Chapter 2
Introduction to Friction

In this chapter we aim at giving a short overview of dry friction, i.e. frictional phenomena where the lubricants effect is negligible. We first present the phenomenological laws derived by experimental observations, then present the rudiments of the (incomplete) theory of friction.Excellent references on these topics are [Per00, PT96, Kri02]. In the process, we comment on the existing literature and draw some conclusions about possible directions for future work, especially for the statistical physics community.

2.1 The Phenomenological Laws of Friction

Consider a solid parallelepiped—as depicted in Fig. 2.1—in contact with a large solid substrate over a surface S (supposed to be flat at the macroscopic scale), with a normal load L (for instance due to gravity), being pulled along the surface via a spring k_0, itself pulled at a fixed velocity V_0. The block's velocity is denoted v. The force F_k of frictional effects was[1] claimed to follow these three laws:

- First law: F_k is independent from the surface area S.
- Second law: F_k is proportional to the normal load: $F_k \propto L$.
- Third law: F_k is independent of the sliding velocity v.

This allows to write a phenomenological equation for the friction force:

$$F_k = \mu_k L \tag{2.1}$$

[1]These laws were stated in the 17th century by Amontons for the first two of them, and in the 18th century by Coulomb for the third one.

F.P. Landes, *Viscoelastic Interfaces Driven in Disordered Media*,
Springer Theses, DOI 10.1007/978-3-319-20022-4_2

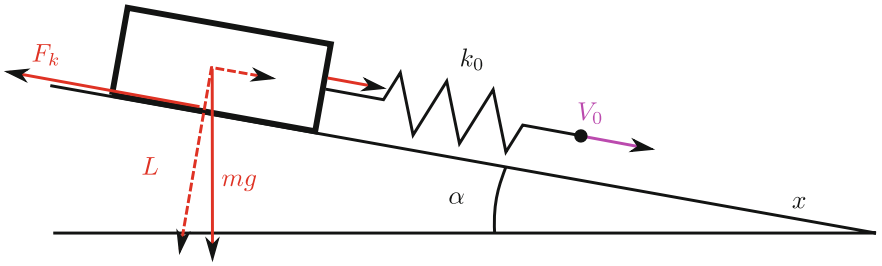

Fig. 2.1 *Solid block* sliding on a solid substrate. Solid parallelepiped sliding on an inclined plane (angle α) at velocity $v = \dot{x}$. The weight can be decomposed in two components, one orthogonal to the surface (the load L), and one parallel to it (which contributes to the pulling). Additional pulling can be provided via a spring k_0, of which the "free" end may be moved at a fixed velocity V_0. The kinetic friction force is denoted F_k

where μ_k is the kinetic (or dynamic) friction coefficient, which depends on the nature of the surfaces in contact along with many other things, but which is here assumed to be independent from S, L and v.

There is one "exception" to the third law which is commonly observed: for the static case ($v = 0$, i.e. when there may be pulling, but without motion) the friction coefficient takes a different value μ_s, larger than the dynamical one: $\mu_s(v = 0) > \mu_k(v > 0)$.

2.1.1 Stick-Slip Motion

Due to the fact that the static ($v = 0$) friction force is higher than the dynamic ($v > 0$) one, a mechanical instability known as "stick and slip motion" can occur, especially when the pulling is provided mainly in a sufficiently flexible way (small k_0) or at sufficiently low driving velocity V_0. As we are going to see, this is something that we experience on a daily basis.

Consider the system pictured in Fig. 2.1, with an angle $\alpha = 0$, for simplicity. The free end of the spring k_0 is denoted w_0 and is driven steadily at a velocity V_0. The spring k_0 can be thought of either as an actual spring through which the driving is performed, or as an effective representation for the bulk rigidity of the solid. As we pull the block from the side, we transmit some shear stress through its bulk. If the solid is driven at constant velocity V_0 directly from a point on its side, the effective stiffness k_0 is proportional to the Young's modulus E and inversely proportional to the height d of the driving point (neglecting torque effects). See Fig. 2.2 for a visual explanation.In the context of a simple table-top experiment as presented here, the solid's stiffness is generally too large for stick-slip to occur, so that the use of an actual spring k_0 to perform driving is useful.

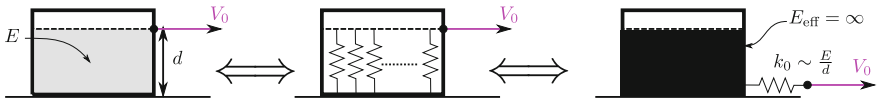

Fig. 2.2 Effective stiffness of the driving spring. *Left* a *solid block* with Young's modulus E is pulled rigidly from some point at a height d, i.e. this point is forced to have the velocity V_0. *Middle* the *solid block* can be pictured as a dense network of springs, related to E. Springs in the *horizontal* directions are not pictured for clarity. *Right* effective modelling by a block with infinitely rigid bulk, pulled by an effective spring $k_0 \sim E/d$

Newton's equations for the center of mass of the block at position x can be written in the dynamic and static cases:

$$m\ddot{x} = k_0(V_0 t - x) - \mu_k L \qquad \text{(dynamic)} \tag{2.2}$$

$$0 = k_0(V_0 t - x) - F_s \qquad \text{(static)} \tag{2.3}$$

where the static friction force F_s adapts according to Newton's second law (Law of action and reaction) in order to balance the pulling force, as long as it does not exceed its threshold: $|F_s| < \mu_s L = (F_s)_{\max}$.

We start with $x(t = 0) = 0, w_0(0) = 0$, and for $t > 0$ we perform the drive, $w_0 = V_0 t$. As long as $|F_s| < \mu_s L$, the block does not move: we are in the "stick" phase.

At time $t_1 = \frac{\mu_s L}{k_0 V_0}$, the static friction force F_s reaches its maximal value $\mu_s L$ and the block starts to slide. This is the "slip" phase. Thus we have the initial condition $x(t_1) = 0, \dot{x}(t_1) = 0$ for the kinetic equation. The solution reads:

$$x(t) = V_0(t - t_1) - \sqrt{\frac{m}{k_0}} V_0 \sin\left(\sqrt{\frac{k_0}{m}}(t - t_1)\right) + \frac{(\mu_s - \mu_k)L}{k_0}\left(1 - \cos\left(\sqrt{\frac{k_0}{m}}(t - t_1)\right)\right). \tag{2.4}$$

It is natural to take a look at the short-time limit of the solid's position:

$$x(t) \underset{t\sim 0}{\sim} \frac{(\mu_s - \mu_k)L}{2m} t^2 + \frac{k_0 V_0}{6m} t^3 - \frac{(\mu_s - \mu_k) L k_0}{24 m^2} t^4 + o(t^4), \tag{2.5}$$

which is increasing at short time, as expected, since $\mu_s > \mu_k$.

As x initially increases faster than $V_0 t$, the driving force from the spring, $(k_0(V_0 t - x))$, decreases over time, so that $\dot{x}$ may reach zero again. If at some point $\dot{x} = 0$, the kinetic friction coefficient is replaced by the static one, and oscillations (and any form of further sliding) are prevented. We can compute the times t_2 such that formally, $\dot{x}(t_2) = 0$:

$$t_2 = t_1 + 2\sqrt{\frac{m}{k_0}}\left(p\pi - \arctan\left(\frac{(\mu_s - \mu_k)L}{\sqrt{m k_0} V_0}\right)\right) \tag{2.6}$$

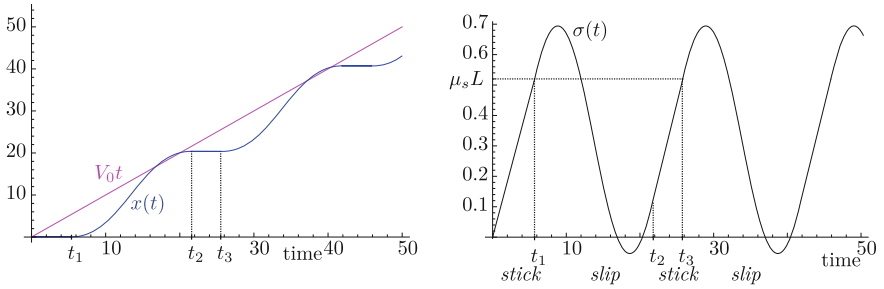

Fig. 2.3 Stick-slip evolution of the block over time. *Left* variations of the center of mass x over time t (*solid blue*) computed from (Eq. 2.4). *Right* saw-tooth evolution of the stress during stick-slip motion. Variations of the stress $\sigma = k_0(V_0t - x)$ (*solid grey line*) computed from (Eq. 2.4). The function V_0t (*dashed purple*) is given for reference. At time t_1, the threshold for the static force is reached and the block starts to move, with a decreased friction force F_k (kinetic). At time t_2, as velocity cancels, one needs to consider the static friction force. Loading then increases until the time t_3 where the threshold of static friction is once again reached. Parameters used for the two figures are: $m = 1$, $V_0 = 1$, $k_0 = 0.1$, $\mu_S L = 0.52$, $(\mu_S - \mu_K)L = 0.2$. Note that the slip phase seems long, but this is due to the parameters used: in particular, with a larger $(\mu_S - \mu_K)$ we get longer stick phases (and—relatively—shorter, sharper slip phases) Here we have a detailed view of the slip phase

where $p \in \mathbb{N}$. The physical solution corresponds to the first positive time that can be obtained, i.e. $p = 1$. At this time, the friction force (that always opposes motion, whichever direction it goes) increases from $\mu_k L$ to $\mu_s L$ and motion stops. The evolution of the block is once again controlled by the static equation of motion (Eq. 2.3), and we are in the "stick" phase.

The system will remain in the stick state until the time t_3 such that $V_0t_3 - x(t_3) = \mu_s L/k_0$. Since the system has no memory (beyond $\dot{x}$), the dynamics at ulterior times is exactly periodic, as shown in Fig. 2.3.

In friction experiments, one usually measures the *total shear stress* or total friction force, which is given by $\sigma = k_0(V_0t - x)$. We present the evolution of $\sigma(t)$ in Fig. 2.3 (right), to be compared with experimental results, e.g. for a mica surface pulled at constant velocity (Fig. 2.4).

The difference between μ_s and μ_k generates a mechanical instability, in which the elastic energy provided by the driving is at times stored (static case, or "stick" phase) and at times released over a short[2] period (kinetic case, or "slip" phase). This is the exact opposite of the more common situation of dissipative forces monotonously increasing with velocity so that a balance between drive and drag naturally yields stable solutions.

[2]Note that in Fig. 2.3, the parameters chosen are such that the stick phase is rather short. For larger $(\mu_S - \mu_K)$ we get longer stick phases, and—relatively—shorter slip phases, since the duration of the slip phase is independent of μ_S, but the loading time grows essentially linearly with it.

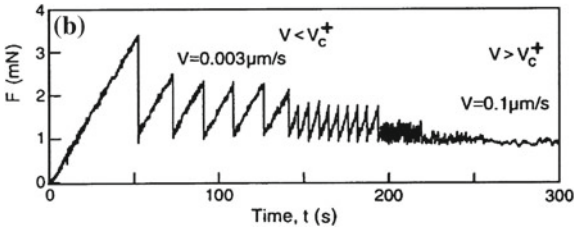

Fig. 2.4 From [Per00]. Stick-slip motion of mica surfaces coated with end-grafted chain molecules (DMPE). The driving velocity is set to a few different values over time, while the stress or friction force (here denoted F, measured in mN) is measured.When the spring velocity (v or V_0) increases beyond v_c^+ the sliding motion becomes steady. Here $v_c^+ \approx 0.1\ \mu s^{-1}$

Scope: Limiting Behaviours in k_0 and V_0

In the limits $k_0 \sim \infty$ or $V_0 \sim \infty$ we can derive simple analytical expressions, which allow to estimate the range of relevance stick-slip motion.

