966) model that considers the statistics of rough surface contact and its numerous modifications (e.g. with fractal surfaces), the real area of contact is almost linearly proportional to the normal load, since the contact takes place only at the top of the asperities. Bowden & Tabor (1939) suggested that the friction force is proportional to the linear area of contact. Together, these two models explain the linearity of Amontons’ law. Surface roughness and heterogeneity often have a similar effect. He et al. (1999) showed that in a somewhat similar manner, chemical heterogeneity (adsorbed molecules) could play a role in friction. Other mechanisms that lead to the linearity of equation (1.1) were also suggested, including hierarchical asperities (Archard 1957), ‘elastic ploughing’ (Ying & Hsu 2005), etc. Various theories of friction explaining the linearity of the Amontons law are summarized in figure 1.

We have proposed a different explanation as to why the linear Amontons law is so remarkably universal (Nosonovsky & Bhushan 2007, 2008a). In many cases, a linear functional dependence, such as $F = \mu W$, is a limiting case of a nonlinear dependence $F = f(W)$ when the latter is studied in a small vicinity of a constant value. This suggests that frictional interactions are an asymptotic limit of a more general case of interactions, and that a non-dimensional ‘small parameter’, $\xi \ll 1$, characterizing the ratio of surface-to-volume forces can be identified in diverse friction mechanisms. This makes linearization inherent in friction. When $\xi \to 0$, we deal with friction and a linear law applies, whereas when
\( \xi \) is not small, we deal with more general interactions. Indeed, friction involves interactions along the interface, which are weak compared with the bulk forces, so that the ratio of the bulk-to-interface force is small. Friction is therefore a nonlinear process, and self-organization can occur during friction.

The attention of scientists was driven in recent years to friction and adhesion at the nanoscale and in biological systems. These studies led to several conclusions. First, it is understood now that friction mechanisms at the atomic scale are quite different from macroscale mechanisms. Furthermore, macroscale frictional properties can hardly be deduced from the atomic-scale interactions, since there is a lot going on in between, at the mesoscale. Thus, friction is essentially a multiscale phenomenon (Nosonovsky & Bhushan 2007a). Second, biological surfaces that adapted to unusual frictional properties (such as very high or low friction or adhesion, or so-called ‘smart’ or controlled adhesion) have hierarchical organization (Nosonovsky & Bhushan 2008a). In addition, unlike most engineered surfaces that are subject to wear and degradation, many biological surfaces have the ability to self-repair/self-heal, and this ability is associated with their hierarchical structure (Fratzl 2007). Mimicking nature, or biomimetics, was suggested as an approach to design engineered surfaces with functional properties that cannot be achieved with the traditional engineering paradigm.

While a comprehensive first-principles theory of friction and wear involving the multiscale character of frictional mechanisms remains to be developed, it is clear that one of the main properties of frictional interactions is energy dissipation. Thermodynamic entropy is the measure of dissipation. In this paper, we present a first-principles thermodynamic model for friction, degradation and wear, surface self-organization and self-healing. The approach is applied to biomimetic self-healing, self-lubricating and self-cleaning surfaces with hierarchical structure or mesoscale patterning.

2. Entropy, degradation and self-organization

Friction and wear are dissipative irreversible processes. Entropy is the measure of irreversibility and dissipation. Therefore, entropy can be used to characterize wear and related degradation processes. On the other hand, in systems that
operate far from the thermodynamic equilibrium, entropy is driven away from the interface faster than excess entropy is produced at the interface, and thus the orderliness at the interface can actually increase, leading to self-organization. These phenomena are investigated theoretically in the present section.

(a) Thermodynamics of degradation during frictional contact

Entropy, \( S \), was introduced in the 1850s by R. Clausius as a measure of irreversibility. It is defined in thermodynamics as

\[
\frac{dS}{dT} = \frac{dQ}{T}, 
\]

where \( T \) is the temperature and \( Q \) is the heat. Entropy is an additive function (the net entropy of a system is equal to the sum of entropies of its parts). When heat \( dQ \) is transferred from a body with temperature \( T_1 \) to a body with temperature \( T_2 \), the net entropy changes by the amount \( dS = -dQ/T_1 + dQ/T_2 \). Thus, if heat is transferred from a hotter body to a colder one \((T_1 > T_2)\), the net entropy grows \((dS > 0)\). This provides a convenient formal basis for the second law of thermodynamics, stating that the net entropy of a closed system either remains constant (for a reversible process) or grows (for an irreversible process).

In 1877, L. Boltzmann suggested a definition of entropy using the statistical thermodynamics approach and the concept of microstates

\[
S = k \ln \Omega, 
\]

where \( k \) is Boltzmann’s constant and \( \Omega \) is the number of microstates corresponding to a given macrostate. Microstates are arrangements of energy and matter in the system, which are distinguishable at the atomic or molecular level, but are indistinguishable at the macroscopic level (Craig 1992). A system tends to evolve into a less-ordered (more random) macrostate that has a greater number of corresponding microstates, and thus the ‘configurational’ entropy given by equation (2.2) grows.

