Chapter 1
Introduction

In the introduction we provide an overview of scientific disciplines and research areas that are relevant to the multidisciplinary topic of the book: self-healing materials, surface science and tribology, nonequilibrium thermodynamics, self-organization during friction, nonlinear effects in friction, composite materials, and biomimetics. We introduce basic concepts and definitions that are used throughout this book.

1.1 Self-Healing in Materials Science and Engineering

Engineering materials are usually divided into several classes according to their chemical structure and properties, such as metals, polymers, ceramics, and their composites. The physical properties of metals are determined by strong nondirectional metallic bonds between their atoms, which give them a combination of strength and formability by plastic deformation as well as a constant value of stiffness. The properties of ceramic materials are determined by strong directional ionic–covalent bonds that result in high stiffness, temperature-resistance, and brittleness. The properties of polymers are determined by two types of bonds: strong covalent bonds in the polymer chain, and weak sacrificial bonds between different chains or different parts of the same chain (Zwaag 2009). In composite materials, at least two materials from different classes are used to obtain a set of properties which is difficult to obtain when only one class of materials is used. One material usually serves as a matrix that holds the material together, and the other is a reinforcement that gives special desired properties to the composite.

The common feature of most engineering materials is that they tend to irreversibly deteriorate with time due to tear and wear, fatigue, creep, brittle fracture and formation of cracks, corrosion, erosion, change of chemical structure, and so on. On the other hand, biological tissues and materials in living nature often have the ability to repair minor and moderate damage due to their ability for healing and regeneration. The biological mechanisms of healing and regeneration are complex

M. Nosonovsky and P.K. Rohatgi, *Biomimetics in Materials Science: Self-Healing, Self-Lubricating, and Self-Cleaning Materials*, Springer Series in Materials Science 152,
DOI 10.1007/978-1-4614-0926-7_1,

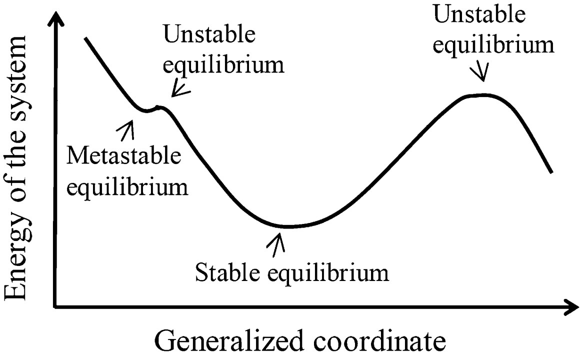


Fig. 1.1 A fragile metastable equilibrium is separated by a small energy barrier. When deterioration (e.g., cracking) occurs, the system leaves the metastable state

and lie beyond the scope of this book. However, in the 1990s a new field of artificial self-healing materials has emerged. These materials, inspired by the ability for self-healing in biological objects, are designed in such a way that they include mechanisms that can at least partially repair damage, such as voids and cracks, and partially or completely restore macroscopic material properties. Usually, a healing agent (often a liquid) is stored in the matrix of the material and when a cavity (a crack or void) is formed, it fills the cavity and closes it due to a chemical reaction or a phase transition (e.g., solidification). The healing can be autonomous (without human intervention) and nonautonomous. The latter require some external intervention, such as heating the material to trigger the repair process (Ghosh 2009; Zwaag 2007).

From the thermodynamic point of view, healing is achieved by shifting the system away from the thermodynamic equilibrium, which causes a restoring thermodynamic force (e.g., diffusion) to drive the system back to equilibrium. The restoring force also drives the healing process that is characterized by a local decrease of entropy (Nosonovsky and Bhushan 2009; Nosonovsky et al. 2009; Nosonovsky 2010a, b). Shifting the system away from equilibrium can be achieved by placing it in a metastable state (e.g., creating an oversaturated solution), so that the rupture breaks the fragile metastable equilibrium, and the system drives to the new most stable state (Fig. 1.1). The metastability can also be achieved by heating that causes a phase transition (e.g., melting, the martensite–austenite transition in metals).

Several strategies of embedding the self-healing properties into engineering materials have been suggested so far (Fig. 1.2). The most successful class of self-healing materials is the polymers, and this is because of their relatively large rates of diffusion and plasticity due to the presence of the cross-molecular bonds. One way to create self-healing polymers is to use thermosetting polymers that have the ability to cure (toughening or hardening by cross-linking of polymer chains), such as the thermosetting epoxy. Epoxy is a polymer formed by a reaction of an epoxide resin with a polyamine hardener. Epoxy can serve as a healing agent that is stored within thin-walled inert brittle macrocapsules embedded into the matrix along with a catalyst or hardener (but separately from the latter). When a crack propagates, the

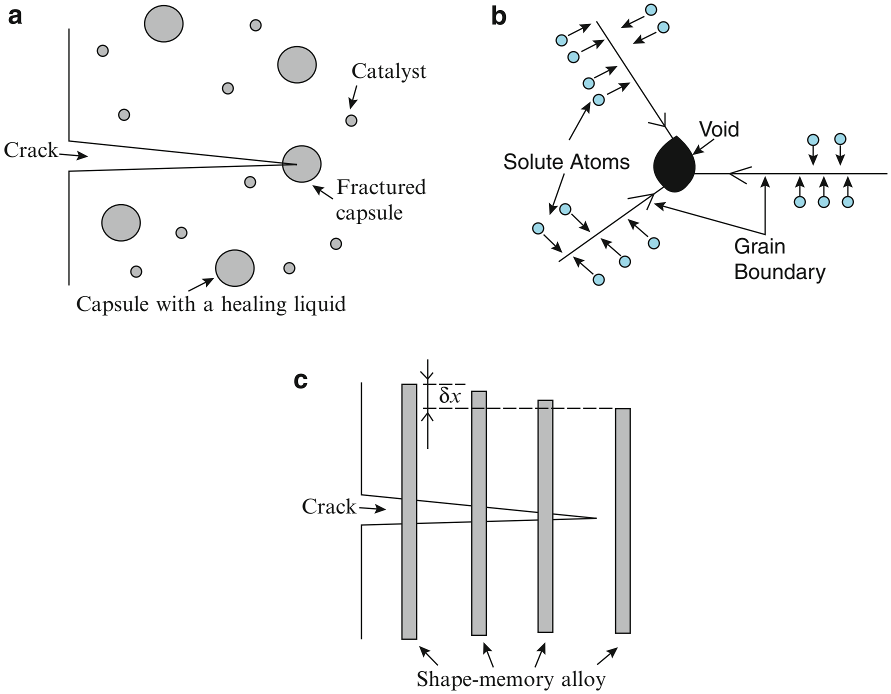


Fig. 1.2 Self-healing in metallic materials by (**a**) encapsulation of a healing agent, (**b**) precipitation in an over-saturated alloy, (**c**) embedding of shape-memory alloy microwires

capsules fracture, the healing agent is released and propagates into the crack due to capillarity. Then the healing agent mixes with the catalyst in the matrix, which triggers the cross-linking reaction and hardening of the epoxy that seals the crack (Fig. 1.2a). A different approach involves thermoplastic polymers with various ways of incorporating the healing agent into the material. In this approach, heating is often required to initiate healing. Self-healing ceramic materials often use oxidative reactions, because products of these reactions, including oxides, can be used to fill small cracks (Zwaag 2009).