Duration of the Slip Phase The duration of the slip phase $t_2 - t_1$ is obtained by developing (Eq. 2.6):

$$\begin{aligned} t_2 - t_1 &\underset{k_0 \sim \infty}{\sim} 2\sqrt{\frac{m}{k_0}}\pi + O(k_0^{-1}) \\ t_2 - t_1 &\underset{V_0 \sim \infty}{\sim} 2\sqrt{\frac{m}{k_0}}\pi + O(V_0^{-1}) \end{aligned} \tag{2.7}$$

This means that the duration of the slip phase vanishes when $k_0 \sim \infty$, but remains finite when $V_0 \sim \infty$.

Duration of the Stick Phase To fully predict how stick-slip behaviour depends on the parameters k_0 and V_0, we need to compare the durations of the slip and stick phases. The recurrent stick phase has duration $t_3 - t_2$, which is different from t_1 because the initial condition we used is different from the system's state at $t = t_2$ (the spring is not extended at all at $t = 0$, it is fully relaxed). Starting from $t = t_2$ with (Eq. 2.3), the static friction force will reach its threshold at the time t_3 such that $V_0 t_3 - x(t_2) = \mu_s L/k_0$ (we used $x(t_2) = x(t_3)$). We thus have

$$t_3 - t_2 = \frac{\mu_s L}{k_0 V_0} + \frac{x(t_2)}{V_0} - t_2. \tag{2.8}$$

It is useless to fully write down the exact value of $x(t_2)$, obtained by injecting (Eq. 2.6) in (Eq. 2.4). Instead, we only give the relevant limits:

$$x(t_2) \underset{k_0 \sim \infty}{\sim} V_0\, 2\sqrt{\frac{m}{k_0}}\pi + O(k_0^{-2}), \qquad x(t_2) \underset{V_0 \sim \infty}{\sim} V_0\, 2\sqrt{\frac{m}{k_0}}\pi + O(V_0^{-2}), \tag{2.9}$$

i.e. the first[3] order term of both developments happens to be the same. In this (common) term, we recognize the previous developments of (Eq. 2.7):

$$\frac{x(t_2)}{V_0} \underset{k_0 \sim \infty}{\sim} (t_2 - t_1) + O(k_0^{-1}), \qquad \frac{x(t_2)}{V_0} \underset{V_0 \sim \infty}{\sim} (t_2 - t_1) + O(V_0^{-1}), \quad (2.10)$$

where the dominant corrections come from (Eq. 2.7). We can inject these expressions in (Eq. 2.8): $t_3 - t_2 = \frac{\mu_s L}{k_0 V_0} + \frac{x(t_2)}{V_0} - t_2 = t_1 + \frac{x(t_2)}{V_0} - t_2$:

$$t_3 - t_2 \underset{k_0 \sim \infty}{\sim} O(k_0^{-1}) \qquad t_3 - t_2 \underset{V_0 \sim \infty}{\sim} O(V_0^{-1}) \quad (2.11)$$

This means that for a sufficiently rigid spring k_0 or a sufficiently high velocity V_0, the duration of the stick phase vanishes.

Existence of Stick-Slip More precisely, we see that in these limits, the duration of the slip phases is always large compared to the duration of the stick phase. For $k_0 \sim \infty$, $T_{\text{slip}} \sim k_0^{-1/2} \gg T_{\text{stick}} \sim k_0^{-1}$. For $V_0 \sim \infty$, $T_{\text{slip}} \sim O(1) \gg T_{\text{stick}} \sim V_0^{-1}$. We can conclude that in these limits, the system looses its stick-slip behaviour. In this very simple model, we did not include any viscous term of the form $-\eta \dot{x}$, and the friction law was assumed to be very simple. The addition of viscosity gives a sharper decrease of the stress in the slip phase, and smooths the displacement, which tends to suppress the stick-slip. In more refined models, one may find a critical value of the spring stiffness, k_0^c (which depends on V_0), as is observed in most experiments.

The steady state can be obtained very simply by assuming a stationary behaviour. Using the kinetic equation: $0 = k_0(V_0 t - x) - \mu_k L$, we get:

$$x(t) = V_0 t + \frac{\mu_k L}{k_0}. \quad (2.12)$$

Examples of Stick-Slip in Everyday Life

There are too many examples of natural occurrences of stick-slip motion to make a comprehensive list here: we are only going to name a few.

The sound of squeaking doors originates from a motion of the hinge of stick-slip kind. The sudden motion during each slip phase produces a sound pulse, and the periodicity of the stick-slip provides sound waves with a rather well-defined frequency. The fact that the phenomenon is not exactly periodic does not prevent us from classifying it as stick-slip, as the driving is still essentially monotonous. We may notice that the computations from the previous section are validated by our everyday experience: the sound of a squeaking door can often be suppressed by

[3] Actually, many higher-order terms are also equal in both developments. This is also true for the developments of $t_2 - t_1$.

opening or closing it fast enough. This is what could be expected from the fact that when $V_0 \sim \infty$, the stick-slip behaviour is suppressed.

The same kind of mechanism applies to grasshoppers which produce their characteristic noise by rubbing their femur against their wings (or abdomen). The physics is essentially the same as for squeaking doors, only at different length scales.

The bow of a violin also produces sound waves in a similar way (but it's a bit more complex, and of course the resonance of the violin's string plays an important role too).

The sudden stop of a car also involves stick-slip. Car brakes tend to squeal when pressed too hard: by the same mechanism as above, the gentle and rather noiseless sweep of the brake pads against the wheel (pure sliding) is then replaced by a high-pitched noise (stick-slip). This could be expected from (Eq. 2.6), where we see that an increase in the load L is similar to a decrease in V_0, thus enhancing stick-slip behaviour. The tires on the road can also (unfortunately) perform a sort of stick-slip: when the brakes are pushed so hard that they lock up the wheels (pure stick in the brake-wheels system), the tires will slide on the road (instead of rolling, i.e. sticking to the road). In that case, the stick state corresponds to tires normally rolling, and the slip state corresponds to a sudden slip on the road, which can induce wear of the tires (loss of material and irreversible deformations) and "skid marks". However, the intermittent behaviour (which defines stick-slip) is usually just due to an intermittent braking, so that the regularly spaced skid marks seen on roads are mostly not directly related to stick-slip, but rather are the consequence of the use of an Anti-lock Braking System.

We quickly mention a case of lubricated friction that has important implications in human health: bones articulations. In this system, stick-slip causes more damage than steady slip, something that can further increase the occurrence of stick-slip [LBI13].

In all of the above examples of stick-slip motion, the whole "parallelepiped" is considered as a single block. But stick-slip actually occurs on many different length scales. Thus, even when the motion of the center of mass seems smooth, local "stick-slips" usually occur at the interface between the sliding solid and its substrate: for instance, groups of molecules or surface asperities can "jump" quickly in a stick-slip like fashion. During "steady" sliding, these local slip events occur asynchronously, so that they essentially average out at the macroscopic level. These local events may be probed indirectly, for instance, by studying the elastic waves emitted from the sliding interface.

We will give more details on these local events and their relevance for macroscopic friction in the following sections, but the impatient reader might jump directly to Sect. 2.2.3.

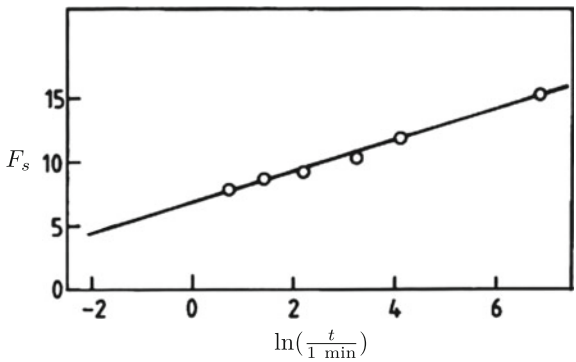

Fig. 2.5 Static friction force versus $\ln(t)$. C.A. Coulomb's data (*circles*) is compared a simple law $A + B\ln(t)$ (*solid line*). Data taken from [Dow79], retrieved from [Per00]

2.1.2 Ageing and Violation(s) of the Third Law

Observations

Ageing in Static Friction As early as the 18th century, C.A. Coulomb measured and observed an increase of the static friction coefficient with the time of contact with the substrate. He found that the time dependence was rather well fit by a law $F_s = A + Bt^{\alpha}$, with $\alpha \approx 0.2$ (see Fig. 2.5). However, in a more modern view one notices that his experimental data is also well fit by $F_s = A + B\ln t$, which is essentially the currently widely-accepted law for the ageing of contact in many materials.[4] From these rudimentary results, we see that the strength of the contact initially increases quickly, but the time to double from ∼10 (arbitrary units) to ∼20 can be extrapolated to be of ∼1 h. More recent results about the ageing of contact at rest can be found e.g. in [BDRF10].

This time-dependence of the static friction with time of stationary contact is very important both in applications and conceptually. It could almost be nicknamed the 4th law of friction, due to its importance.

Velocity Weakening The third law is actually quite incorrect: how could friction be independent from the sliding velocity v, and at the same time, have a singularity at $v = 0$? Upon closer inspection there is no singularity, but a smooth behaviour connecting the $v = 0$ and the very small velocity regimes (as one would expect from intuition), via a friction force which decreases when the velocity increases (a rather counter-intuitive observation). Typically, in the case of steady-state motion, the velocity-dependent friction law can be expressed in its most simplified form by:

$$\mu_k = \mu^* - A\ln\left(1 + \frac{v}{V^*}\right). \tag{2.13}$$

[4] The common way to write this equation nowadays is rather $\mu_s = A + B\ln(1 + t/t_0)$, a notation that better preserves the need for homogeneity and the hatred for divergences.

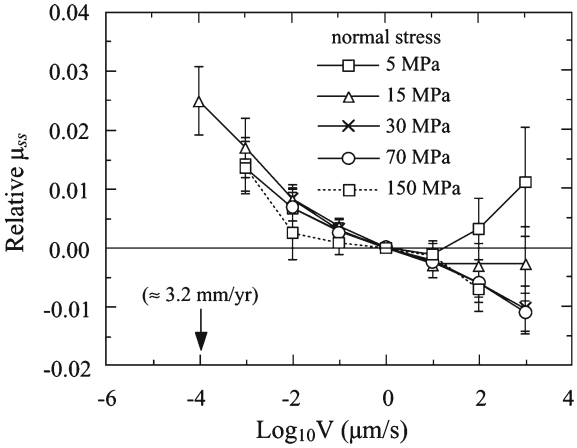

Fig. 2.6 From [KBD93]. The relative variations of the steady state friction coefficient μ_{ss} at different velocities. *Each curve* corresponds to a normal stress (*symbols*). For loads larger than 30 MPa, a logarithmic velocity weakening can be detected (approximately a *straight line* with negative slope). For smaller loads of 5 and 15 MPa, there is velocity strengthening for $v > V^*$

We further discuss the physical interpretation of this equation in the next subsection (p. 16). For bare granite (see [KBD93]) parameters values range in the scales: $\mu \sim O(1)$ (typically $\mu^* \approx 0.6$), $V^* = 1\ \mu\text{m s}^{-1}$, and $A \sim O(10^{-2})$. These parameters can be extracted from experiments where steady-state sliding is obtained for various velocities, at different loads or other external conditions varied. Keeping the same setup for different velocities, one is especially interested in the relative variations of the Steady State friction coefficient $\mu_{ss} = \mu - \mu^* = -A \ln(1 + v/V^*)$, as shown in Fig. 2.6.

However, for most materials this continuous decrease can only be observed at very small velocities ($\sim 10\,\mu\text{m s}^{-1}$, see Fig. 2.6), and one needs rather good instruments to detect it in the lab. This also explains why it was not detected earlier. An example of the crucial role of this weakening of friction with increasing driving velocity is found at the level of Earth's tectonic plates: as the imposed driving $\sim V_0$ is very small, plates perform stick-slip motion, with the slip phases corresponding to earthquakes. The fact that friction is decreasing up to a limit velocity means that any initial motion of the plate triggers an instability which drives it up to this limiting velocity. Understanding this instability of the statics is an important aspect of geophysics. In the geophysicists' community, this decrease of friction with velocity is known as the *velocity-weakening* effect.