The concept of entropy can be applied to friction, surface degradation and self-organization. Manufacturing transforms raw materials into highly organized components, while ageing and degradation tend to return them into their natural disordered state (Bryant et al. 2008). The entropy production rate is given by

\[
\frac{dS}{dt} = XJ, 
\]

where \( J \) is the thermodynamic flow rate and \( X \) is a generalized thermodynamic force.

Consider now frictional sliding with the velocity \( V = dx/dt \), applied normal load \( W \) and friction force \( F = \mu W \). The work of the friction force is equal to the dissipated energy, and we will assume now that all dissipation energy is converted into heat

\[
dQ = \mu W \, dx. 
\]

Substituting equation (2.4) into equation (2.1) yields the rate of entropy generation during friction

\[
\frac{dS_t}{dt} = \frac{\mu WV}{T}, 
\]

with the flow rate \( J = V \) and the thermodynamic force \( X = \mu W/T \).
Bryant et al. (2008) suggested a degradation measure $w$ that is a parameter associated with a particular degradation mechanism, so that the rate of degradation is given by

$$\frac{dw}{dt} = B \frac{dS_f}{dt} = YJ,$$

(2.6)

where $Y = BX$ is the generalized degradation force and $B$ is a constant degradation coefficient (a material property).

For the wear process, it is natural to take wear volume as the degradation measure $w$. Combining $\dot{w} = B \dot{S}_f$ (equation (2.6)) with equation (2.5) yields the wear rate

$$\frac{dw}{dt} = B \frac{\mu WV}{T}.$$

(2.7)

For plastic contact, equation (2.7) can be rewritten by setting the wear coefficient $k = \mu HB/T$ as

$$\frac{dw}{dt} = k \frac{WV}{H},$$

(2.8)

where $H$ is the hardness of a softer material in contact. For elastic contact, $k = \mu E^*(\sigma/\beta^*)B/T$ yields

$$\frac{dw}{dt} = k \frac{WV}{E^*(\sigma/\beta^*)},$$

(2.9)

where $E^*$ is the effective elastic modulus; $\sigma$ is the standard deviation of the rough profile height; and $\beta^*$ is the correlation length of the profile roughness (in a sense, $\sigma$ is the height and $\beta^*$ is the length of a typical asperity). Equation (2.8) is widely used as the Archard equation for adhesive wear, whereas equation (2.9) is known as the Bhushan (2002) equation.

The non-trivial part of this derivation is that the wear volume per unit time is linearly proportional to the friction force. This implies that (i) a constant value of energy dissipated by friction is consumed for wear debris generation and (ii) a constant amount of energy is consumed, in average per wear particle volume. Wear and friction are essentially the same dissipative process, as the energy consumed for creating wear particles is related to the work of friction.

(b) Thermodynamics of self-organization during frictional contact

In §2a, we investigated irreversible degradation at the surface (wear) during friction. However, the opposite process of self-organization can also take place.

(i) Self-organization

A different approach to use the concept of entropy was developed in the 1980s by a group of Russian tribologists, including N. Bushe, L. Bershadsky, I. Gershman and others, who applied non-equilibrium thermodynamics to the frictional contact (see the review in Fox-Rabinovich & Totten 2006; Fox-Rabinovich et al. 2007). It is well known that when a frictional process
starts, after a certain transition period (run-in), it usually enters a steady-state regime. The contacting bodies tend to adjust to each other. Various manifestations of this can be viewed as evidence of self-organization.

While the net entropy grows in most systems in accordance with the second law of thermodynamics, some thermodynamic systems may lead to an increasing orderliness and self-organization (Prigogine 1961). These are thermodynamically open systems that operate far from thermodynamic equilibrium and can exchange energy, matter and entropy with the environment. Many of these self-organizing systems (such as the Bénard cells in boiling liquid and oscillating chemical reactions) were known a long time ago; however, the universality and generality of the processes involved in these systems was understood only with the works of Prigogine (1917–2003). It is believed that this ability for self-organization of physical systems led to the formation of complex hierarchical chemical and biological systems.

The flow of heat, entropy and material away from the interface during dry friction and wear can lead to self-organization when so-called ‘secondary structures’ form (Fox-Rabinovich & Totten 2006). The secondary structures are either patterns that form at the interface (e.g. stick and slip zones) or those formed as a result of mutual adjustment of the bodies in contact. The entropy production rate reaches its minimum in the self-organized state. Therefore, the self-organization is usually beneficial for the tribological system, as it leads to the reduction of friction and wear.