It is much more difficult to heal metallic materials than polymers, because metallic atoms are strongly bonded and have small volumes and low diffusion rates. Currently, there are three main directions in the development of self-healing metallic systems. First is the formation of precipitates at the defect sites that immobilize further growth until failure. Van der Zwaag (2009) and co-workers called this mechanisms “damage prevention” because the idea is to prevent the formation of voids by the diffusion of the atoms to form precipitate from an oversaturated but under aged solid–solid solution (alloy). The driving mechanism for the diffusion is the excess surface energy of microscopic voids and cracks that serve as nucleation centers of the precipitate which plays the role of the healing agent (Fig. 1.2b). As a result, the newly formed void is sealed by the deposit of atoms in the form of precipitates before it grows and thus minimizes the creep and fatigue. Manuel and Olson (2007) and co-workers used another approach: reinforcement of an alloy matrix with a microfiber or wires made of a shape-memory alloy (SME), such as nitinol (NiTi). SMA wires have the ability to recover their original shape after some deformation has occurred if they are heated above the phase transformation temperature. If the composite undergoes crack formation, heating from the surface will activate the shape recovery effect of the SMA wires, and close the cracks (Fig. 1.2c). The third approach is to use a healing agent (such as an alloy with a low melting temperature) embedded into a metallic solder matrix, similarly to the way it is done with the polymers. However, encapsulation of a healing agent into a metallic material is a much more difficult task than in the case of polymers. The healing agent should be encapsulated in microcapsules which serve as diffusion barriers and which fracture when a crack propagates.

1.2 Surface Science and Tribology

Surface science is the study of physical and chemical phenomena that occur at the interface of two phases (solid–liquid, solid–gas, and solid–vacuum) or of different substances of the same phase (solid–solid, liquid–liquid) (Adamson 1990). Since properties of matter can change rapidly at the interface, it is convenient to assume that the interface is a 2D surface in a sense that every point at the interface can be characterized by only two parameters. In reality, every interface has nonzero thickness and the bulk properties change gradually at the interface; however, the thickness is so small compared to the two other dimensions, it can often be neglected.

An important characteristic of every surface or interface is the surface free energy, γ. In the bulk of the body, chemical bonds exist between the molecules or atoms. To break these bonds, certain energy is required. The molecules or atoms that do not form the bonds have higher potential energy than those that form the bonds. Molecules or atoms at the surface do not form bonds on the side of the surface and thus they have higher energy (Fig. 1.3). This additional energy is called surface or interface free energy and is measured in energy per area; in the SI system, $\mathrm{J\ m^{-2}}$ or $\mathrm{N\ m^{-1}}$. In order to create an interface (e.g., to form a vapor bubble inside boiling water), the energy should be supplied, which is equal to the area of the interface multiplied by the interface free energy (Nosonovsky and Bhushan 2008a).

Any system tends to achieve a position that corresponds to its minimum energy. For the stable existence of the interface, it is required that the free energy of formation of the interface be positive, so that accidental fluctuations would not result in the dispersion of one material into the other. This is the case for most solid–solid, solid–liquid, solid–gas, and liquid–gas interfaces. However, an interface between two gases or between miscible liquids does not offer opposition to the dispersion (Adamson 1990).

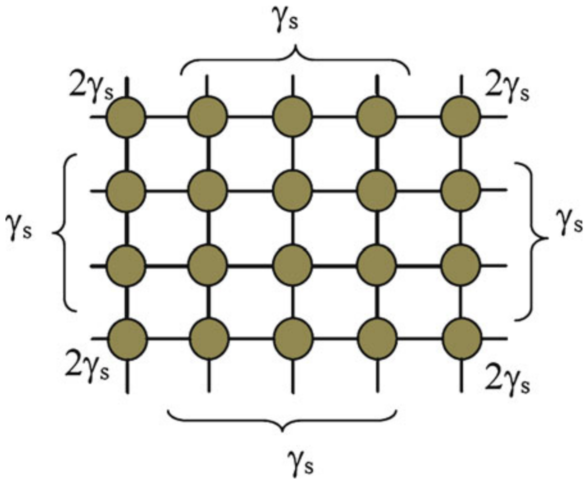


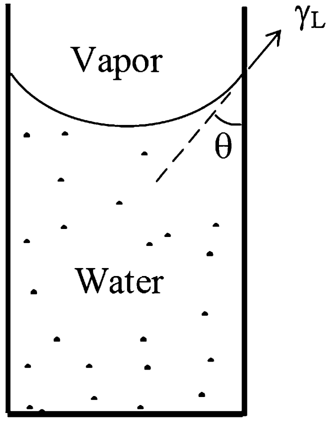
Fig. 1.3 Molecules at the surface have fewer bonds with neighboring molecules than molecules in the bulk. Therefore, surface molecules have higher energy. Edge molecules have even fewer bonds and higher energy, than surface molecules

The surface tension force is the force which should be applied to the solid–liquid–air contact line (the triple line) to expand the solid–liquid interface. The surface tension is measured in N m^{-1} and in many senses it is equivalent to surface free energy. The concept of the surface tension was introduced in 1805 by Thomas Young (1773–1829) and almost simultaneously by Pierre-Simon Laplace (1749–1827), while the idea of free surface energy was suggested by Josiah W. Gibbs (1839–1903) in about 1870. The accurate thermodynamic definition of the surface free energy involves a distinction between the concepts of the Gibbs free energy (the useful work obtainable from an isothermal isobaric thermodynamic system) and Helmholtz free energy (the useful work obtainable from a closed thermodynamic system).

The obvious manifestation of the free surface energy is found in the capillarity effect (Fig. 1.4), defined as the ability of a substance to draw another substance into it (Adamson 1990; Rowlinson and Widem 1982). When the size of a liquid droplet or a channel is much smaller than the so-called capillary length, given by $l_c = (\gamma/\rho g)^{1/2}$, where $g = 9.81\ \text{m s}^{-2}$ is the gravitational acceleration, γ is the free surface energy, and ρ is the density of liquid, the surface energy dominates over the gravity potential energy and the corresponding capillary forces dominate over the weight. For water at room temperature, $\gamma \approx 72\ \text{mN m}^{-1}$, $\rho \approx 1{,}000\ \text{kg m}^{-3}$, and $l_c \approx 2.7$ mm.

Interest in the surface phenomena is stimulated by current advances in nanoscience and nanotechnology. While the volume of a body is proportional to the third power of its linear size, the surface area is proportional to the second power of the linear size. With decreasing size of an object, the surface-to-volume ratio grows and surface effects dominate over the volume effects. This is why all surface phenomena, such as capillarity, adhesion, friction etc., become increasingly important for small objects (Nosonovsky and Bhushan 2008a), and the behavior changes below a certain size scale.