Velocity Strengthening Let's mention also the *velocity strengthening* regime (where friction increases with velocity) which is expected to occur at a higher velocity (which depends on other parameters as the load): see the right part of Fig. 2.6. It is tempting to attribute velocity strengthening to viscous or hydrodynamical effects due to lubricants. Actually, in the presence of lubricants the hydrodynamic theory predicts a friction force going as $\sim v^2$ at high Reynolds numbers (i.e. at high velocities). Furthermore, velocity strengthening can appear at much lower velocities via mechanisms completely independent from hydrodynamics. A more reasonable explanation for velocity strengthening is the wear, which increases roughly linearly

with velocity. Wear may also produce an abundance of granular materials between the surfaces, which may also dissipate more energy by increasing the contacts and the sliding-induced deformation. In this thesis we are only interested in the small velocity regime, and it is enough to know that beyond some limiting velocity, friction starts to increase instead of decreasing. For a presentation of additional experimental results on various materials displaying velocity strengthening and some arguments to explain its origin, see [BSSBB14].

The Rate- and State-Dependent Friction Law(s)

From these diverse observations came the need to have a single constitutive law (or empirical law) that would encompass both the observed time dependence of static friction *and* the velocity dependences of kinetic friction (velocity weakening or strengthening). We now present this general phenomenological law.

In the general case of non stationary sliding velocity $v(t)$, the friction coefficient can be expressed in terms of the so-called *rate and state friction law* [Die79, Rui83], where "rate" simply refers to the time derivative ($\dot{x} = v$) of the position and "state" refers to an internal variable which represents the quality of the contacts between the sliding solid and its substrate, $\theta(t)$ (also sometimes denoted $\phi(t)$). A widely used form for the evolution of the variables μ, θ is:

$$\mu = \mu^* + a \ln\left(\frac{v}{V^*}\right) + b \ln\left(\frac{V^*\theta}{D_c}\right) \tag{2.14}$$

$$\frac{\partial \theta}{\partial t} = 1 - \frac{v\theta}{D_c} \tag{2.15}$$

where typically, $V^* = 1\,\mu\mathrm{m\,s}^{-1}$, $\mu^* \approx 0.5$, $D_c \sim 1$–$10\,\mu\mathrm{m}$, and a, b are dimensionless constants that need to be fit for each particular data set, but typically range in $a, b \sim O(10^{-3})$. This is what is often called a "constitutive relation" for friction. We may note that (Eq. 2.14) is undefined at $v = 0$. This can be problematic for computations, but this is compatible with the definition of friction as the normalized shear strength of a surface: there must be some slip at some scale for it to be measured. Anyhow, (Eq. 2.14) is sometimes rewritten as

$$\mu = \mu^* + a \ln\left(1 + \frac{v}{V^*}\right) + b \ln\left(1 + \frac{V^*\theta}{D_c}\right) \tag{2.16}$$

to tackle this issue.

The above law is just one of several possible rate-dependent and state-dependent friction laws (RSF laws). Many variations are possible for the evolution of the state variable θ. Keeping (Eq. 2.14), we can have two other RSF laws by using one of these evolution equations for θ:

$$\frac{\partial \theta}{\partial t} = 1 - \left(\frac{v\theta}{D_c}\right)^2, \tag{2.17}$$

$$\frac{\partial \theta}{\partial t} = -\frac{v\theta}{D_c} \ln\left(\frac{v\theta}{D_c}\right). \tag{2.18}$$

Each of these will give different behaviours when looking in details, but some of the main features are shared:

- In the steady state ($\partial_t \theta = 0$), we obtain $\theta^{ss} = D_c/v$. Injecting it into (Eq. 2.14), we get the steady state friction coefficient $\mu^{ss} = \mu^* + (a-b)\ln(v/V^*)$. Depending on the sign of $a-b$, we will get velocity weakening or strengthening.
- In the case of zero velocity ($v = 0$), θ is a monotonously increasing function of time. For instance, starting from $\theta(0) = 0$, (Eq. 2.15) gives $\theta(t) = t$. This allows to account for the reinforcement of static friction over time.

These two shared features exactly answer to the initial need to reconcile static and dynamic observations.

2.2 The Microscopic Origin of Friction Laws

Up to now, we have approached friction purely phenomenologically. At this point, the reader should be thrilled to learn about the fundamental mechanisms of friction. How come the friction force is not extensive in the surface of contact? What is the role of the load, and how come the dependence is exactly linear? What are the mechanisms for ageing, in the static and dynamical cases? Are they related? Can we find the form of the velocity-weakening law, "from scratch"?

We are only going to give a few clues about these questions, since definitive answers are not always available: even though it has progressed a lot in the last 30 years, tribology still has many challenging questions to be answered. Although we only present an overview of a sub-part of tribology, we will try to explain clearly the link between length scales, and how "elemental" objects and phenomena emerge from smaller and more fundamental ones. This simple yet rather accurate description of friction is in large part due to Archard [Arc57], with important improvements being very well summarized in [PT96, Per00].

However, we won't explore much the nano-scale aspects of friction here: for reviews on nano-scale models of friction and experimental results on nano-tribology, see [VMU+13, CBU13]. The resource letter [Kri02] contains accessible references to the relevant literature, as references are sorted and somewhat commented. Besides, in this thesis we are interested in dry friction as opposed to lubricated friction: we explain how we may dismiss lubrication in Appendix A.1.1.

2.2.1 Preliminary: What is the Atomic Origin of Friction?

Small friction forces have been observed even for contacts of very few atoms: thus, it is natural to wonder about the atomic origin of friction. At the quantum level, there is no equivalent of "friction forces" between atomic clouds. What prevents sliding at the atomic level are all the sorts of bond-formation mechanisms: chemical bonding, Hydrogen bonds, van der Waals forces, etc. At a larger level,[5] wet contacts develop capillary bridges, which are essentially liquid bonds developing due to surface tension and geometrical constraints.

In any case, the existence of bonds between surfaces in contact is an obstacle to the relative sliding of surfaces: in order to move, these bonds may first deform and at some point, break. For a bond to break, the local force has to reach a certain threshold, i.e. there is an energy barrier or activation energy needed to perform local motion. The macroscopic friction force thus emerges from these local energy barriers that have to be overcome to allow motion, so that the friction force is proportional to the number of bonds:

$$F \propto N_{\text{bonds}}. \tag{2.19}$$

The intermittent nature of bonding at a local level is sometimes seen as a sort of local stick-slip occurring at the micro or nano scale (depending on the characteristic size of the bond). However this is just an analogy: for most surfaces the local state (bound/unbound) is far from being periodic, and it is controlled mainly by surfaces' properties (and not inertia or internal stress).

Once a bond is broken, the energy is generally not recovered: in general, no new bond is formed right after breaking. Various detailed dissipation mechanisms can account for this "loss" of energy, the main ones being excitation of electrons and creation of phonons. The energy lost in these processes can be converted into mechanical energy (elastic and plastic deformations) or directly into heat. The dissipative nature of macroscopic friction originates from the irreversible part of these processes (even elastic oscillations dissipate energy via phonons).

Conclusion: Friction is Adhesion Aside from the relationship $F \propto N_{\text{bonds}}$, the main point of this very short discussion is that dry friction at the atomic scale can be reduced to *adhesion* (in the broad sense). In other words, the continuum mechanics friction force simply emerges from the adhesion properties of the particles in contact at the solid-substrate interface.

Outline

In this Sect. 2.2, we will explain the three phenomenological laws with arguments based on simple microscopic mechanisms.

[5] Note that we do not identify asperities and bonds. Bonds can be single-atomic contacts, whereas the term asperity commonly denotes micro-scale contacts. Some bonds (as wet contacts) can be of the μm length scale, as asperities. We discuss these nuances in Sect. 2.2.3.

Since friction orginates from the adhesion of atoms that are actually in contact, the geometry of each surface is crucial. We start our analysis by defining the main kinds of surface profiles in Sect. 2.2.2, in particular we define the notion of *algebraic roughness*, and provide experimental evidence of the strong roughness of most surfaces. This allows to understand naturally why the friction force is independent from the *apparent contact area*, as specified by the first law.

In Sect. 2.2.3, we discuss how the *real contact area* evolves, and how different mechanisms (elastic or plastic deformation) for its evolution all lead to a linear dependence in the load (second law). We also quickly discuss the role of fracture.

In the last Sect. 2.2.4 we show how the third law is actually violated in experiments, explain why it is almost correct at the human scale and present some hypothetical microscopical mechanisms explaining this violation.

2.2.2 Roughness

In the common sense, the *roughness* of a surface or texture is "how much the height profile deviates from the average height", and it is often taken as a binary measure: things are either smooth or rough. However, this "definition" implicitly promotes human length scales as references: for a height profile with a large spectrum of wavelengths, the human senses (tactile or visual) can only perceive variations over length scales larger than some threshold. Additionally, large wavelength variations are often considered as irrelevant for roughness "to the eye".

The concept of roughness as an objective measure of the texture properties of a surface is used in various areas of science and engineering, so that depending on the subject, its definition changes. In engineering, the variation of the profile at small enough length scales is called roughness, at larger scales it is called *waviness*, and at even larger scales it is called *form*. This is in contrast with the roughness as understood in most statistical physics works, where roughness is a measure embracing all length scales (as in *fractals*), i.e. where no particular length scale is favoured.

However, all definitions of roughness share a common goal: to reduce the tremendous amount of information contained in any given height profile $\{h(x, y),\ (x, y) \in \mathcal{D}\}$ to a few scalar variables at most—ideally just one, which would then be called "the roughness". The aim is of course to retain as much information as possible in these few variables. Depending on the symmetries expected from the profile, some definitions will be more or less fit for this purpose.

Corrugation (or the False Roughness)

The concept of *corrugation* is in the neighbourhood of roughness. In common terms, corrugation is either "the process of forming wrinkles" or "how much wrinkling there is" at the surface of something. Corrugation refers to how much some profile departs from being perfectly flat (as roughness does), but it implies the idea of periodicity

or pseudo-periodicity for the height function h. Typical examples of profiles where corrugation rather than roughness is relevant are:

- Top surface of a pack of hard spheres (e.g. glass beads), whether they are in perfect order (hexagonal lattice) or not.
- Surface of an atomically smooth substrate (e.g. mica surface): the electronic potential of the atoms forms regular bumps. The shape is essentially the same as for ordered glass beads, at a different length scale.
- Underwater sand close to the shore can form a corrugated profile with characteristic lengths of a few cm.
- Rail tracks tend to from quasi-periodic corrugations when excited at certain wavelengths. This increases the wear of tracks, because the "bumps" are extremely work-hardened, and thus *fragile*. See [Per00], p. 41.
- Fingerprints, or friction ridges, are "wrinkles" atop the fingers, which allow for a good perception of textures. See [SLPD09] or [WCDP11] for more details on the role of corrugation in tactile perception.

The crucial discrepancy between the concepts of corrugation and roughness is that the latter carries the idea of randomness, whereas the former one is usually a synonym for periodic behaviour.

In the case of the contact of two atomically flat surfaces (i.e. flat at the atomic scale, without any one-atom bump or hole), there is a small corrugation due to the crystalline lattice. If the two lattices have lattice parameters (the length of one cell of the lattice) a and b such that a/b is an irrational number, they are said to be *incommensurate*. In this case, the perfect fit of the two lattices is impossible, because locations of strong bonding due to correspondence of sites of both lattice will be rare: in this case the relative corrugation "potential" may play an important role. The locations for strong bonding will appear to be random, but are indeed determined by the relative corrugation of the two surfaces. Many friction models use this sort of corrugation to produce seemingly disordered, or random surfaces. One has to be careful with this interpretation, because this chaotic behaviour due to the incommensurate nature of substrates is "not very random". If the ratio of lattice parameters $a/b \in \mathbb{Q}$, then the two lattices are said to be *commensurate*, and then the interaction between the two will be quite strong, since the number of strong bonding sites will be extensive with the lattices size. We will discuss the case of commensurate surfaces a bit later, in Sect. 2.2.3.