(ii) Entropy rate with account for the thermal conductivity

The net entropy growth rate for the frictional sliding of rigid bodies is given by equation (2.5). However, if instead of the net entropy, the entropy per surface area at the frictional interface is considered, the rate equation becomes more complicated. Heat \( dQ \) is generated at the interface in accordance with equation (2.4). The heat is dissipated away from the interface in accordance with the heat conduction equation

\[
\lambda \frac{\partial T}{\partial z} = \mu W V, \tag{2.10}
\]

where \( z \) is the vertical coordinate (distance from the interface) and \( \lambda \) is the heat conductivity. Consider a thin layer near the interface with thickness \( d_z \). The temperature drop across the layer is \( dT = (\mu W V / \lambda) dz \) (figure 2). The ratio of the heat released at the interface, \( dQ \), to that radiated at the bottom of the layer, \( dQ' \), is equal to the ratio of the temperatures at the top and bottom of the layers

\[
\frac{dQ'}{dQ} = \frac{T - \mu W V \ d_z / \lambda}{T}. \tag{2.11}
\]

Therefore, the energy released at the subsurface layer of depth \( dz \) is given by

\[
d_q = \frac{dQ - dQ'}{dz} = dQ \frac{\mu W V}{\lambda T} = \frac{(\mu W V)^2}{\lambda T}. \tag{2.12}
\]

Thus, the entropy in the subsurface layer, \( dS/dt = dq/T \), is given by

\[
\frac{dS}{dt} = \frac{(\mu W V)^2}{\lambda T^2}. \tag{2.13}
\]

Phil. Trans. R. Soc. A (2009)
The difference between equations (2.5) and (2.13) is that the latter takes into account the thermal conductivity and that equation (2.5) gives the net entropy rate, while equation (2.13) gives the rate of entropy in the subsurface layer. Note that the form for the thermodynamic flow is now $J = \mu W V$, and the thermodynamic force is $X = \mu W V / (\lambda T^2)$.

(iii) Criterion for self-organization

The stability condition for the thermodynamic system is given in the variational form by

$$\frac{\partial}{\partial t} (\delta^2 S) \geq 0,$$

where $\delta^2 S$ is the second variation of entropy. When equation (2.14) is not satisfied, the system is driven away from equilibrium, which creates the possibility for self-organization. Another form of equation (2.14) is given by (Fox-Rabinovich et al. 2007)

$$\frac{\partial}{\partial t} (\delta^2 S) = \frac{1}{2} \delta^2 \left( \frac{(\mu W V)^2}{\lambda T^2} \right) = \delta X \delta J = \delta (\mu W V) \delta \left( \frac{\mu W V}{\lambda T^2} \right) \geq 0. \quad (2.15)$$

Equation (2.15) is a powerful tool to study frictional contact. It involves the coefficient of friction, thermal conductivity and the sliding velocity. In the case of any interdependence between these values, the stability of the system should be analysed. Let us assume first that the coefficient of friction depends on $V$ and $T$, while the thermal conductivity depends on $V$

$$\mu = \mu(V, T) \quad \text{and} \quad \lambda = \lambda(V). \quad (2.16)$$

The stability condition given by equation (2.15) takes the form of

$$\frac{1}{2} \delta^2 S = \delta X \delta J = \frac{W^2}{\lambda T^2} \left( \frac{\partial \mu}{\partial V} V + \mu \right) \left( \frac{\partial \mu}{\partial V} V + \mu - \frac{\mu V}{\lambda} \frac{\partial \lambda}{\partial V} \right) (\delta V)^2$$

$$- \frac{2 V^2 W^2 \mu}{\lambda T^3} \frac{\partial \mu}{\partial T} (\delta T)^2 \geq 0. \quad (2.17)$$

The stability condition is violated either if the coefficient of friction grows with temperature ($\partial \mu / \partial T > 0$) or if, in the first term, the parentheses have different signs. The latter is possible if $\partial \mu / \partial V < 0$, whereas $\partial \lambda / \partial V > 0$, or if $\partial \mu / \partial V > 0$, 

Figure 2. Heat flow away from the frictional interface.
whereas $\frac{\partial \lambda}{\partial V} < 0$. When the stability condition is violated, the tribological system is likely to enter the self-organizing regime, with reduced friction and wear.

When the coefficient of friction and the thermal conductivity depend upon a material’s microstructure, it may be convenient to introduce a parameter $\psi$ that characterizes the microstructure of the surface (for example, the density of a micropattern), so that

$$\mu = \mu(\psi) \quad \text{and} \quad \lambda = \lambda(\psi). \quad (2.18)$$

The stability condition given by equation (2.15) takes the form of

$$\frac{1}{2} \delta^2 S = \frac{V^2 W^2}{T^2} \frac{\partial \mu}{\partial \psi} \left( \frac{1}{\lambda} \frac{\partial \mu}{\partial \psi} - \frac{\mu}{\lambda^2} \frac{\partial \lambda}{\partial \psi} \right) (\delta \psi)^2 \geq 0. \quad (2.19)$$