Fig. 1.4 The rise of water in a capillary tube due to the surface tension



Tribology is a technical discipline that studies contacting surfaces in relative motion, including adhesion, friction, lubrication, and wear. As opposed to the surface science, tribology is an application-oriented multidisciplinary area, which involves mechanics, physics, chemistry, materials science, fluid dynamics, heat transfer, biology, and other related areas. The word "tribophysics" (from the Greek word *tribos* "to rub") was coined in the 1940s by David Tabor (1913–2005), who was working in Australia (Dowson 1998). The intention of Bowden and Tabor was to alleviate the status of the friction, lubrication, and wear science and to attract public attention to the importance of these subjects for technology. The first official use of the term "tribology" was in 1966, in the "Jost report," when the British government reported on research in this area.

Although the word "tribology" appeared only about 40 years ago, people paid attention to friction and lubrication from ancient times. There is evidence that water lubrication may have been used in ancient Egypt as early as 2300 BC or 1800 BC, as shown in two examples (Nosonovsky 2007a). Oil lubrication is mentioned in the Bible in the King Saul story (1020 BC), animal fat may have been used as a lubricant for ancient Egyptian and Chinese chariots, and there is a list of lubricants in the treatise of Roman author Pliny the Elder (first century BC) (Dowson 1998; Nosonovsky 2007a). The emergence of the modern study of friction and lubrication is related to the activity of Leonardo da Vinci (1452–1519), Guillaume Amontons (1663–1705), and Charles August Coulomb (1736–1806), who formulated the empirical rules of friction.

It should be emphasized that despite the decades of intensive tribological studies in the twentieth century and the hundreds of years of investigation of friction, there is no general or axiomatic theory of friction. It is still viewed as a complex phenomenon dependent on different and unrelated mechanisms of dissipation, and governed by empirical laws, such as the Coulomb–Amontons law. Despite that, the Coulomb friction force plays a fundamental role in theoretical and applied mechanics, where frictional boundary conditions or constitutive laws are common (e.g., for the granular flow when the interaction between the grains of the material is governed by the dry friction law). The case of wear is different; although wear is almost as universal as friction, it is rarely considered an inherent part of models of mechanical contact. Instead, there are several phenomenological models of wear, the most popular of which is the Archard's law, which states that the volume of the worn material is proportional to the sliding distance, normal load, and a coefficient called "the wear coefficient," divided by hardness of a softer material in contact. Despite numerous models developed in the past 50 years that attempt to correlate tribological properties with material microstructure, there is no reliable way to always predict quantitatively the coefficient of friction or the wear coefficient from any microstructural material parameters. The reason for this is that most relevant interactions occur at the mesoscale, rather than at the atomic scale, and it is very difficult to take these mesoscale interactions into consideration.

The community of researchers who call themselves tribologists includes both practical engineers, chemists, and surface scientists who work on various aspects of surface characterization and modification, lubrication, etc., as well as mechanical engineers and mathematicians who work on contact mechanics. The boom of tribology occurred in the last third of the twentieth century stimulated by the space and arms race of the two superpowers (the USA and the Soviet Union). However, despite the fact that today many scientists, especially in the USA, consider the field of tribology (or at least the use of the term "tribology") shrinking, there are still many researchers who identify themselves as "tribologists." Modern tribology concentrates on such issues as rough surface topography, contact mechanics, adhesion, mechanisms of dry and lubricated friction, hydrodynamic (thick film) and boundary (thin film) lubrication, bearings, lubricant chemistry and additives, wear, surface texturing, and medical and biotribology.

Since the 1990s the new field of nanotribology has emerged due to the advances in nanotechnology. The idea of nanotechnology was suggested in 1960 by physicist Richard Feynman (1912–1985) who pointed out that no laws of physics prohibit manufacturing of very small devices, which would be able to perform many tasks considered earlier impossible. It took more than 40 years until Feynman's vision began to materialize, when the field of nanotechnology emerged in 1990s, stimulated by the discoveries of carbon nanotubes, C_{60} molecules (fullerenes), graphene (graphite monolayer), and quantum dots (Poole and Owens 2003; Nosonovsky and Bhushan 2008a).

Since the importance of surface effects is inversely proportional to the size of the investigated object, it is not surprising that tribological studies became an important area in nanotechnology. Nanotribology is concentrated on the study of the adhesion, friction, lubrication (in particular, by self-assembled molecular monolayers) and wear at the nanoscale (Nosonovsky and Bhushan 2008a). The main instrument currently used in nanotribological research is the atomic force microscope (AFM) and its variations such as the friction force microscope and various other scanning probe microscopes. In the AFM, a small cantilever (typical length is 100 μm or less) with a very sharp tip (typical radius is 10–30 nm) can scan the surface of a sample, the position of which is controlled with a great accuracy by a piezoelectric tube (Fig. 1.5).

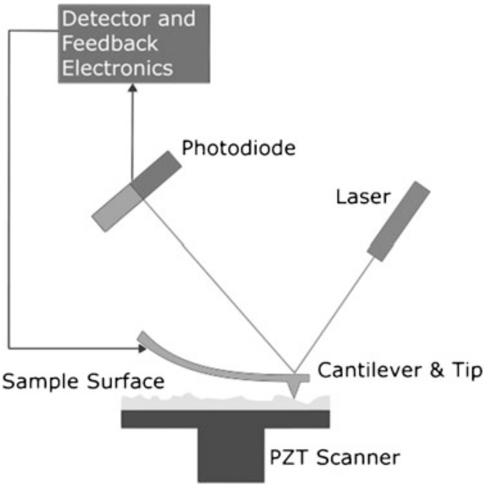


Fig. 1.5 The principle of the atomic force microscopy

The cantilever bends when in contact with the surface or under effect of forces (such as the adhesion and friction force) and its deflection is measured with a reflected laser beam. If the stiffness of the cantilever is known, it can be converted into the force acting upon the tip and thus small adhesion and friction forces can be measured with high accuracy.

1.3 Nonequilibrium Thermodynamics and Self-Organization

1.3.1 Classical Thermodynamic Potentials

Many physical processes result in irreversible energy dissipation. Examples include plastic deformation, friction, and viscosity. The energy during the dissipative process is converted into heat. According to the Second law of thermodynamics, formulated by Rudolf Clausius (1822–1888), heat, Q, cannot of itself pass from a colder to a hotter body. The mathematical formulation is the entropy S, defined as $\mathrm{d}S = \delta Q/T$, can only increase (in an irreversible process) or remain constant (in reversible processes). When heat δQ is transformed from a body at temperature T_1 to that at T_2, the change of entropy is $\delta Q(1/T_2 - 1/T_1)$, so the entropy grows when heat is transmitted from a hot to a cold body. The second law has a statistical nature and states that a system tends to transfer from a more ordered state to a less ordered state, which is more probable statistically. The state of thermodynamic equilibrium, at which the temperatures of the contacting bodies are equal ($T_1 = T_2$), corresponds to the less ordered, most probable state and to the highest entropy. At the nanoscale, when the typical energy of the system is comparable with kT, the second law can be violated due to small fluctuations, which can lead to local reductions of the entropy of the system (Anisimov 2004).