Overhangs The formalism used above (and below) implicitly assumes that the surfaces we consider do not have *overhangs*, i.e. for any point $(x, y) \in \mathcal{D}$ of the surface considered the function h is uni-valued (not multi-valued). Another way to see this is to say that at any point, the local angle between the surface and the base-plane is less than or equal to $\pi/2$. In case a surface actually has overhangs, many detection apparatus would measure a "regularized" surface (as shown in panel c and d of Fig. 2.7).

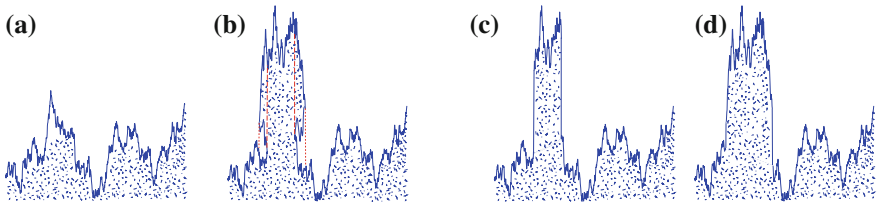

Fig. 2.7 Various height profiles $h(x)$. The *solid part* is pictured by *small dots*. **a** "Normal profile", without overhangs. **b** Profile with one overhang. Two regularisations are suggested by *dashed* and *dotted red lines*. **c** A first regularization of profile (**b**), as suggested by the *dashed lines*. **d** A second regularization of profile (b), as suggested by the *dotted lines*

Width Described by a Single Scale: The Finite Roughness

For essentially "flat" profiles or more generally in engineering applications (where only a certain range of length scales are relevant for friction), one may resort to simple measures of the height profile $h(x, y)$ in terms of its first moments or of some extremal values. The underlying assumption is that the variations of h are "finite", i.e. the moments of the distribution $h(x, y)$ (or even its cumulants) are finite, i.e.[6] $h \in \mathbf{L}^2(\mathbb{R}^2)$. We will see later how well this condition should be fulfilled for this sort of measures to be accurate.

Let us now precisely define a few measures of roughness. Consider a finite (but macroscopic) sample, defined by the domain $\mathcal{D} \subset \mathbb{R}^2$. Suppose that the raw profile h is sufficiently regular: $h \in \mathbf{L}^2(\mathcal{D})$. To extract relevant variations of the height profile, we will generally subtract its average to h. We use $\overline{X}$ to denote the space average of any quantity X: $\overline{h} \equiv \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} h(x, y)\mathrm{d}x\mathrm{d}y$. The most common measures of roughness are given by the following functions of h.

- The (average of the) absolute value: $R_a[h] \equiv \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \left|h(x, y) - \overline{h}\right| \mathrm{d}x\mathrm{d}y$.
- The root mean squared R_{RMS} or width: $w[h] = \sqrt{\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \left|h(x, y) - \overline{h}\right|^2 \mathrm{d}x\mathrm{d}y}$.
- The maximum height of the profile: $R_t[h] = \max_{(x,y)\in\mathcal{D}}(h) - \min_{(x,y)\in\mathcal{D}}(h)$.

Additional measures of the properties of a surface are e.g. the skewness and the kurtosis of the profile, which come naturally as higher moments of the height function, seen as a probability distribution.

Relevance These kind of measures—taken as simple real values—are well fit for engineering applications, where the roughness needs only to be assessed on a definite range of length scales, and for which the variations are usually mild in this range. In the case of small variations, the observables defined above are well-behaved, in particular they are essentially independent of the sample size. However, in the more general context of the physics of friction, these measures fail to account for the rich behaviour of the surfaces we may be interested in, and more specifically, they can

[6] This notation indicates that the function h is a square-integrable function on $\mathbb{R}^2$: $\int_{\mathbb{R}^2} |h|^2 < \infty$.

strongly depend on the sampling size. Instead of looking at these functionals of h as simple real variables, it is preferable to consider them as functions of the sampling length, and to extract a few relevant quantities from these functions.

In particular, in the case of numerous natural surfaces, these indicators would explode: the root mean square or width measurement for instance, w, would essentially diverge, if the distribution h were to increase as a power-law. We are about to see that this is indeed the case, at least in the applications we have in mind.

Self-Affinity: The Algebraic Roughness

As can be observed for silicon nitride ceramic balls observed at the micrometer scale (see Fig. 2.8) the height profile of rather smooth objects can actually be quite irregular. We give a view of a rough surface from a toy model in Fig. 2.8 (central and right panels). This toy profile has large relative variations over a large range of length scales. Here we want to provide the tools for describing such kind of profiles. Defining new tools will also allow us to characterize more precisely experimental observations.

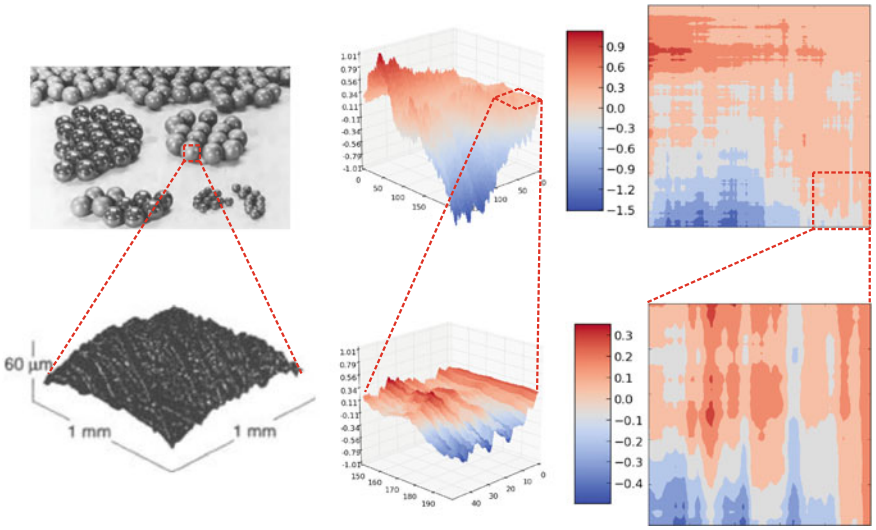

Fig. 2.8 *Left* silicon nitride balls (used for bearings), finished (very smooth) and "rough lapped" (rougher). We zoom ($\sim\times 100$) on one of the rougher balls (*below*), and realize that the landscape is much rougher than it seemed, using a height resolution $\sim 10\,\mu$m (*Images retrieved from* [Per00], *originally from* [Cun93]. *Central* (respectively *right*) *panel*: 3D view (resp. "heat map" colouring) of the height profile for a toy model of surface (arbitrary units). We zoom ($\sim\times 3$) on a seemingly flat section, which reveals a rather irregular microscopic landscape upon closer inspection (*below*), similar to the large scale one. Note that the preferred directions of our toy-surface (present at various scales) are an artefact of the generating procedure, they are not expected to be so strong for real materials

First, we want to give clear definitions of the mathematical terms used, then see a few examples of surfaces that can be characterised using these definitions, and finally explain how we can quantitatively describe these surfaces efficiently, which will yield a natural definition of the (algebraic) roughness.

Self-Similarity (and related definitions) Numerous objects have the property that they "look the same" at various length scales. Here we make this idea more precise by defining a few mathematical properties related to this idea. Additional details are available in Appendix A.1.2.

Let us first define the property of self-similarity. A function of two variables $g(x, y)$ is said to be self-similar if an only if (iff) it satisfies:

$$g(x, y) = \Lambda_1 \Lambda_2 g(\Lambda_1^{-1} x, \Lambda_2^{-1} y), \qquad \forall \Lambda_{1,2} > 0, \ \forall (x, y). \tag{2.20}$$

This is a *re-scaling*, and it correspond intuitively (e.g. for $\Lambda > 1$) to do two things at the same time: "zoom out" in the x- and y-directions and to magnify (or also "zoom in") in the g-direction. Self-similarity is a very stringent constraint, since the re-scaling in different directions has to be exactly the same.

A more general property defining objects with "similar" appearance at different length scales is *self-affinity*. A function of two variables $g(x, y)$ is said to be self-affine iff:

$$g(x, y) = \Lambda_1^{b_1} \Lambda_2^{b_2} g(\Lambda_1^{-1} x, \Lambda_2^{-1} y), \qquad \forall \Lambda_{1,2} > 0, \ \forall (x, y), \tag{2.21}$$

where b_1, b_2 are the self-affinity or scaling exponents related to the affine transformation. This may be referred to as "anisotropic" self-affinity, but this wording is misleading, because even for $b_1 = b_2 \neq 1$, we already have an affine transformation (and not a similarity transformation).[7] We see that self-affinity is an anisotropic transformation which contains self-similarity as a special case ($b_1 = b_2 = 1$).

Self-affinity is a rather general property, however it is interesting to note that it only allows to compare fully deterministic objects. If we are interested in a random process, we need an additional definition: *statistical self-affinity*. This is especially relevant to characterize a real surface (which is highly heterogeneous, i.e. random). A surface profile is said to have a roughness exponent ζ when it is statistically self-affine, i.e. when:

$$g(x) \overset{\text{Law}}{=} \Lambda^{\zeta} g(\Lambda^{-1} x), \qquad \forall \Lambda > 0, \ \forall x, \tag{2.22}$$

where the equality is "in Law" (for the random variables as distributions, not realization per realization).

[7] Please note that in part of the literature, these two concepts are sometimes mistaken for one another, or simply melted and seen as equivalent. When considering functions, it seems quite natural that the ordinate and abscissa do not share the same scaling exponent, so that considering self-affinity seems very natural. However, when considering geometrical objects such as self-similar or self-affine objects, the distinction becomes important. Not all *fractals* are *self-similar fractals*.

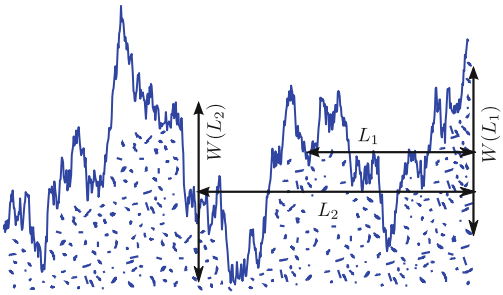

Fig. 2.9 Illustration of the width and its dependence on the sample length L. Depending on the definition of the width (or "roughness"), the precise value of $w(L)$ will defer. However, for an algebraically rough surface, all definitions will display a roughness exponent ζ such that $w(L) \sim L^{z}eta$

A generic example of mathematically well-defined stochastic process which is statistically self-affine is the *fractional Brownian motion* (fBm). To give the interested reader more insight into statistcal self-affinity, we study the fBm in Appendix A.1.3.

Structure Factor

To describe height profiles with the statistical self-affinity property, one needs to extend the tools previously introduced. For instance, the root mean squared w ("width") of the height profile $h(x)$ is the square root of the second moment of the distribution computed in Eq. (A.9). For a surface being statistically self-affine at least over the range $x \in [0, L]$ with a roughness exponent ζ, we thus have a width $w[h, L] = L^{\zeta}$ (see Fig. 2.9 for a concrete illustration). This is obviously a problem, since an observable that explicitly (and much strongly) depends on the sampling size is clearly ill-defined.

The solution is to acknowledge the self-affine nature of the surface, and to use the exponent ζ to define the roughness, which is possible since

$$\zeta \underset{L \gg 1}{\sim} \frac{\ln(w[h, L])}{\ln(L)} \tag{2.23}$$

does not depend on the precise value of L, as long as $L \gg 1$. However, it is important to note that not all rough surfaces are exactly statistically self-affine with a unique exponent over all length scales. There are usually cutoffs (lower and upper) to the self-affine behaviour, and the exponent may even have two distinct values over two distinct ranges! Thus, in order to be valid for a wider class of rough profiles, this definition of roughness needs to be extended.