The stability condition can be violated if

$$\frac{\partial \mu}{\partial \psi} \frac{\partial \lambda}{\partial \psi} < 0. \quad (2.20)$$

It is known from non-equilibrium thermodynamics that, when the secondary structure is formed, the rate of entropy production reduces (Fox-Rabinovich et al. 2007). Therefore, if equation (2.20) is satisfied, the frictional force and wear can reduce. By selecting appropriate values of $\psi$ (e.g. the density of a micropattern), the condition of equation (2.20) can be satisfied. Note that the wear rate is related to the rate of surface entropy production

$$\frac{dw}{dt} = B \frac{dS}{dt} = YJ. \quad (2.21)$$

It is important to stress that the approach discussed in this section is not just a mere mathematical exercise, because it has been used successfully for many applications, such as ceramic and alloy hard coatings, analysis of material compatibility and many others (Fox-Rabinovich & Totten 2006). While it is well known that a tribological process enters a steady-state phase or a limiting cycle after a period of run-in and adaptation, the analysis is useful for selecting material parameters (such as the derivatives of $\mu$ and $\lambda$ by $V$, $T$ and $\psi$) that lead to desirable tribological behaviour.

(c) Multiscale structure of entropy

In §2b(iii), we studied the case when the excess entropy is driven away from the surface. A different situation will be investigated here. Many systems with dissipation have a multiscale structure, for example, microtextured surfaces or biological and biomimetics surfaces (Nosonovsky & Bhushan 2007b, 2008b). In these systems, entropy can be produced at one hierarchy level while consumed at a different level.

For example, three levels of hierarchy can be distinguished when grain growth in metals is considered (Nosonovsky & Esche 2008). Metal crystals form grains, characterized by different orientations of the lattice. The typical grain size is in the range of micrometres. There is an additional energy associated with grain
boundaries due to misorientation of neighbouring grains. Therefore, it is energetically profitable for larger grains to grow and to absorb smaller grains, thus reducing the total number of grains and the total boundary area. The grain growth is a random thermally activated diffusion-like process (Nosonovsky & Esche 2008). At the macroscale, the bulk properties, such as the yield strength and hardness, are essential; however, these properties are dependent upon average grain size, which is a parameter defined by mesoscale (microscale) structure. At the molecular scale (or nanoscale), energy barriers for grain growth and random thermal fluctuations exist. The grain growth is a self-organizing process system, which naturally evolves from a disordered to an ordered state, while being driven by random thermal fluctuations with an apparent violation of the second law of thermodynamics. To explain the paradox, the multiscale structure of entropy should be considered.

The net entropy can be presented as the sum of entropies associated with the structures and process at the corresponding scale levels

\[ S_{\text{net}} = S_{\text{macro}} + S_{\text{micro}} + S_{\text{nano}}. \]  

At the macroscale, no change of entropy occurs (\( \delta S_{\text{macro}} = 0 \)). At the mesoscale, the grain growth prevails; therefore, with the increasing grain size, the mesoscale entropy decreases due to the increased orderliness of the system (\( \delta S_{\text{meso}} < 0 \)). At the nanoscale, energy barriers for grain growth and random thermal fluctuations exist. Every time an atom is reoriented, a certain amount of energy is dissipated because the energy barrier must be overcome. This results in an increase in the system’s temperature (note that the heat is usually removed and the system is not closed), and entropy is created at the nanoscale (i.e. \( \delta S_{\text{nano}} > 0 \)). In other words, the random fluctuations at the nanoscale lead to an orderliness increase at the mesoscale, which is compensated by the entropy increase at the nanoscale. Computational models, such as Monte Carlo simulations of grain growth, usually take into account only the mesoscale level; not surprisingly, the orderliness of the system in these models increases in an apparent contradiction to the second law of thermodynamics (Nosonovsky & Esche 2008).

The same analysis applies to surface process (friction or wetting). These involve macroscale properties, such as the friction coefficient or water contact angles, the mesoscale properties (for example, when the surface has a pattern) and nanoscale properties related to the molecular interactions. Entropy can be created at the nanoscale while consumed at the mesoscale, leading to self-organization. Unlike in the case discussed in §2b(iii), the excess entropy is not radiated from the surface, but consumed at the nanoscale.

3. Self-organization and dynamic effects

In §2c, we discussed entropy production during friction and investigated the opportunity of self-organization when entropy is driven away from the surface or created at the nanoscale while consumed at the meso/macroscale. In this section, we will consider a specific type of self-organization called self-organized criticality (SOC) and the role of frictional dynamic instabilities in self-organization.