Ludwig Boltzmann (1844–1906) suggested in 1877 a definition of entropy using the statistical thermodynamics approach and the concept of microstates, so that entropy is proportional to the integral of the number of microstates $S = k \ln \boldsymbol{\Omega}$, where k is Boltzmann's constant and Ω is the number of microstates that correspond 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). Lower entropy corresponds to a more ordered state, and 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 grows.

Heat transfer and material transfer (diffusion) play a central role in many processes. Entropy is a thermodynamic potential that characterizes heat transfer while the mass transfer is characterized by the chemical potential, μ. The chemical potential is defined in such a way that particles tend to move from regions of high chemical potential to regions of low chemical potential.

For a rigorous definition of thermodynamic quantities, the concept of a thermodynamic system is the key. The system is considered isolated if there is no exchange of heat and mass with the environment and no work is done. The state of a thermodynamic system at any moment of time is characterized by several variables, so that the energy of a system depends on these variables. The total internal energy $U(S, V, N)$ is a function of the number of particles N, volume V, and entropy S. The temperature, pressure, and chemical potential are given by partial derivatives of U as

$$T = \frac{\partial U}{\partial S}, \quad P = -\frac{\partial U}{\partial V}, \quad \mu = \frac{\partial U}{\partial N}. \tag{1.1}$$

The change of the total energy is the given by

$$\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V + \mu\mathrm{d}N. \tag{1.2}$$

The quantities S, V, and N are generalized coordinates that characterize the state of the system in a unique way, while T, P, and μ are corresponding generalized forces, namely, the temperature, pressure, and chemical potential. When atoms or molecules of more than one type are present in the system, $N_1, N_2, \ldots, N_k$, several corresponding chemical potentials, $\mu_1, \mu_2, \ldots, \mu_k$, should be introduced as well.

Several laws of thermodynamics can be formulated now. The law of mass conservation states that in an isolated system the number of particles remains constant, or

$$\mathrm{d}N = 0. \tag{1.3}$$

The First law of thermodynamic is the law of energy conservation, and it states that in an isolated system the internal energy does not change,

$$\mathrm{d}U = -P\mathrm{d}V + T\mathrm{d}S + \mu\mathrm{d}N = 0. \tag{1.4}$$

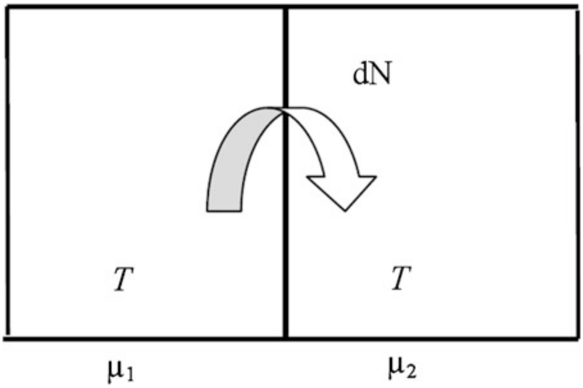


Fig. 1.6 Mass transfer between two systems at the same temperature T but different chemical potentials, $\mu_1 > \mu_2$

The Second law of thermodynamics states that entropy of an isolated system can only increase (in the case of an irreversible process) or remain constant (in the case of a reversible process) so that

$$\mathrm{d}S = \frac{P}{T}\mathrm{d}V - \frac{\mu}{T}\mathrm{d}N + \frac{1}{T}\mathrm{d}U \geq 0. \tag{1.5}$$

Suppose that a thermodynamic system is divided into two parts (subsystems) with the temperatures T_1 and T_2, so that the subsystem can exchange heat between themselves but not with the surrounding environment. The immediate consequence of Second law, as it was stated above, is that heat flows from the hotter to the colder system, since $\mathrm{d}S = \left(\frac{1}{T_2} - \frac{T}{T_2}\right)\mathrm{d}U > 0$, and, therefore, $T_2 > T_2$. In a similar manner, one can show that matter flows from the subsystem with a higher chemical potential to that with a lower one, assuming $T_1 = T_2 = T$, since $\mathrm{d}S = \left(\frac{\mu_2}{T} - \frac{\mu_2}{T}\right)\mathrm{d}N > 0$, and thus $\mu_1 > \mu_2$ (Fig. 1.6).

The volume, entropy, and the number of particles are called extensive thermodynamic properties of a system, since they are proportional to the mass of the system or the amount of substance. On the other hand, the pressure, temperature, and chemical potential are called intensive properties, since they are not dependent on the mass of the system or the amount of substance. Furthermore, the pairs P and V, T and S, μ and N are called "conjugate variables." In many cases, it is more convenient to present the state of a thermodynamic system as a function of pressure and temperature. For that, the thermodynamic potentials called Helmholtz free energy $A(V,T,N) = U - TS$ and Gibbs free energy $G(P,T,N) = U + PV - TS$ can be used instead of the internal energy. The differentials of A and G are defined as

$$\begin{aligned} \mathrm{d}A &= -S\mathrm{d}T - P\mathrm{d}V + \mu\mathrm{d}N, \\ \mathrm{d}G &= -S\mathrm{d}T - V\mathrm{d}P + \mu\mathrm{d}N. \end{aligned} \tag{1.6}$$

Note that the transformation from $U(S, V, N)$ to $A(V,T,N)$ and to $G(P,T,N)$ for a thermodynamic system is analogous to the transition from strain energy to the stress (or "complimentary") energy of a mechanical system. Indeed, for a system consisting of a mass M on a spring with the spring constant of k, the potential energy as a function of displacement from the state of equilibrium, x, is $U(x) = kx^2/2$ while the complimentary energy as a function of spring force F, is given as $\tilde{U}(F) = FX - U(x)$. Of course, for a linear system ($F = kx$) the values of U and $\tilde{U}$ coincide $\tilde{U}(F) = F_2/2k$. The force F and displacement x are the conjugate variables of a mechanical system.

The state of a thermodynamic system is called homogeneous if its properties, such as temperature, pressure, and chemical potential, do not depend on the spatial coordinates. The state is called stationary if the properties do not depend on time. If the properties do not depend on both time and spatial coordinates and no entropy is produced, the state of the system is called equilibrium. Throughout this book we deal with both thermodynamic and mechanical processes and with two types of equilibriums: thermodynamic and mechanical equilibrium. Mechanical processes are usually slower than thermodynamic processes and thus in most cases we assume that the system has enough time to achieve a thermodynamic equilibrium with a constant temperature, pressure, and chemical potential.