A very general observable that helps measuring the roughness of a given height profile is the *structure*[8] *factor* $S(q)$. This is not a roughness, since it is not a scalar, but a function (which inherently contains more information than a single scalar). The

[8]Originally, the concept was used in crystallography, where *structure* obviously refers to the crystalline structure. The idea of looking at the spectrum in Fourier space, and at the typical energy of each mode has since spread in many disciplines.

idea is simply to look at the energy associated to each mode in the spectrum of the height distribution. For a d-dimensional profile $h(\mathbf{x})$, assuming periodic boundary conditions (for simplicity) in a system of lateral length L, the averaged structure factor is defined as:

$$S(\mathbf{q}) \equiv \overline{\frac{1}{|\mathcal{D}|}\left|\int_{\mathcal{D}} \mathrm{d}^d\mathbf{x}\, h(\mathbf{x})\, e^{-i\mathbf{q}\mathbf{x}}\right|^2} \tag{2.24}$$

$$= \int_{\mathcal{D}} \mathrm{d}^d\mathbf{x}\, \overline{h(\mathbf{x})h(\mathbf{0})}\, e^{-i\mathbf{q}\mathbf{x}} \tag{2.25}$$

where $\mathbf{x}$ is the d-dimensional coordinate, $\mathcal{D}$ is the domain considered and where translational and rotational invariance ensure that the (spatial) frequency $S(\mathbf{q})$ only depends on $q = |\mathbf{q}|$, via $q = 2\pi n/L$, $n \in \mathbb{N}$. The average $\overline{X}$ is the average of X over many samples. For any self-affine process with exponent $b = \zeta$, we have $h(x) \sim x^\zeta$ up to a random phase so that we get:

$$S(q) \sim q^{-(d+2\zeta)}, \tag{2.26}$$

so that aside from finite size effects (at short and large wavelengths), it is a pure power-law (see e.g. [KRGK09]). The measure of the structure factor is a robust way to estimate roughness. A nice feature of $S(q)$ is that if the profile considered is actually not self-affine, or if it has two regimes with different exponents of self-affinity, it can be seen immediately, as for example in Fig. 2.11.

From now on, we will be interested solely in this last sort of roughness, so that "rough" will refer to statistically self-affine surfaces, and ζ may be called the roughness. Except when explicitly stated otherwise, the surfaces we will consider are rough over a large range of length scales.

We will discuss examples of rough interfaces produced by theoretical models in later sections. For an example of concrete use of the structure factor and some precise results on the roughness of a one-dimensional elastic line in disordered medium, see [FBK13].

Experimental Examples of Rough Surfaces

Now that we have defined the appropriate tools, we can discuss real observations more seriously than with Fig. 2.8. In Fig. 2.10, the roughness of some surfaces of brittle materials (close to some cracks) is observed. If Fig. 2.11, the roughness of two-dimensional surfaces is measured for various materials, and we see how the structure factor can help to determine to what extent a surface is really self-affine. From these examples of self-affine surfaces, we begin to understand why the friction force is independent from the apparent contact area: since most surfaces are very rough, they can touch each other only at few points. If friction truly happens only

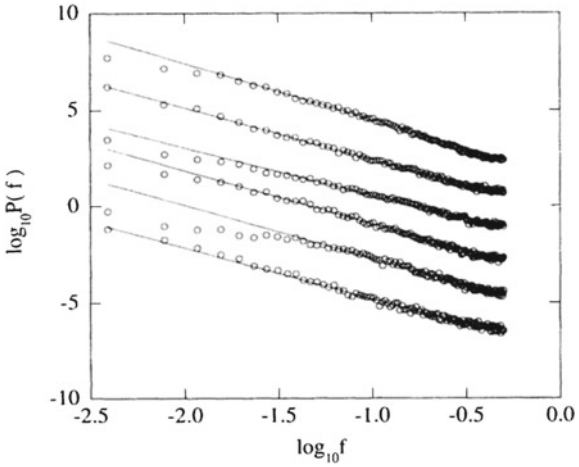

Fig. 2.10 From [MlyHHR92]. Roughness of surfaces of six different brittle materials, close to the fracture area (*crack*). Measurement of the height profile along one-dimensional cuts in the direction perpendicular to the crack. The "power spectrum" $P(f)$ of the profile is exactly what we defined as the structure factor $S(q)$. The log-log plot shows the dependence of $P(f)$ in the wavelength or space frequency f. The roughness ζ is extracted from the fit $P(f) \sim f^{-(1+2\zeta)}$

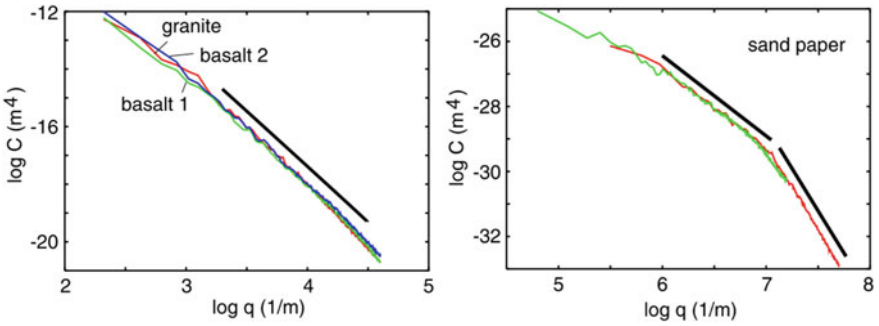

Fig. 2.11 From a recent and excellent review on roughness, namely [PAT+05] (©IOP Publishing. Reproduced by permission of IOP Publishing. All rights reserved). Optical measures (*left panel* and *green curve* of *right panel*) are combined with AFM (Atomic Force microscope) measurements (*red curve* of the *right panel*). The correlation function can be identified with the two-dimensional structure factor, here denoted $C(q)$. A fit is done to evaluate the fractal dimension, which is found to be $D \approx 2$ for basalt and granit (*left panel*) and $D \approx 2.2$ for sandpaper at $\log q < 7$ (*right panel*). This corresponds (for these 2D surfaces) to roughness given by $\zeta = 3 - D$. Notice how there are two regimes for sandpaper, which are easily identified thanks to the use of the structure factor

where the surfaces meet, it must be proportional only to this real contact area, which we now expect to be much smaller than the apparent one.We will explain this clearly in the following section.

2.2.3 Real Contact Area

We have seen that the apparent contact area has probably little to do with the real one, and that only the latter is involved in friction. Here we want to compute this real contact area from macroscopic measurements.

Most people have the idea that "smooth surfaces slide better". So, let's imagine the extreme case of two perfectly flat, clean and commensurate surfaces. What would happen if we were to put them in contact, and then apply some shear? The answer is that we would simply observe cold welding, i.e. the boundary atoms would form bonds between the two surfaces. Bonds could be chemical, or just van der Waals forces.[9] If at least one of the materials has some impurities, the shear stress necessary to obtain some strain (deformation) would be essentially the yielding stress of the weaker of the two materials, and the shear would occur in the bulk of it, instead of occurring in the contact plane. This simplistic example illustrates how friction would be incredibly huge, if contact was to truly occur on the complete *apparent area of contact*. Notice that in this ideal case, "friction" would be proportional to the apparent contact area. From now on when we discuss the contact area, it will be implicitly assumed that we do *not* refer to this apparent area of contact.

Stepping back a little from this very extreme example, if a surface is flat except from few *asperities*[10] of approximately the same height, one may expect that the very few "true" contact points will allow for very low friction. However, imagine this surface is slowly driven down towards another one with similar design (or completely flat). As soon as the macroscopic load would be a bit more than zero, the local pressure at the asperities would quickly become enormous, since it goes as the inverse of total (true) area of contact. This would result on the plastic yielding of asperities, i.e. in irreversible deformations at the atomic level, instead of reversible elastic deformation. The "peaks" would be crushed, flattened, so that in the end we would have the flat solids separated by few spots of one-layer flattened asperities, resulting once again in a large contact area. Furthermore, if the distance between the two flat solids is indeed of only one atomic diameter, the van der Waals interactions might once again play some role by further increasing the macroscopic adhesion force.

Thus, we see that very smooth—nearly atomically smooth—surfaces, contrary to popular belief, do not slide well. Another common idea is that very rough surfaces slide badly. Actually, this one is true: for a surface with macroscopic height oscillations, i.e. "macroscopic corrugation" or *form* (or *waviness*), the energy barriers that one needs to overcome to slide through are so high that they prevent any easy sliding. Even if the microscopical properties of the solids are such that the microscopic friction coefficient is small, for corrugated profiles, the surfaces will be interlocked

[9] The relevance of van der Waals forces at the *nano*scale has been questioned recently in [MTS09]: "friction is controlled by the short-range (chemical) interactions even in the presence of dispersive [van der Waals] forces".

[10] Asperities, *contacts* or *junctions* are all words that designate the small "bumps" at the top of any surface, which are responsible for the *true* contact between solid and substrate. For a rough surface, they are the top "peaks" of the profile.

with one another, and the macroscopic friction force will be high. This is the case for "roughcast" (or for "pebbledash"): even with a good microscopical surface treatment, two such surfaces rubbed against each other would still slide very badly. In this sense, the engineering definitions of the waviness and form are appropriate to eliminate the large length scales contributions to friction, which can involve mechanisms other than "small scale" friction.

Asperities at the Microscale

As it has been mentioned earlier, asperities are the small "bumps" on top of a surface which are responsible for the *true* contact between solid and substrate. By definition, a *contact* is the point where the two surfaces meet and where bonds can form. The concept of *junction* involves the idea of welding, which is made easier by the high pressures at the asperities. For a rough surface, asperities are typically the top "peaks" of the profile.

It is important to notice at this stage that bonds and asperities are not the same thing. On the one hand, the notion of *bond* covers length scales from the atomic size (a few Ångströms, $\sim 10^{-10}$ m) to capillary bridges (up to fractions of mm, $\sim 10^{-4}$ m). A bond is an elementary unit: it can get weaker or stronger due to external conditions, it can break, but it does not have relevant sub-elements. On the other hand, the notion of asperity refers to an entity generally described by continuum mechanics: the contact between two asperities is of a size such that in the range of loading conditions studied, it can not merge with a neighbouring one. Typically, the radius of the contact area of an asperity is $\sim 10\,\mu$m.

On a first approach, asperities can be seen as the building blocks of the contacts responsible for friction. Then, the *true contact area* or *asperity contact area*[11] can be considered to be the whole area of contact between asperities, as depicted in Fig. 2.12a. A refined approach consists in considering the inner dynamics of the contact. Then, the real contact area or atomic contact area is just the sum of the individual contact area of each atomic bond (See Fig. 2.12b) The difference between these two approaches has been pointed out in [MTS09], and opens promising avenues for a better understanding of friction, especially for nanoscale objects.

However, the notion of asperity is often not only sufficient, but more relevant than that of bond, for several reasons. First, the fact that the real contact area is not equal to the apparent asperity area is not truly an issue, since in calculations it is (often) automatically the real contact area which is involved. Second, asperities are the (pseudo) elementary blocks which pin the surfaces together: their scale appears as a natural length scale in many aspects of friction, and is way more practical to handle than the atomic scale. Consequently, it is often sufficient to study their dynamical behaviour alone (elastic and plastic deformations). Third, asperities are large enough that one can apply most continuum mechanics to them: this is very handy. Hence, we will mainly discuss the behaviour and dynamics of asperities in what follows. For a

[11] What we call asperity contact area used to be consider the true contact area.