Phil. Trans. R. Soc. A (2009)
Self-organized criticality and avalanche dynamics

SOC is a concept in the theory of dynamic systems that was introduced in the 1980s (Bak 1996). The best-studied example of SOC is the ‘sandpile model’, representing grains of sand randomly placed into a pile until the slope exceeds a threshold value, transferring sand into the adjacent sites and increasing their slope in turn (figure 3). Placing a random grain at a particular site may have no effect, or it may trigger an avalanche that will affect many sites at the lattice. Thus, the response does not depend on the details of the perturbation. Note that the scale of the avalanche is much greater than the scale of the initial perturbation; thus, the avalanche belongs to the upper level of hierarchical organization. Unlike the self-organized systems studied in §2, SOC systems are constantly ‘tuned’ to the critical behaviour. The system tends to return to the state when an avalanche can be initiated. In the case of the sandpile, the slope of the sandpile tends to reach the critical threshold value. There are typical external signs of an SOC system, such as the power-law behaviour (the magnitude distribution of the avalanches) and the ‘one-over-frequency’ noise distribution (Bak 1996). The concept has been applied to such diverse fields as physics, cellular automata theory, biology, economics, sociology, linguistics and others.

The term ‘criticality’ originates in the physics of phase transitions, which provides another way of viewing SOC. Many physical systems have a critical point, i.e. a point at which a distinction between two phases vanishes. A typical length of fluctuations (referred to as the correlation length) tends to grow up to infinity near the critical point. Simple scaling relationships between various parameters of the system in the vicinity of the critical point can usually be established. These relationships are governed by power laws with certain critical exponents. For example, the critical point of water is at $T_c = 374^\circ C$ and $P_c = 218$ atm, and the distinction between liquid and gas water at these conditions disappears, so that no energy is needed to convert liquid water into vapour. At the critical point, the energy barrier vanishes. The systems with SOC have a critical point as an attractor, so that they spontaneously reach the vicinity of the critical point and exhibit power-law scaling behaviour. Since SOC allows a system to reach criticality spontaneously and without tuning the controlling parameter, it was suggested that it plays a major role in the spontaneous creation of complexity and hierarchical structures in various natural and social systems (Bak 1996). SOC was suggested to be responsible for landslides and earthquakes, because it is known that the number of earthquakes and their amplitude are related by a power law. In other words, a number of earthquakes with an amplitude greater than a certain level in a given area during a given period are related to that level by a power law. During earthquakes, the stress between two plates is accumulated for a long time and released suddenly in a catastrophic event, which is similar to the sandpile avalanche (Turcotte 1999).

In the case of dry frictional sliding, it has been suggested that a transition between the stick and slip phases during dry friction may be associated with the SOC, since the slip is triggered in a similar manner to the sandpile avalanches and earthquake slides. Zypman et al. (2003) showed that, in a traditional pin-on-disc experiment, the probability distribution of slip zone sizes follows the power law. In a later work, the same group found nanoscale SOC-like behaviour during atomic force microscopic studies of at least some materials.
(Zypman et al. 2003; Buldyrev et al. 2006). Thus ‘stick’ and ‘slip’ are two phases, and the system tends to achieve the critical state between them: in the stick state, elastic energy is accumulated until slip is initiated, whereas energy release during slip leads, again, to the stick state.

During wetting, SOC also apparently plays a role in wetting behaviours (di Meglio 1992; Nosonovsky & Bhushan 2008). Frictional stick–slip motion and liquid spreading are examples of ‘avalanche’-like behaviour, when a small input into the system leads to a big change. For example, a droplet on an inclined surface starts its motion when the tilt angle exceeds a certain critical value, so that the gradient of gravity exceeds the energy barriers associated with the pinning of the triple line. Wetted and dry states are two stable states with an energy barrier between them (figure 4). When the barrier vanishes, the liquid spreads. For example, when a sessile droplet is placed on an inclined surface, increasing the tilt angle increases the energy gradient due to the gravity, which overcomes the energy barrier at a certain critical value. The liquid starts to spread.

(b) Frictional dynamic instabilities and self-organization

Frictional sliding can lead to several types of instabilities due to the velocity dependence of the coefficient of friction, destabilization of interface elastic waves, thermal expansion and the effect of wear. These instabilities can potentially result in the formation of self-organized ‘secondary structures’, e.g. in the form of a train of slip pulses that may lead to the reduction of friction. The best-known manifestation of these instabilities is ‘squealing brakes’ (Kincaid et al. 2003). The stability condition for frictional sliding is that the second variation of dissipated energy is positive

$$\delta^2 Q \geq 0.$$  \hspace{1cm} (3.1)

Taking into account equation (2.1), note that equation (3.1), in most cases, is satisfied together with equation (2.14). Thus, frictional instabilities have potential for creating self-organized secondary structures.
(i) Instabilities due to velocity dependence of friction

When the coefficient of friction is dependent on sliding velocity, equation (2.14), together with equation (2.4), yields

\[ \frac{1}{2} \delta^2 S = \delta \left( \frac{\mu W}{T} \right) \delta(V) = \frac{W}{T} \frac{\partial \mu}{\partial V} (\delta V)^2 \geq 0. \]  

(3.2)

Thus, if \( \partial \mu / \partial V \geq 0 \) (the positive viscosity), the motion is stable; however, if \( \partial \mu / \partial V < 0 \) (the ‘negative viscosity’), the motion is unstable, and secondary structures can form. The negative viscosity effect is often found in systems with dry friction. The physical meaning of this conclusion is clear: if the friction force decreases with increasing velocity, the velocity will further grow, leading to the instability (Rice & Ruina 1983; Nguyen 2003). The process will continue until it will leave the linear region and enter a limiting cycle, which is likely to be manifested in the stick–slip motion. Such stick–slip motion can be viewed as a self-organized secondary structure.