1.3.2 Nonequilibrium Thermodynamics

While classical thermodynamics studies systems at equilibrium, nonequilibrium thermodynamics received a lot of attention in the twentieth century. For a nonequilibrium thermodynamic system, the dependencies of thermodynamic parameters on time and spatial coordinated should be investigated. For that end, all intensive parameters can depend on time and coordinates $T(x, y, z, t)$, $P(x, y, z, t)$, and $\mu(x, y, z, t)$. For extensive parameters, their volume densities are considered $u(x, y, z, t)$, $s(x, y, z, t)$, and $c(x, y, z, t)$, where u is the energy density, s is the entropy density, and c is the concentration (number of particles per unit volume). Furthermore, flow vectors of thermodynamic quantities are introduced, so that the law of mass conservation is now given by

$$\frac{\partial c}{\partial t} + \nabla J^{\mathrm{c}} = 0, \tag{1.7}$$

where J^{c} is the mass flow. The First law of thermodynamics is given by

$$\frac{\partial u}{\partial t} + \nabla J^{\mathrm{e}} = 0, \tag{1.8}$$

where $J^{\mathrm{e}} = q$ is the heat flow q. Finally, the Second law of thermodynamics is given by

$$\frac{\partial s}{\partial t} + \nabla J^{\mathrm{S}} \geq 0, \tag{1.9}$$

where J^{S} is the entropy flow.

In the approach developed by Lars Onsager (1903–1976), generalized thermodynamic forces Y_i and thermodynamic flows J_i are associated with every generalized coordinate q_i, so that the rate of entropy density production is given by

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{1}{T}\sum Y_i J_i. \tag{1.10}$$

For example, for the heat flow the force is $Y_T = \nabla\frac{1}{T} = -\frac{1}{T^2}\nabla T$ (here used $\frac{\partial}{\partial x}\frac{1}{f(x)} = -\frac{1}{f(x)^2}\frac{\partial f(x)}{\partial x}$), for the mass flow is $Y_i = -\nabla\frac{\mu_i}{T}$, thus when entropy is produced due to heat and matter flow, using (1.7)–(1.8) we find

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\frac{q}{T^2}\nabla T + \frac{J^C}{\left(\frac{T}{\mu_i}\right)^2}\nabla\left(\frac{T}{\mu_i}\right) = \frac{K}{T}(\nabla T)^2 + D\frac{\mu}{T}\left(\nabla\frac{T}{\mu_i}\right)^2, \tag{1.11}$$

where $q = -K\nabla T$ (Fourier law of heat transfer) and $J^{\mathrm{C}} = -D\nabla\left(\frac{T}{\mu i}\right)^2$ (Fick's law of diffusion). Note that entropy rate is always positive. The generalized forces must have the same tensor rank as the flows, e.g., both should be vectors (or scalars, or tensors of the 2nd rank, etc).

In general, flows can be complicated functions of the generalized forces, $J_i = J_i(Y_1, Y_2, \ldots, Y_n)$; however, in the linear approximation, the flows are related to the forces by a linear equation

$$J_{\mathrm{k}} = \sum_i L_{ki} Y_i, \tag{1.12}$$

where L_{ki} is called the Onsager (phenomenological) coefficient (de Groot and Mazur 1962). Many physical linear empirical laws, such as the Ohm's law of electrical resistance, the Fourier law of heat conduction, and Fick's law of diffusion can be interpreted as linear relationships between generalized forces and flows.

Note that (1.12) provides the possibility of coupling of various irreversible processes inside the system, or their interaction with each other in the case when nondiagonal Onsager coefficients are different from zero. Due to coupling, a process can occur without its primary driving force or it may move the process in a direction opposite to the one imposed by its own driving force. For example, in thermodiffusion a species diffuses not because of a concentration gradient but because of a temperature gradient. When a species flows from a low to a high-concentration region, it must be coupled with a compensating process. The principles of thermodynamics allow the progress of a process without or against its primary driving force only if it is coupled with another process. However, one cannot simply assume that any two phenomena are coupled. Phenomenological coupling, if it exists, must be observed in nature. In this book, we interpret self-healing, which is observed in specially prepared materials with an embedded healing mechanism, as a result of coupling between deterioration and a force which affects healing.

1.3.3 Self-Organization

Ilya Prigogine (1917–2003) and his co-workers used the methods of nonequilibrium thermodynamics to investigate self-healing. They showed that so-called dissipative systems may lead to increasing orderliness and self-organization (Prigogine 1961). These systems are thermodynamically open and they operate far from thermodynamic equilibrium and can exchange energy, matter, and entropy with the environment. The dissipative systems are characterized by spontaneous symmetry breaking and formation of complex structures, where interacting particles exhibit long-range correlations. Examples of such systems are the Bénard cells in boiling liquid and oscillating chemical reactions. Many of these systems were known a long time ago; however, the universality and generality of the processes involved in these systems was understood only with the works by Prigogine. In 1945 Prigogine suggested that a nonequilibrium system tries to minimize its rate of entropy production and chooses the state accordingly. This condition was criticized by Landauer (1975), who argued that minimum entropy production is not in general a necessary condition for the steady state, and that the most favorable state of the system cannot be determined based on the behavior in the vicinity of the steady state, but one must consider the global nonequilibrium dynamics. Prigogine was awarded the Nobel Prize in chemistry in 1977 for his contribution to the theory of dissipative structures. Later studies have developed new techniques for analyzing pattern selection (Leppänen 2004).

It is believed that this ability for self-organization of physical systems led to the formation of complex hierarchical chemical and biological systems. Nonequilibrium dissipative systems may lead to the hierarchy, and their investigation involves the study of instability and loss of symmetry. Self-organization is related to an enormous reduction of degrees of freedom and entropy of the macroscopic system, consisting of many nonlinearly interacting subsystems, which macroscopically reveals an increase of order.

The related broad field is sometimes called complexity science, and the systems that are studied are between perfect order and complete randomness. Synergetics is an interdisciplinary science which explains the formation and self-organization of patterns and structures in systems far from thermodynamic equilibrium. These disciplines employ theoretical concepts similar to those of mesoscale physics, e.g., the Landau–Ginzburg functional and the order parameter (Haken 1993).

Even prior to the work of Prigogine, famous mathematician Alan Turing (1912–1954) investigated self-organization in the form of pattern-formation in the reaction-diffusion systems. He showed that a simple mathematical model describing spontaneously spreading and reacting chemicals could give rise to stationary spatial concentration patterns of fixed characteristic length from a random initial configuration. Turing (1952) proved rigorously that a chemical state, which is stable against perturbations in the absence of diffusion, may become unstable to perturbations in the presence of diffusion. The diffusion-driven instability is initiated by arbitrary random deviations of the stationary state and results in stationary spatially periodic variations in the chemical potentials, i.e., chemical patterns. His intention was to explain, at least qualitatively, the formation of biological patterns such as the stripes of a zebra or spots of a cheetah. Turing patterns were first observed in chemical experiments as late as 1990. Although the biological relevance of Turing's work is controversial even today, it had a significant impact on the development of nonlinear dynamics and nonequilibrium physics. Turing gave the first detailed description of a mechanism that can generate order in a nonequilibrium system. The self-oscillating chemical reactions were discovered in the 1950s–1960s by Belousov and Zhabotinsky (BZ) in Russia. They studied a mix of several ingredients, in which the ratio of concentration of the cerium (IV) and cerium (III) ions oscillated, causing the color of the solution to oscillate between a yellow solution and a colorless solution. It is remarkable that Belousov was not able to publish his discovery for a long time, because it was believed that such reactions are not possible; the BZ reactions became known in the West only in the 1970s. Unlike the patterns in the BZ oscillating chemical reactions, Turing patterns are stationary in time (Leppänen 2004).