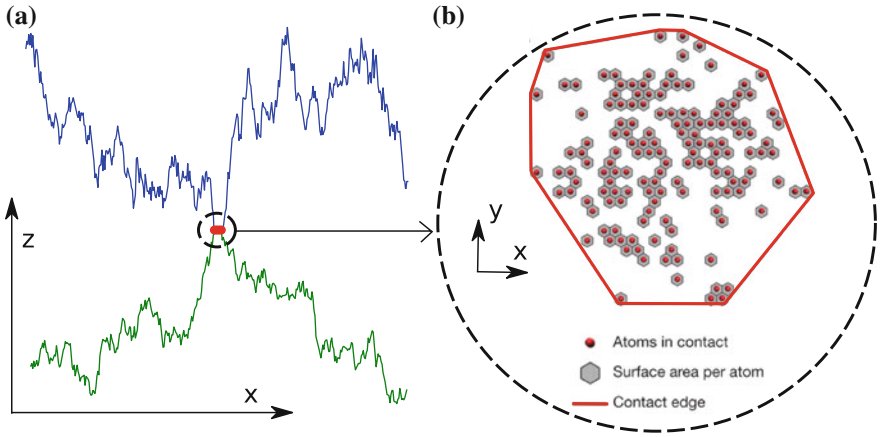

Fig. 2.12 *Left* two profiles with algebraic roughness ($\zeta = 0.5$) enter in contact. The junction is highlighted in *red*. *Right* schematic view (from [MTS09]) of the junction, from *above*. Over the area of the junction (the "real contact area"), not all space is actually covered in bonds. The atomic bonds (*red dots*) actually cover only the *grey area*. From outside, the contact area is naturally mistaken for the contact edge (*solid red line*), i.e. for the convex hull enclosing all the atomic bonds. In most studies, the "real" or "true" contact area implicitly refers to this convex hull, not to the *grey area*

review on nanoscale models of friction and experimental results on nano-tribology, see [VMU+13], or the resource letters [Kri02] which contains accessible references to the literature.

Role of Plastic Yielding at the Solid-Substrate Interface

Consider a substrate upon which we set an object of which the lower surface is rough in the sense defined earlier (i.e. it has a statistically self-affine surface). As we approach the solid[12] from above, at first there is only a single asperity in contact. At this asperity, the pressure p_1 over the (real) contact area A is given by $p_1 = L/A$, where L is the *macroscopic* load. For a typical asperity of diameter $a \sim 10\,\mu\text{m}$, we have an asperity area $A \approx 10^{-10}\,\text{m}^2$. For a load given by the weight of 1 kg, $L \approx 10\,\text{N}$, so that $p_1 \approx 100 \times 10^9\,\text{N/m}^2$. For reference, the yield stress[13] for diamond is $\sim 80 \times 10^9\,\text{N/m}^2$, and for steel it is between 1 and $7 \times 10^9\,\text{N/m}^2$ (it depends on the quality of the steel). As the pressure in the contact area is larger than

[12] At this point, it does not matter to know precisely the profile of the substrate: whether it is flat or rough with the same exponent as the upper solid, we can subtract the two profiles and consider the result as the effective profile for the solid, and consider the effective profile of the substrate to be flat.

[13] The yield stress is the stress that one needs to apply in order to obtain plastic yield. In the context of these estimations, the relevant quantity is the *penetration hardness* or *indentation hardness*. The typical measure protocol is that of Vickers: on the sample, an indentation is performed with a tetrahedron in diamond. The stress needed to perform the indent is the indentation hardness.

the yield stress, this single asperity must yield plastically, i.e. it is smoothly crushed by the upper solid.

As the upper solid goes further down, it will encounter other asperities, which will increase the contact area. As long as the pressure remains larger than the yield stress, the solid will deform plastically. When the contact area is large enough to strike a balance between pressure at asperities and yield stress, plastic deformation will stop. This gives us a natural formula for the real contact area:

$$A_{\text{real}} = \frac{L}{\sigma_c}, \tag{2.27}$$

where σ_c is the yield stress or indentation hardness of the softer of the two materials. To be concrete, let's continue with our mass of 1 kg, on top of a table of the same steel (or any other stronger material). Let's assume it is made of steel with $\sigma_c = 10^9\,\text{N/m}^2$. The real contact area is then $A_{\text{real}} = 10^{-8}\,\text{m}^2$, which is completely independent from the apparent contact area. Surprisingly, this corresponds to only $\approx$100 asperities of unitary area $\sim 10^{-10}\,\text{m}^2$. We may compare this contact area with the apparent one A_{app} by assuming the steel to be shaped as a parallelepiped, for instance with the dimensions 10 cm $\times$ 10 cm $\times$ 1 cm (the density of steel is $\rho \approx 10\,\text{g/cm}^3$). In this case we have $A_{\text{real}} = 10^{-4}\,\text{cm}^2 \ll A_{\text{app}} = 100\,\text{cm}^2$, or also $A_{\text{real}}/A_{\text{app}} = 10^{-6}$, i.e. the real contact area is only a tiny fraction of the apparent one. See Fig. 2.13 for an illustration.

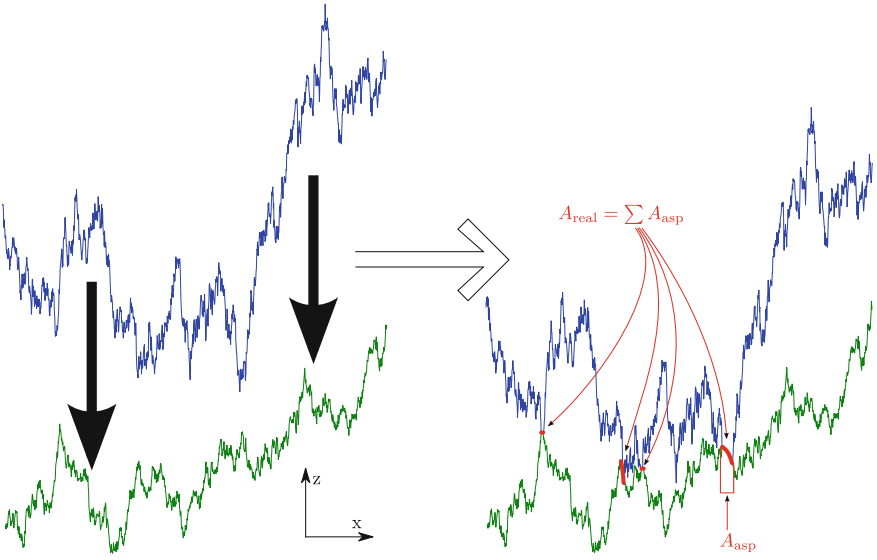

Fig. 2.13 Schematic description of two rough surfaces (*left*, $\zeta = 0.5$) squeezed together. They make more and more contacts (*right*, highlighted in *red*) until $A_{\text{real}} = L/\sigma_c$. The area of each asperity, A_{asp}, is the projection (*dotted red line*) of the contact onto the (x, y) plane. This area is much smaller than the total area

When we slide a solid over a "fresh" area, or when the frictional wear changes the asperity landscape, and more generally as soon as some surfaces meet for the first time, the picture presented above will also be valid. As we have seen earlier (Eq. 2.19), $F \propto N_{\text{bonds}}$. The number of bonds is essentially proportional to the area of real contact, so that in the end, $F \propto N_{\text{bonds}} \propto A_{\text{real}} \propto L/\sigma_c$, i.e. we found Amonton's second law.

In the above cases, we have assumed that the elastic deformations of the materials are negligible. This is perfectly correct as long as we start from a state with few contacts: the pressure is so high that local strain is large, and most of the deformation is plastic. Another way to put it is to say that contacts are in a state of incipient plastic flow, i.e. that they are at their plasticity threshold (or way beyond). When we are around the equilibrium state with $A_{\text{real}} = L/\sigma_c$, however, elastic deformations can become relevant.

Role of Elastic Deformation at the Solid-Substrate Interface

In several cases, it is elasticity rather than plasticity which controls the evolution of the surface area. In the friction of rubber, the very low elastic modulus makes it very difficult to plastically deform the rubber, so that elastic forces prevail.[14] For a surface that is very smooth, in the sense there are many asperities at the top with approximately equal height, one may expect the real contact area to be larger than what is expected from the plastic yield reasoning.

Another natural question is to ask what happens in the following "extra-load" experiment. In the "extra-load" setup, we set our steel block onto a (hard and flat) table (the load is $L = L_1$), then press it with an extra load of $1\,\text{kg} \Rightarrow L_1 = 10\,\text{N}$ (i.e. we double the total load), then remove the extra load ($L = L_1$). According to our reasoning, the asperities have been crushed to a point where $A_{\text{real}} = 2\,L_1/\sigma_c$, so that we would naively expect the real contact area to be double of what is expected from the simple, current load L_1 (such an effect of memory of the previous loading is actually not observed, not to this extent at least). We have just produced a "very smooth" surface as mentioned above, since the top asperities have all the same height.

In all these cases the naive analysis implies that the real area of contact is no longer proportional to the load, i.e. that Amonton's second law is violated. However, in all these cases the stress in the asperities can be quite high, since it is only bounded from above by the plastic yielding limit σ_c. With values of the local stress up to σ_c, the elastic deformations of contacts can and will play an important role. To compute the real area of contact and in particular its dependence on the load, we will need to consider the elastic deformations of asperities. In this application of linear elasticity theory, we will consider adhesion forces negligible compared to the elastic tensile stresses (even though it is precisely adhesion which is responsible for friction!). The

[14] See [Per01] for a study of this extreme case that is rubber friction. Be careful that the theory has evolved since, in particular one should consult [PAT+05] for accurate results.

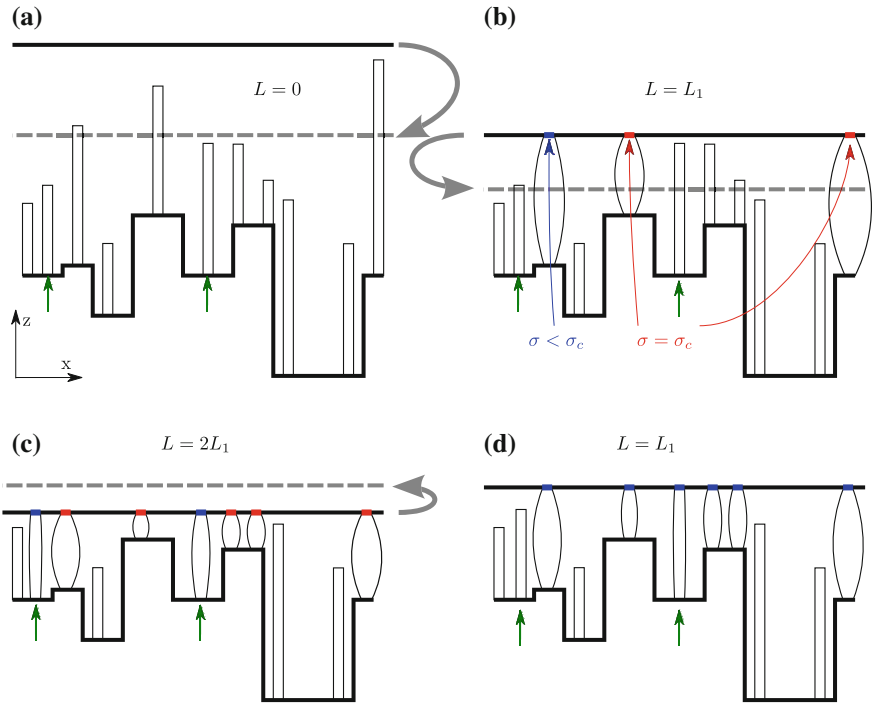

Fig. 2.14 Schematic description of the "extra-load" (thought) experiment we propose. The *upper solid* is considered infinitely tough compared to the lower one ($\sigma_c^{up} \gg \sigma_c^{low}$). Two particularly interesting asperities are highlighted by *green arrows* along the evolution. **a** Load is zero, all asperities are intact. **b** Load is L_1, some asperities deform elastically (*blue*, $\sigma < \sigma_c$), others also yield plastically (*red*, $\sigma = \sigma_c$). **c** Load is increased to $2L_1$: additional elastic and plastic deformations occur. **d** Load is decreased back to L_1: the *upper solid* is not pushed back to its initial position. Asperities that were subject to very high stress can release a lot of it by pushing the *upper solid* up. Asperities that were subject to moderate stresses go back to their original shape (*left green arrow*), or are only slightly compressed (*right green arrow*)

role of adhesion for elastic solids with rough (random) surfaces has been included in recent works as [PSS+08], where the law $F \propto L$ is still predicted.

We now discuss the elastic response of two simple models of asperities: cylindrical asperities of which the extremity is considered flat, and spherically ended asperities (where the asperities are not elongated enough to be able to neglect the shape of the asperity extremity).