(ii) Elastodynamic instabilities

Let us assume now that the coefficient of friction is constant; however, the contacting body can be deformed elastically. The sliding velocity and interface pressure can now be time dependent and consist of a constant and oscillating component.

**Figure 4.** (a) A water drop on an inclined surface. (b) As the tilt angle \( \alpha \) grows, the critical state can be reached with a disappearing energy barrier between the wetted and dry states, so that the liquid advances.
\[ V = V_0 + V_a \cos(\omega t) \]  \(3.3\)

and

\[ P = P_0 + P_a \cos(\omega t + \phi), \]  \(3.4\)

where \( V_0 \) and \( P_0 \) are average values and \( V_a \) and \( P_a \) are amplitudes of local velocity and pressure, respectively; \( \omega \) is the frequency; and \( \phi \) is the phase shift between the velocity and pressure. The average energy dissipation rate due to friction per unit surface area over one period of oscillation, \( T_0 \), is given by

\[ Q = \mu \int_0^{T_0} PV \, dt. \]  \(3.5\)

The energy dissipated at the interface can be smaller than the work of the friction force, \( \mu P_0 V_0 \), which leads to the instabilities found by Adams (1995) and Martins et al. (1995). The Adams–Martins instabilities occur during frictional sliding of two rough or smooth elastic half-spaces with a constant coefficient of friction (Nosonovsky & Adams 2004). The stability condition for the entropy rate is given by

\[ \frac{1}{2} \delta^2 \dot{S} = \frac{\mu}{2T} \delta^2 \left( \int_0^{T_0} PV \, dt \right) = \frac{\mu}{T} \int_0^{T_0} \frac{\partial P}{\partial V} \, dt \left( \delta V \right)^2 \geq 0. \]  \(3.6\)

These self-excited oscillations are confined to a region near the sliding interface and can eventually lead to either partial loss of contact or to propagating regions of stick–slip motion (slip waves). In a slip wave, a region of slip propagates along the interface, which is otherwise at the stick state. When a train of slip pulses propagates, two bodies shift relative to each other in a ‘caterpillar’ or ‘carpet-like’ motion (figure 5). This microslip can lead to a significant reduction of the observed coefficient of friction, as the slip is initiated at a shear load much smaller than \( \mu W \) (Nosonovsky & Adams 2001; Bhushan & Nosonovsky 2003).

(iii) **Thermoelastic instabilities**

Another type of friction-induced instability is thermoelastic instability (Barber 1969). Heat is generated during friction, and it leads to the thermal expansion of the material, which increases the contact pressure. The increased pressure results in increased friction force and excess heat generation, i.e. the instability. The stability condition for the entropy rate is given by

\[ \frac{1}{2} \delta^2 \dot{S} = \frac{\mu}{2T} \delta^2 \left( \int_0^{T_0} PV \, dt \right) = -\mu \int_0^{T_0} \frac{V \partial P}{T^2 \partial T} \, dt \left( \delta T \right)^2 \geq 0. \]  \(3.7\)

(iv) **The role of wear**

Another mechanism that may provide instability is the coupling between friction and wear. As friction increases, so does the wear, which may result in an increase in the real area of contact between the bodies and in a further increase in friction. The sliding bodies adjust to each other, and the process is known as frictional self-organization. On the other hand, wear produces smoothening of the surface distorted by the thermoelastic instability mechanism, and thus the wear and thermal expansion are competing factors, with the wear leading to the stabilization of sliding and the thermal expansion leading to the destabilization (figure 6).
4. Biomimetic self-healing and self-organized surfaces

In §§2 and 3, we discussed two situations that can lead to self-organization at the frictional interface: a simple entropy flow from the interface and a more complicated situation that involves a hierarchical organization with entropy production and flow at one level (e.g. the molecular scale) and entropy

Figure 5. The pattern of propagating stick and slip zones can lead to friction reduction, as $F<\mu W$.

Figure 6. Various mechanisms can create positive or negative feedbacks that lead to instabilities during friction.

Phil. Trans. R. Soc. A (2009)
consumption at another level (e.g. the mesoscale). The self-organizing and, in particular, self-healing surfaces are common for biological objects. Since it has been suggested that engineers can borrow some design principles from biology, we will discuss, in this section, typical features of biological and biomimetic self-healing and self-organized surfaces and their relationship to the mechanisms of entropy production and consumption, which have been investigated in the preceding sections.