An important example of self-organization that has been studied extensively by P. Bak (1996) is "self-organized criticality," which implies that a system tends to spontaneously achieve a critical point-like behavior. Self-organized criticality is typical for many systems where energy is accumulated for a long time, and the catastrophic release of energy is triggered by a random event, such as earthquakes, landslides, snow avalanches, or collapses of piles of granular materials. The classic example is the "sand pile model," representing a pile of sand where adding a single grain of sand can have no effect, or it can trigger a catastrophic avalanche. Self-organized critical behavior has certain characteristic quantitative features (such as the power law of amplitude distribution of the avalanches and "one-over-frequency noise"), which can be indicative of the presence of the self-organized critical behavior without a detailed analysis of the dynamics of the system. Apparently, self-organized criticality plays a role in such effects as stick–slip dry frictional motion and wetting of rough solid surfaces.

1.4 Self-Organization During Friction

Sliding friction is a nonequilibrium process that results in the dissipation of large amounts of energy and the flow of heat, entropy, and material move away from the frictional interface. Therefore, sliding friction does not always lead to wear and deterioration, but has potential for self-organization. It is well known that when the sliding is initiated, friction and wear are usually high during the initial "running-in" period. However, with time, the surfaces "adjust" to each other due to wear, relaxation, and other processes, so friction and wear decrease. This is perhaps the simplest example of frictional self-organization, which also shows why the system that underwent the transition to a self-organized state has a lower friction and wear rate.

This self-organization can occur in the form of so-called “selective transfer” which was investigated by Russian tribologists Kragelsky and Garkunov. In the 1950s, they studied the technical state of airplane chassis which included a bronze-steel lubricated frictional system, and found that a protective copper film can form, which reduces the wear to very small values. The copper film is formed by the anodic dissolution of bronze (an alloy of copper and tin with additive elements). The additives, such as iron, zinc, aluminum, and tin, dissolved in the lubricant, while the copper formed a film on the surfaces of the contacting materials. The film is in a dynamic equilibrium, while the contacting layers are worn and destroyed, new layers of copper are formed, resulting in the virtual absence of wear and the friction force reduced by an order of magnitude. A similar effect can be achieved through the diffusion of a copper ion dissolved in a lubricant. The authors called this effect the “selective transfer” of copper ions, and the protective effect, “servovit film” (Garkunov 2004). More common terms in modern literature are the “in situ formation of self-lubricating films” and “the third body effect,” indicating that film is formed in situ and becomes the “third body” trapped between the two contacting bodies.

Thermodynamically, the system is driven away from equilibrium by frictional sliding. The system reacts producing a generalized thermodynamic force that acts to restore the equilibrium, which causes a diffusion (precipitation) of copper ions.

A different approach to using the concept of entropy was developed since the 1980s by a group of Russian tribologists, including N. Bushe, L. Bershadsky, Kostetsky, I. Gershman, and others, who applied nonequilibrium thermodynamics to the frictional contact (Fox-Rabinovich and Totten 2006). The stability condition for the thermodynamic system is given in the variation form by Fox-Rabinovich et al. 2007

$$\delta^2 S > 0, \tag{1.13}$$

where $\delta^2 S$ is the second variation of entropy. When (1.13) is satisfied, the system is at equilibrium. However, when this condition is violated, the system is driven away from equilibrium, which creates the possibility for self-organization and therefore, to reduce friction and wear. In the self-organized state, friction and wear can be significantly reduced. If the coefficient of friction and the thermal conductivity depend upon a material’s microstructure, it may be convenient to introduce a parameter ψ that characterizes the microstructure of the surface (e.g., the density of a micropattern), so that $\mu(\psi)$, $\lambda(\psi)$. These dependencies are substituted into the stability condition (1.13) in order to determine whether the self-organization is possible. For example, in nanocrystalline AlTiN coatings for sever operating conditions (e.g., hard coatings for cutting tools) both the thermal conductivity and the coefficient of friction decrease with decreasing size of grains, which leads to self-organized regime with low friction and wear (Fox-Rabinovich et al. 2007). We understand such a self-organized regime as one of the modes of self-lubrication.

The term “self-lubrication” implies the ability of parts and components to operate without lubrication, and refers to several methods and effects, in addition to the reduced friction or wear mentioned above of. Among these methods are the use of self-lubricating coatings that are either hard (to reduce wear) or with low surface energy (to reduce adhesion and friction). Besides coatings, self-lubrication refers to the development of metal-, polymer-, or ceramic-based self-lubricating composite materials, often with a matrix that provides structural integrity and a reinforcement material that provides low friction and wear. The nanocomposites have become a focus of this research, as well as numerous attempts to include carbon nanotubes and fullerene C_{60} molecules.

Simple models assume that these large molecules and nanosized particles serve as "rolling bearings" that reduce friction; however, it is obvious now that the mechanism can be more complicated and involve self-organization. Dynamic self-organization is thought to be responsible for self-lubrication in atomic force microscopy experiments with the atomic resolution. A protective layer can be formed through the chemical reaction of oxidation or a reaction with water vapor. For example, a self-lubricating layer of boric acid (H_3BO_3) is formed as a result of a reaction of water molecules with a B_2O_3 coating (Erdemir et al. 1990). Another type of self-lubricating material involves lubricant embedded into the matrix, e.g., inside microcapsules that rupture during wear, and release the lubricant. Surface microtexturing provides holes and dimples that can serve as reservoirs for lubricant is another method of providing self-lubrication. In addition, we should mention that self-lubrication is observed in many biological systems (e.g., human joints), and that the term "self-lubrication" is used also in geophysics where it refers to the abnormally low friction between tectonic plates that is observed during some earthquakes.

1.4.1 Frictional Dynamic Effects

Another type of self-organizing behavior during friction is related to dynamic frictional effects, such as vibrations and waves induced by frictional sliding. Two types of elastic waves can propagate in an elastic medium: shear and dilatational waves. In addition, surface elastic waves may exist, and their amplitude decreases exponentially with the distance from the surface. For two slightly dissimilar elastic materials in contact, the interface waves (Rayleigh waves) may propagate at the interface zone. Their amplitude decreases exponentially with the distance from the interface. Whereas the interface waves occur for slightly dissimilar (in the sense of their elastic properties) materials; or very dissimilar materials, waves would be radiated along the interfaces, providing a different mechanism of pumping energy away from the interface (Nosonovsky and Adams 2001). Self-excited oscillations, such as Rayleigh waves, are confined to a region near the sliding interface and can eventually lead to either partial loss of contact, or to the propagation of 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. This microslip can lead to a significant reduction of the observed coefficient of friction, as the slip is initiated at a shear load, which is much smaller than the coefficient of friction times the normal load (Nosonovsky and Adams 2001; Bhushan and Nosonovsky 2003; Nosonovsky and Bhushan 2005a, b).