Model I: Cylindrical Asperities Here I give a schematic description of what happens in the "extra-load" experiment by considering the asperities as essentially cylindrical. In this limit, the contact area at each asperity is either 0 (no contact) or $A_{\text{asp},1}$ (typical area of one micro-scale asperity).

When we increase the load up to $2L_1$, asperities are crushed so that $A_{\text{real}} \approx 2L_1/\sigma_c$ (See Fig. 2.14c). When we then decrease the load to $L = L_1$, the pressure no longer

overcomes the yield stress, so that plastic flow is no longer possible. Yet, there is still some high stress concentration in the asperities: instead of having the compressive stress of the bulk, $\sigma_{zz}^{\text{bulk}} \approx L/A_{\text{app}}$, asperities are subject to a compressive stress $\sigma_{zz}^{\text{asp}} \in [0, \sigma_c]$, with an average:

$$\langle \sigma_{zz}^{\text{asp}} \rangle \equiv p_1 = \frac{L}{A_{\text{real}}} \gg \sigma_{zz}^{\text{bulk}}. \tag{2.28}$$

This stress corresponds to a compression of each asperity along z by a compressional strain ε (dimensionless variable) initially given by:

$$\varepsilon = \frac{d}{z_0} \propto E\sigma_{zz}^{\text{asp}}, \tag{2.29}$$

where d is the elastic displacement of the asperity, z_0 its initial length,[15] E the Young's modulus of the material. Depending on its initial length z_0 (length after plastic flow), each asperity is more or less compressed, as pictured in Fig. 2.14c.

Qualitatively, when the asperities are relieved from the extra load, those which were more compressed (larger d/z_0) are also those which de-compress more: they rise, thus "lifting" the solid upwards. Those which were less compressed (smaller d/z_0) "rise" less and can thus lose contact in the process. (See Fig. 2.14d).

We denote δd the "rise" of each asperity, so that the total lift of the upper solid is equal to $\langle \delta d \rangle_s$, where the average is over the surviving[16] contacts at the end of the process. As the rise of each surviving asperity is automatically equal to the macroscopic one, we have also $\delta d = \langle \delta d \rangle_s$ (only z_0 is a random variable, drawn independently for each asperity). To give a rough estimation of the dependence of the real contact area in the load, we make the assumption that the total "rise" of the asperities is negligible, i.e. that $\delta d \ll \langle z_0 \rangle_s$. Thus for a surviving asperity the local change $\delta\sigma_{zz}^{\text{asp}} \propto \delta d/\langle z_0 \rangle_s$ in compressive stress is negligible: $\sigma_{zz}^{\text{asp}} \propto d/z_0 \approx \text{const.}$ For these asperities we have an average compressive stress $p_1 = \langle \sigma_{zz}^{\text{asp}} \rangle_s \approx \text{const.}$, i.e. $L = p_1 A_{\text{real}} \propto A_{\text{real}}$, i.e. Amonton's second law is respected. We may notice that since the asperities are cylindrical, the unitary contact area $A_{\text{asp},1}$ is constant, and we have the more precise relation $L \propto N_{\text{asp}}$. We note that the approximation $\delta d \ll \langle z_0 \rangle_s$ is especially well respected for very rough profiles, where z_0 has a large distribution. This is clear if we consider the asperities which lose contact: if their z_0 is very large, a smaller rise δd will be enough to kill contact.

---

[15] For now, we assume elongated asperities in the z direction, in the sense that their contact does not depend on compression. Examples of such ideal shapes are cylinders or parallelepipeds, that can be modelled by a simple spring. Examples of cases we exclude with this assumption are the spherical and cylinder-with-rounded-tips shapes.

[16] For the asperities which lose contact (or "die"), the variation of d is even smaller than for the surviving ones (it is less than the rise of the upper solid). However, this smaller rise corresponds to a drop of the compressive stress from σ_{zz}^{asp} to zero, since contact is lost. This explains how some load bearing can be "forgotten", despite the surviving contacts being subject to an approximately constant pressure.

The conclusion is that the main effect of decreasing (resp. increasing) the load in the elastic regime is to remove some contacts (resp. create new ones).

Model II: Spherically Ended Asperities Another way to consider asperities is assimilate them as spherical bumps, as depicted in Fig. 2.15. Let's start with a single contact. Assuming purely elastic deformation and no adhesion, the Hertzian theory of contact mechanics predicts, for a sphere pressed into a half-space, a non linear dependence of the contact area with the load: $A_1 \propto L^{2/3}$. The non linearity may seem surprising, given that we only used linear elasticity theory.

The qualitative explanation is very simple: as loading increases, the contact area increases from a point to a disk of increasing radius. The average pressure in the contact area is the macroscopic load divided by the contact area: it starts very large, which allows for a large indent depth d, but as indent increases, so does the contact area, which reduces the local pressure. At the end of the day, even though the indentation is always proportional to local pressure, the geometry is such that the overall dynamic is non linear in the load L.

However at the macroscopic scale, the linear dependence is most commonly observed. A linear dependence emerging from the non linear law is found in a simple model of spherical asperities. There is a classical derivation of the area of contact and load for this model due to Greenwood, nowadays available in Chap. 2 of [PT96]. We reproduce here the main line of Greenwood's argumentation.

Consider the centres of the spherically shaped bumps (of radius R) as depicted in Fig. 2.15: the centres' heights constitute a profile $\Phi(z)$ ($z = \mathcal{H}$ being the height of the flat plane onto which asperities are pressed). At each bump, applying Hertz theory, the bump is compressed a distance $d = z - \mathcal{H}$, leading to an (unitary) contact area $A_1 = \pi R d$ and a load (borne by this single asperity) $L_1 = (4/3)E^* R^{1/2} d^{1/2}$,

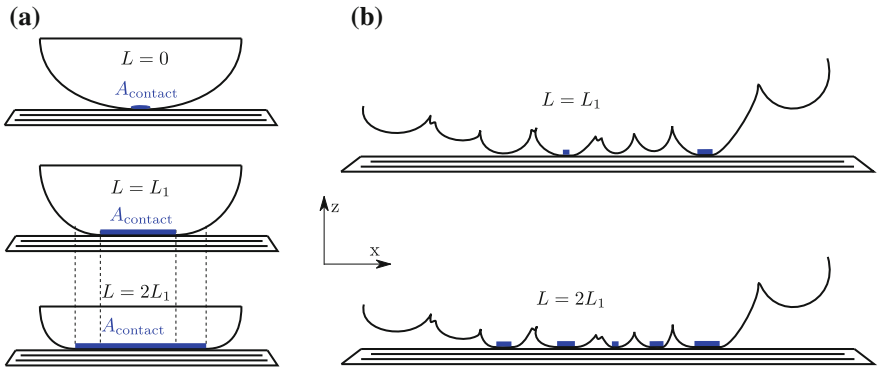

Fig. 2.15 Schematic description of spherically shaped asperities or "bumps", in purely elastic compression (*not* to scale). **a** A single bump is compressed onto a rigid substrate. At zero load, the area of contact is a single point. At higher loads, the area of contact is elastically deformed (*blue highlight*) and is not proportional to the load: $A_{\text{contact}} \propto L^{2/3}$. **b** A surface is modelled by spherical bumps. On average, the area of contact is proportional to the load

where E^* is the reduced Young's modulus.[17] With N being the number of bumps in the sample, we have the number of contacts n, area and load given by:

$$n = N \int_{\mathcal{H}}^{\infty} \Phi(z)\mathrm{d}z \tag{2.30}$$

$$A_{\text{real}} = N\pi R \int_{\mathcal{H}}^{\infty} \Phi(z)(z-\mathcal{H})\mathrm{d}z \tag{2.31}$$

$$L = (4/3)E^* R^{1/2} \int_{\mathcal{H}}^{\infty} \Phi(z)(z-\mathcal{H})^{3/2}\mathrm{d}z \tag{2.32}$$

We assume a rapid decay of the function $\Phi(z)$, which seems reasonable for "flat" solids. We can take any decay, e.g. Φ can be Gaussian or more simply, $\Phi(z) \simeq e^{-\lambda z}$ is fine. This allows to compute:

$$n = \frac{N}{\lambda} e^{-\lambda\mathcal{H}} \tag{2.33}$$

$$A_{\text{real}} = \frac{N\pi R}{\lambda^2} e^{-\lambda\mathcal{H}} \tag{2.34}$$

$$L = \frac{E^* R^{1/2}\pi^{1/2}}{\lambda^{5/2}} e^{-\lambda\mathcal{H}}, \tag{2.35}$$

from which Greenwood concludes that $A \propto L$. The problem with this reasoning is that it assumes a rapidly decaying profile Φ: for very flat surfaces, it is well acceptable. However, for surfaces with roughness at several length scales, the relevance of this model has been questioned, for instance in [PSS+08]. Indeed, this model only accounts for roughness at a single length scale: the elastic deformation of larger regions (e.g. made of several bumps) is implicitly considered to be zero, because this larger length scale is implicitly ignored.

The Hertzian theory of contact applied on spherical asperities has played an important historical role, and is still valid for "Gaussian" or flat surfaces. This non linearity in the response of spherical contacts is also interesting for the study of granular materials. In regimes where the "balls" merely touch each other, it can be crucial to account for the non linear response (see for example [GTvHV12] or the review [AT06] for more details).

There has also been some observations of sub-linear dependence of the friction in the load, in some particular contexts [BGK06, Per00]. This kind of non linear dependence at the macroscopic scale is typically obtained when the unitary contact area depends on the local load, i.e. in all sorts of rounded or triangular shapes, but also when additional forces (e.g. van der Waals or capillary forces) produce geometrical arrangements which strongly depend on the load, at rather large length scales. The idea of a single-asperity with rounded extremity is also sometimes used as a

[17] The reduced Young's modulus is defined as a combination of the two materials Young moduli E_1, E_2 and their poisson ratio ν_1, ν_2 via: $1/E^* = (1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2$.

rudimentary model of AFM tip, in this case the single-asperity tip is the macroscopic system. The relevance of this sort of behaviour was summarized very early by Archard, whom asserted that [Arc57]:

> If the primary result of increasing the load is to cause existing contact areas to grow, then the area of real contact will not be proportional to the load. But if the primary result is to form new areas of contact, then the area and load will be proportional.

The Question of Fracture

The breaking of junctions is fundamentally a fracture process: as we have said earlier, at the asperity scale, the high pressures result in cold welding, so that the separation of the two surfaces occurs through rupture. The Fineberg's group recently developed a real-time visualization method of the real area of contact during the sliding of the blocks [RCF04, RCF07, BDRF10, SF14], see also the review [VMU+13]. This method shows that the transition from static to kinetic friction is controlled by the collective behavior (and fracture) of the ensemble of asperities that form the interface between the two solids. In particular they identify three different kinds of coherent crack-like fronts that govern the onset of slip [RCF04]. In a recent study [SF14], it was shown that the slowest of these three fronts indeed governs the rupture, under certain conditions: at driving velocities such that the rupture velocity is lower than the Rayleigh wave speed, the predictions from Linear Elastic Fracture Mechanics are in quantitative agreement with experiments.

In what follows we will discuss the question of the relevance of brittle fracture on domains much larger than a single junction, or which involve some loss of material (wear), which is a different question from that discussed by Fineberg and collaborators. As asperities and the surrounding domains are subject to high stresses and various geometrical constraints, one may naively expect mesoscopic fracture to be commonplace, especially during sliding. We are going to see some reasons for why fracture is not so common at the scale of micro asperities, but also how it can still be relevant in some cases.

In the static case (with no driving being performed), the asperities are subject to very high compressive stresses, which tend to decrease the probability of brittle fracture.[18] This is because the ductility (essentially the maximal plastic deformation possible before fracture) generally increases with the (hydrostatic) pressure. An intuitive but hand-waving argument for this is that high pressure tends to close the micro-cracks, vacancies and other voids generated by the plastic flow in the bulk of the solid. As these defaults are responsible for fracture (which always starts from the largest crack in the region under stress), their relative closing by pressure tends to diminish the occurrence of fracture.