**(a) Self-healing surfaces in biology**

Many materials and surfaces produced in living nature can achieve remarkable properties exceeding materials produced by humans. For example, spider silk has a breakage energy per unit weight two orders of magnitude greater than steel, and the abalone shell, a composite of calcium carbonate and organic material, is 3000 times more fracture resistant than a single crystal of the pure mineral (Smith et al. 1999). Furthermore, living creatures do not have access to high temperatures and pressures that humans use to produce, for example, metals. The key to these remarkable properties is the hierarchical organization of biological tissues and materials. Unlike man-made products that are built according to final design specifications, biological tissues are created with certain algorithms and adapt to the environment. One consequence of that is the ability of many tissues for self-repair and self-healing.

Healing is a complex process that involves both mechanical aspects (such as a steady replacement of a soft and tough material with stiffer and less tough tissue during bone healing) and biological stimuli (Fratzl & Weinkamer 2007). Healing also requires input of substances through the liquid (blood) circulating in the vascular system, which is difficult to reproduce in man-made materials.

In recent years, molecular-level mechanisms of self-healing have become the subject of active investigation. Fantner et al. (2006) found that the so-called sacrificial bonds in structural molecules play a significant role to increase the fracture toughness of biomaterials. Sacrificial bonds are the bonds that break before the main structural link (often the molecular backbone) is broken (Smith et al. 1999). These bonds are frequently weaker than the covalent bonds of molecular backbones and they provide a reversible, molecular-scale energy-dissipation mechanism. In many biological examples, the breaking of sacrificial bonds has been found to be reversible, thereby additionally providing a ‘self-healing’ property to the material. The force that causes an elastic response of a freely joint polymer molecule is often called the ‘entropic force’. This is a general name of forces caused by the system’s tendency to increase its entropy, irrespective of the character of a particular underlying microscopic interaction. Thus, the increase in entropy at the molecular level leads to reversible self-healing at the macroscopic level.

**(b) Biomimetic applications**

We will consider several ways of implementing self-organization and self-healing mechanisms. These include microcapsule encapsulation for self-healing and self-lubrication, surface patterning for self-lubrication and self-cleaning, and nanocomposites.
Encapsulation for self-repair and self-lubrication

Encapsulation has been used for crack damage repair for many years, e.g. epoxy capsules. When the crack propagates, the capsule ruptures and liquid adhesive is released that can heal the crack (figure 7). Crack propagation is an irreversible process, because when intermolecular bonds are broken, the energy $\gamma$ is released irreversibly, so the entropy amount $S_{\text{cr}} = \gamma / T$ per unit length is produced. Another way to see it is to consider the configurational entropy, rather than the thermal entropy. The ideal state without cracks corresponds to the minimum number of microstates and thus to the lowest possible configurational entropy. The crack can be formed in many different ways, and the cracked macrostate corresponds to many microstates and higher configurational entropy. In a similar manner, when a capsule ruptures and its content is released, the configurational entropy grows because mixing occurs. The configurational entropy growth of the mixing of two substances is given by

$$\Delta S_{\text{conf}} = -R(n_1 \ln X_1 + n_2 \ln X_2),$$

where $n_1$ and $n_2$ are the amounts in moles of two pure substances; $X_1$ and $X_2$ are mole fractions in the solution; and $R$ is the gas constant (Craig 1992). Part of this excess entropy can be consumed for healing the bonds at the crack. The net configurational entropy grows; however, the growth is not due to the cracking, but due to the microcapsule rupture and an irreversible decrease in their number.

If $N$ capsules are ruptured to heal the crack of length $L$, then the total entropy production is given by

$$\Delta S_{\text{total}} = N \Delta S_{\text{conf}} - L \frac{\gamma}{T} > 0.$$ 

This is a result of the multiscale structure of the material: the microcapsules at the microscale level help to heal the broken molecular bonds. The decomposition of the net entropy according to equation (2.22) explains the situation. It is noted that we, of course, considered a highly simplified model of crack healing, which in reality may include several stages, such as, in the case of polymers, surface rearrangement and approach, wetting, diffusion and randomization (Wool 2008).
The biomimetic idea of ‘artificial leucocytes’ that use microcapsules filled with nanoparticles has been suggested. The microcapsules consist of an elastic shell and enclose a solution of nanoparticles. These nanoparticles can diffuse from the interior of the capsule into the host fluid. While an imposed pressure gradient drives the capsule to move along the surface, the capsule moves slower or even gets ‘stuck’ at the edge of the damaged area. After a significant fraction of released nanoparticles have reacted with the damaged area, the adhesive strength increases and the capsule continues its motion along the surface (Balazs 2007). More complicated biomimetic solutions, such as artificial ‘blood flow’ vascular networks, have also been suggested.