A linear model (e.g., one based on the linear elasticity) can be used to analyze the stability and onset of nonlinear vibrations; however, in order to find the resulting regime a complicated nonlinear analysis is needed. The train of stick–slip propagating regions can be viewed as a self-organized structure that reduces frictional dissipation.

When discussing the nonlinear model of friction, it is noted that despite the apparent simplicity of the Coulomb friction law, incorporating it with fundamental theories of mechanics, such as dynamic elasticity, causes a number of difficulties. This is in part due to the intrinsic nonlinearity of the Coulombs law with respect to the sliding velocity (the friction force jumps from zero at zero velocity to a finite value at a nonzero velocity). There are a number of paradoxes related to the combining of the Coulomb friction with the mechanics of an absolutely rigid or deformable body. These include the Painlevé paradoxes and other types of paradoxes that lead to the nonuniqueness or nonexistence of a solution. In a strict mathematical sense, the Coulomb friction is inconsistent not only with rigid body dynamics, but also with the dynamics of elastically deformable bodies (Nosonovsky and Bhushan 2008a).

One of the manifestations of that inconsistency is frictional dynamic instability, in the form of self-excited vibrations where the amplitude grows with time (Adams 1995). These instabilities exist due to a positive feedback (coupling) of friction with thermal expansion, wear, or negative velocity depending on friction. However, such instabilities can exist even when there is no velocity-dependence of friction, no thermal expansion, and wear is present. These instabilities are a consequence of energy being pumped into the interface as a result of the positive work of the driving force. As a result, the amplitude of the interface waves grows with time, with the unlimited rate of growth of short wavelength (the ill-posedness). In a real system, of course, the growth is limited by the limits of applicability of the linear elasticity and linear vibration theory. In order to regularize the ill-posedness, it has been suggested to use a dynamic friction law instead of the Coulomb friction law (Ranjith and Rice 2001). One example of such a dynamic friction law is the state-and-rate friction law, according to which the coefficient of friction at first grows every time the load changes, and then decreases to the steady-state value (which by itself decreases with increasing sliding velocity). Dynamic friction laws have a high potential for self-organization dynamics.

1.5 Composite Materials for Tribological Applications

Composite materials are often used when the production of materials with desired tribological properties, such as low friction and wear, is needed. Composite materials usually consist of two or more components, with one material serving as a matrix and others as reinforcements that provide special properties, such as high strength, to the matrix.

An important class of composite materials is the metal matrix composites (MMCs). An MMC is a composite with at least two constituent parts, one of which (the matrix) is a metal whereas the other (the reinforcement) is either a different metal or another material, such as a ceramic. When more than two different materials are incorporated simultaneously into the matrix, the composite is called a *hybrid composite*. The reverse structure to an MMC composite, a ceramic matrix with a metallic reinforcement and a high percentage of the ceramic phase, is called a *cermet*.

MMCs are created by dispersing a reinforcing material into a metal matrix. More than two components can be required, for example, when the need arises to coat the surface of the reinforcement material to prevent a chemical reaction with the matrix. Thus, carbon fibers are commonly used in an aluminum matrix to synthesize composites with low density and high strength. However, a chemical reaction of carbon with aluminum generates a brittle and water-soluble compound, Al_4C_3, at the interface between the matrix and the fiber. To prevent this reaction, the carbon fibers are coated with nickel or TiB_2.

The matrix is the monolithic material into which the reinforcement is embedded, and the matrix is generally a continuous phase. This means that there is a path through the matrix to any point in the material, unlike two materials sandwiched together. In structural applications, the matrix is made of a lighter metal such as aluminum, magnesium, or titanium, since it provides a compliant support for the reinforcement. For high-temperature applications, cobalt and cobalt–nickel alloy matrices are common.

The reinforcement material is embedded into the matrix. The reinforcement does not always serve a purely structural task (reinforcing the compound), but is also used to change physical properties such as wear resistance, friction coefficient, or thermal conductivity. The reinforcement can be either continuous or discontinuous.

1.6 Biomimetics

Biomimetics (or biomimicry) means mimicking biological objects in order to design artificial objects with desirable properties (Bar-Cohen 2005). The word is attributed to biophysicist Otto Schmitt (1913–1998) who was a polymath and, among other problems, attempted to produce a physical device that explicitly mimicked the electrical action of a nerve. Since the late 1950s the term “biomimetics” (Vincent et al. 2006) became popular. Another term that is frequently used, “bionics,” was coined in 1960 by Jack Steele of the USA air force. The term “biomimetic” has replaced, to a certain extent, the term “bionics” in scientific literature. The latter became quite popular in science fiction. The word biomimetics first appeared in Webster’s dictionary in 1974 and is defined as “the study of the formation, structure or function of biologically produced substances and materials (as enzymes or silk) and biological mechanisms and processes (as protein synthesis or photosynthesis) especially for the purpose of synthesizing similar products by artificial mechanisms which mimic natural ones". The idea of creating artificial devices mimicking living nature has inspired humankind since its very early days. Leonardo da Vinci studied birds' flight and designed machines, but never made any (Vincent et al. 2006). The motif is found in the Greek myth of Daedalus and Icarus, who mimicked bird wings to create an artificial flying device. Mechanical birds, animals, and even humans are found in many ancient legends, for example, the Talmud speaks about mechanical golden animals and peacocks in the throne of King Solomon, as well as an artificial human-like dummy (later called Golem) created by Cabbalistic magic.

Biological objects have been studied actively by physicists and chemists before the emergence of biomimetics. However, biomimetics goes further than just biophysics and bioengineering, which only study biological objects, since its objective is to imitate the objects with desirable properties. There is a growing interest toward biomimetics and growing public awareness of the field, which is stimulated in part by a general interest in "green" (environmentally friendly) manufacturing. An example of such growing interest is the database of biomimetic solutions for various engineering problems that can be found at the Web site www.AskNature.org.

The idea behind biomimetics is that nature's technical solutions, achieved by thousands of years of evolution, are perfect or at least better than those which contemporary engineering technology can suggest. This concept may be applied to various areas of engineering, for example, the artificial intelligence and neural networks in information technology are inspired by the desire to mimic the human brain. The existence of biocells and deoxyribonucleic acid (DNA) serves a source of inspiration for nanotechnologists who hope to 1-day build self-assembled molecular scale devices. In the field of biomimetic materials, there is also a whole area of bio-inspired ceramics based on seashells and other materials.

It is important to distinguish between simple *similarity* in the functionality of a biological and an engineered object (e.g., a plane is similar to a bird because it flies and has wings and a tail, a tank is similar to a turtle because the tank has armor and the turtle has a shell) and a true borrowing of *principles* and approaches used by nature in the design of artificial devices.

The best-known and most practically successful biomimetic effect related to surface science is the lotus effect. The term was coined in recognition of the fact that lotus plant leaves are water-repellent, and have the ability to emerge clean from dirty water due to a special structure of their surface. The term refers to the superhydrophobicity and self-cleaning abilities of the plant due to surface roughness. Although an accurate definition of superhydrophobicity still remains a matter of argument in the literature, the consensus is that a superhydrophobic surface should have a high contact angle with water (higher than 150°) and low contact angle hysteresis, so that a water droplet placed on a superhydrophobic surface can easily flow upon it.

Besides the lotus effect, a number of ideas have been suggested so far in the field of biomimetic surfaces (Gorb 2005; Bar-Cohen 2005; Favret and Fuentas 2010). These include the gecko foot, which has a very high and adaptive adhesion, the moth eye that does not reflect light, the sharkskin which can flow underwater suppressing turbulence, the water strider leg that can stay dry, the darkling beetle that can collect dew using hydrophilic microspots, and the sand skink that can reduce friction using nanothresholds. The common feature found among many of these surfaces is that they have hierarchical roughnesses, with rough details ranging from nanometers to millimeters. This observation inspired our earlier study of multiscale frictional dissipative mechanisms in combination with hierarchical surfaces (Nosonovsky and Bhushan 2008a).

The field of self-healing materials is often considered a biomimetic study, because a number of ideas to mimic the biological mechanisms of healing have been suggested. This includes, for example, creating a vascular system for the circulation of a healing agent, or using nanoparticles as artificial leucocytes that can deliver and unload a healing agent to a damaged zone (Balazs 2007; Gebeshuber et al. 2008). However, biological mechanisms of healing are very complex and often involve many factors acting in parallel. There is a growing understanding in the community of materials scientists who work on the development of self-healing materials that biological mechanisms of healing cannot be directly borrowed for artificial materials. For instance, the review by Van der Zwaag (2009) with the symptomatic title "Self-healing behavior in man-made engineering materials: bioinspired but taking into account their intrinsic character" ends with the conclusion "while Nature has shown an extreme diversity in microstructures and microstructural systems, and, subsequently, has shown a wide diversity of healing mechanisms, it is unwise to try to copy these healing mechanisms in man-made engineering materials in a direct manner. As shown in this chapter, engineering materials have their own characteristics, and, in designing self-healing behavior in such materials, their intrinsic or natural character has to be taken into account."

Biological materials and surfaces are known to have remarkable properties. For example, the silk fiber of the spider web has specific strength comparable with or even exceeding steel. In order to produce such strong materials as steel, humans need to use large facilities with high temperatures and pressures. A spider produces its silk at ambient temperature and pressure, while the length of a continuous fiber exceeds the size of the spider thousands of times. Such remarkable properties are achieved due to a hierarchical organization of tissues produced by living organisms. The hierarchy provides flexibility and the ability to adapt to changing environments. Biological tissues are created in accordance with certain recursive algorithms contained in their DNA, and not with final design specifications or blueprints as is the case in the conventional engineering design paradigm. These recursive algorithms lead to hierarchical organization. In the case of functional surfaces, hierarchical organization is manifested in the presence of structures with different characteristic length scales. These structures usually provide the ability to simultaneously protect the surface from

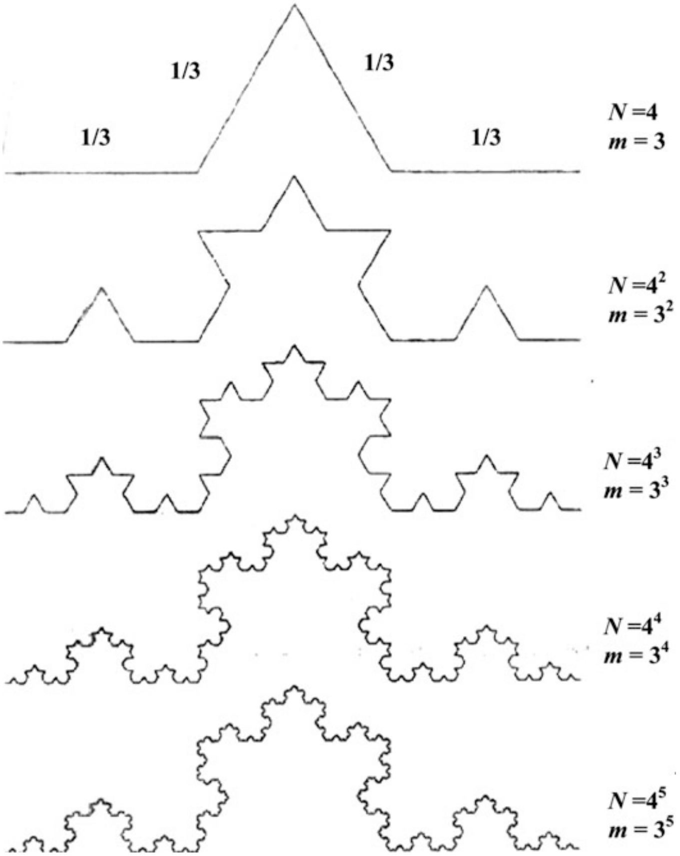


Fig. 1.7 The fractal curve (so-called Koch curve) with a fractal dimension $D = \ln 4/\ln 3 = 1.26$. The curve is built by an iterative procedure so that at every step its length l is increased by the factor of 4/3. If the curve is zoomed by three times, it length is increased by $4 = 3^D$ times

several undesirable factors having different characteristic length scales. Hierarchical structures play important roles in both the lotus effect and the gecko effect.

Not surprisingly, hierarchical structures are of great importance for biomimetic engineering. The concept of hierarchy is different from the concept of scale in a sense that hierarchy implies a complicated structure and organization. Hierarchical surfaces are built from elements of different characteristic lengths, organized in a certain manner (Fig. 1.7). This organization leads to certain functionality. Many examples of these surfaces are found in biology and will be considered in this book. An important class of hierarchical systems is the fractal objects. Self-similar, or fractal structures, can be divided by parts, each of which is a reduced-size copy of the whole. Unusual properties of the self-similar curves and surfaces, including their noninteger dimensions, were studied by mathematicians in the 1930s. The word *fractal* was coined in 1975 by Benoit Mandelbrot, who popularized the concept of self-similarity, and showed that fractal geometry is universal in nature and engineering applications (Mandelbrot 1983). The fractal concepts were applied to rough surfaces. In the late 1980s to early 1990s, the fractal geometry approach was introduced in the study of the engineering of rough surfaces.

1.7 Summary

In the introductory chapter, we discussed some fundamental concepts related to self-organization and surface science. In the following chapters, we review self-healing, self-lubricating, and self-cleaning materials.