Another approach is to use polymer chains in thin films. For sufficiently large nanoparticles that are compatible with the host matrix, the mobile chains expel the nanoparticles from the polymer film and into the cracks. The entropic force is responsible for this effect, as the chains gain greater conformational degrees of freedom by driving the particles to the surface and into the cracks (Balazs 2007). Gupta et al. (2006) observed this effect experimentally and showed that large nanoparticles dispersed in a polymer matrix migrate to a crack generated at the interface between the polymer and a SiO₂ layer.

Another area that required a multiscale analysis is the composite materials that are often used to alter tribological performance. This includes polymer composites (White et al. 2001), polymer–nanoparticle mixtures (Lee et al. 2003), nanocomposite hard coatings for severe tribological conditions (Fox-Rabinovich et al. 2007) and others.

(ii) Textured surfaces for self-lubrication

Besides using the microcapsules for repairing damaged surfaces, they can also be used to supply lubricant on the surface and thus to reduce friction and wear. As friction and wear increase, the capsules rupture, releasing a lubricant, which, in turn, decreases the friction and wear. Kumar et al. (2006) used a self-healing steel corrosion protection coating system with urea formaldehyde microcapsules (50–150 μm diameter) containing several types of film forming compounds (healants) and corrosion inhibitors mixed into commercially available coatings systems.

Another way to supply reservoirs with lubricant is surface texturing. While most engineering and natural surfaces are rough, it may be advantageous not to stay with the random roughness, but to texture a surface in a certain manner so that the useful properties of the surface, such as, for example, low friction and wear, and load capacity, improve. Surface texturing has become an object of intensive study in recent decades. Various techniques are used for surface texturing, including machining, ion beam texturing, etching, lithography and laser texturing. The texturing usually produces a large number of micro dimples on a surface that are effective in combination with lubrication. The dimples can serve as microhydrodynamic bearings, reservoirs for lubricant or traps for wear debris. Surface texturing is commonly used in magnetic storage devices and microelectromechanical systems to prevent adhesion and stiction (sticking of two components to each other due to adhesion). It is also used in the automotive industry for cylinder liner honing. At this point, most studies in the area of texturing are experimental and
concentrate upon finding the optimum size and distribution of the dimples. Thus, Wang et al. (2007) investigated the effect of dimple size (of the order of dozens of micrometres) and depth (below 1 μm) on sliding friction under boundary lubrication conditions. They found that, for a constant dimple surface, smaller and shallower dimples are more advantageous. Fabrication techniques for creating micro/nanoroughness include lithography (photo, E-beam, X-ray, etc.), etching (plasma, laser, chemical, electrochemical), deformation, deposition and others.

(iii) Self-cleaning

Self-cleaning in some respect is similar to self-lubrication. Self-cleaning is associated with the lotus effect: roughness-induced superhydrophobicity, water and dirt-repellency (figure 8a; Nosonovsky & Bhushan 2007a, b, 2008a). A superhydrophobic surface repels water and contaminants due to its micropatterned structure that provides heterogeneity altering surface properties such as the contact angle (figure 8b). Thus, an inhomogeneous microstructure (low configurational entropy) prevents contamination (growing entropy through mixing of the contaminant).

Phil. Trans. R. Soc. A (2009)
5. Conclusions

Entropy production rate provides a measure of dissipation during friction and wear. Unlike many other models, the entropy-based models of friction do not predict the value of the coefficient of friction. Instead, the entropy production rate is calculated for the coefficient of friction that may depend upon temperature, sliding velocity and other parameters. While, in most cases, irreversible dissipation and degradation occur at the frictional interface, in certain situations, self-organizing behaviour takes place. This is because friction is a non-equilibrium process. The self-organization is usually beneficial for friction and wear reduction because the entropy production rate is minimized in a self-organized state. It is therefore important to identify values of parameters that would lead to self-organization. We investigated two cases of self-organization: first, when entropy is driven away from the interface and second, when entropy is produced and driven away at the molecular scale while, at the macro/microscale, self-organization is observed. We suggested to decompose the entropy into macroscale, microscale and nanoscale fractions. Such decomposition allows to investigate self-organization taking place at one scale level of the system while dissipation occurs at another scale level. We also discussed specific mechanisms of self-organization: SOC and dynamic frictional instabilities.

Self-organization is most notorious in biological objects that have the abilities of adaptation and self-healing. It is therefore important to investigate and mimic natural mechanisms of self-healing at the interface. In biological systems, the decrease in entropy at the macroscale is usually compensated by its increase at the molecular level, so the same multiscale mechanism applies that is used in self-lubricating, self-healing and self-cleaning surfaces, such as those using microcapsules, polymer chains or surface micro/nanopatterning. The suggested thermodynamic model that uses the stability criterion for the rate of entropy production is therefore applicable to a broad range of interface processes, from wear and degradation to friction reduction due to self-organization to self-healing, self-organization and self-lubrications in sophisticated biomimetics interfaces. The model should be useful for the design of biomimetic self-healing surfaces.

References