During sliding, the shear stress at the contact points can become enormous (as for the compressive stress, this is due to the small contact area). At the level of

[18]The term brittles refers to "pure" fracture (without plastic deformation) as opposed to ductile fracture. We have already considered the plastic behaviours previously.

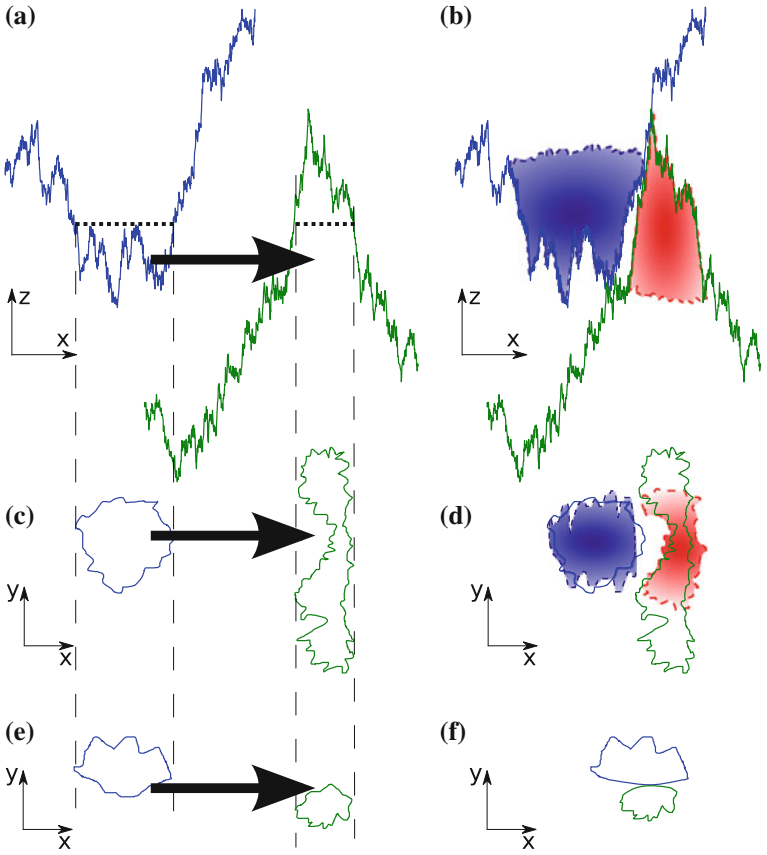

Fig. 2.16 Schematic picture of the situation of interlocking, fracture and elastic deformation scenarios. **a**, **b** Sectional view of two asperities meeting under external driving (*solid arrows*). **c**–**f** View from above of different scenarios. **c** Situation of interlocking. Two scenarios of fracture are suggested (**b**, **d**), with the broken zone shaded in *blue* (*left*) or *red* (*right*). **e** Situation of weak interlocking: elastic deformations can be enough to let asperities go through. **f** With some elastic (and a bit of plastic) deformation, asperities stay in their way

a single contact, as presented in Fig. 2.16d, the response will be to simply break the smallest possible cross-section of the welded asperities, which we identify as a simple junction breaking, which is not the point discussed here. However, in the configuration of "interlocking" (see Fig. 2.16), an asperity is subject to a high shear stress in a direction orthogonal to z (the main compressive stress). Thus one may expect the small asperities to easily break by brittle fracture, rather than deform elastically or plastically (we described these two mechanisms above). This is only half true.

On the one hand, Griffith's criterion shows that the critical stress for brittle fracture is directly controlled by the size of the largest micro-crack in the sample. So for very

small samples, this threshold stress will typically be very high: the smaller is the sample, the smaller its largest crack.[19]

On the other hand, we have to remember that roughness is expected at all scales. At large scales, all sorts of geometries can create locally high stresses on the "asperities" domains, which are always much larger than the typical unitary asperity contact area initially blocked (see Fig. 2.16d). In the context of tectonic plates for instance, the interlocking of "asperities" can involve locked areas over lengths ranging from the centimetres up to several meters (or more), with widths in the same range. For such large domains, the size of the largest crack available can become quite large, so that the elastic stresses will easily trigger macroscopic fracture. This threshold force needed for fracture contributes to the friction force.[20]

In both cases, it is interesting to note that a principle of selection is at play. The domains with the largest cracks (low fracture threshold) break first, and as slip occurs, only the hardest domains remain, so that during a slip phase, the prevalence of fracture typically decreases after a certain slip length D. We will come back to this in more detail in the next Sect. 2.2.4.

In the context of geophysics, it has been noticed that rocks are usually much less ductile than the materials commonly considered in tribology (metals, etc.), so that at equal external conditions, they break much earlier. Thus interlocking and the associated fracture process is expected to be quite important. An attempt at explaining friction as a process controlled mainly by the fracture of asperities was made in 1967 [Bye67], but the application of this theory has been limited to geophysics, where the presence of wear particles in large proportions makes such an hypothesis more likely.

To conclude, mesoscopic fracture plays a minor role in the dynamics of sliding[21] friction, regarding most applications. However when the system is either large, made of rocks or a combination of both, fracture can become equally relevant as adhesion in explaining friction. In geophysical applications, a comprehensive model for the sliding of plates would necessarily acknowledge the role of fracture. Let us recall that at the level of a single asperity, fracture is omnipresent, regardless of the nature of the material and of the external conditions. This fact is the basis for numerous works on friction [RCF04, RCF06, RCF07, BDRF10, BDF11, SF14].

[19] This is very intuitive, and the interested reader may try to make this reasoning more quantitative by using the branch of probability theory called Extreme Value Statistics (EVS).

[20] How does this force scales with the real contact area as determined in the previous section (from the elastic deformations)? Interlocking happens only where contacts are made, otherwise the large "bumps" would simply go by. Because of that, the density of number of interlocked domains is still proportional to the real contact area (itself proportional to the load). Then, the area of domains (which is roughly proportional to the fracture energy) depends on the real contact area in a rather intricate way, through the roughness exponent ζ. This is beyond the scope of this thesis.

[21] And of course, fracture plays an even smaller role in the dynamics of static friction.

Wear

In the context of friction, *wear* is usually defined quantitatively as the volume of particles which separates from one of the two surfaces *during sliding*. The separated particles may wander freely between the two surfaces (in this case, we may call them *debris*) or re-attach to the other surface. In both cases, wear corresponds to a change in the surfaces in contact (for engineering applications, it is sometimes only the net amount of debris which is relevant). In a first approximation, wear is proportional to the work performed by the friction force, hence it is proportional to the sliding length and friction force (but not directly to the velocity). Interlocking and the subsequent fracture obviously causes some wear. Let's quickly discuss a few other mechanisms which enter in the definition of "wear".

A mechanism which is slightly different from plain fracture and also causes wear is that of *adhesion wear*. When two asperities enter in contact and form a junction, depending on the micro-structure of each asperity close to the junction plane, the breaking of the bond may occur elsewhere than in the welding plane, so that one of the asperity keeps a piece of the other one. This part can either stay in place or get quickly separated from the asperity (due to the weakness of the joint): in both cases, we have some wear. This is a possible mechanism of wear, which has much to do with adhesion, hence the name. Note than the debris created in this way, or which are already present, can also re-attach to one of the two surfaces, thus "regenerating" the surface profile.

The term *abrasive wear* is used when one of the two surfaces is much harder than the other, in terms of plastic yielding stress σ_c. In this case, when an asperity of the harder material indents the other, it can plough a gutter (see Fig. 2.17) into it (instead of being deformed or break by brittle fracture). In a sense, ploughing is essentially plastic yielding along the surface plane, except that it can happen locally even when we are no longer in the plastic yield stress regime, macroscopically.

Conclusion

During sliding, the elastic response of asperities is twofold: for a part, asperities deform similarly as in the static case (see Fig. 2.16f), and for the other part they interlock. The interlocking of asperities can be overcome elastically if the height of the asperity involved is small enough. Fracture naturally appears as a limit of the elastic behaviour. And again, during sliding, the plastic response of asperities is twofold: to some extent, asperities deform similarly as in the static case by yielding against each other one at a time, but they may also plough long gutters into the opposite surface.

2.2.4 Ageing of Contact and Its Consequences

We have explained the first two laws up to here: because of high roughness the apparent area has little to do with the real one which is truly responsible for friction, and

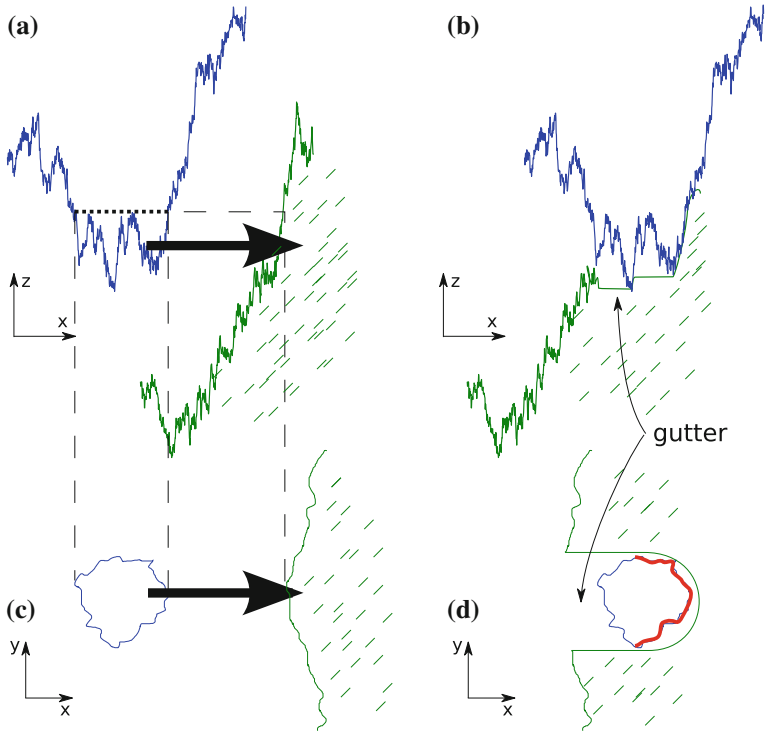

Fig. 2.17 Very schematic picture of the ploughing scenario. **a**, **b** Sectional view of two asperities meeting under external driving (*solid arrows*). **c**, **d** View from above of the ploughing scenario. If the *upper* (*blue*) material is much harder (larger yield stress σ_c) than the *lower* one (*green*), an asperity of the former may plough a gutter into the latter. The zone of the *upper solid* subject to the highest stresses is highlighted in *red* (**d**)

for various reasons this area generally ends up being proportional to the load. However regarding the third law and its corrected version (the Rate- and State dependent Friction laws), we have given no clue about the possible mechanisms yet.

The fact that RSF laws work very well (see Sect. 2.1.2) is not so surprising: with at least three fitting parameters (μ^*, a, b) and somewhat five (counting D_c and V^*), it is rather easy to "fit the data". This picture becomes more satisfying when several of these free parameters can be bound to some underlying physical mechanisms. We are about to interpret θ more precisely than just a "state" variable, and D_c much more precisely than a simple fitting/normalisation parameter. Other parameters can also be interpreted, but with some caution.

Microscopic Origins

Ageing: Definition Here, we are going to see that static friction and more precisely microscopical contacts display *ageing*, and we will give the link with the macroscopic RSF laws. Let's start with definitions.

We define the notion of ageing as the opposite of *stationary*: a system which displays ageing has some of its properties which change over time (i.e. they are not stationary). A process with ageing necessarily has some long-term memory (typically a power-law decay of the autocorrelation function over time).

A corollary is that with perfect knowledge of the microscopic dynamics of a process with ageing, one can typically estimate the "age" of a sample from a snapshot observation at some given time (i.e. a measure at a single time); the age being the time spent evolving from some default (known) starting configuration, to the observed one. The most common example of materials displaying ageing are glasses (see the course 7 of [ABC+02] or [Bir05] for an introduction on the topic, and [AKBC+98] for a discussion of an experimental example).

Creep In our study of the formation mechanisms of the real contact area we have omitted the aspect of temporal evolution. At first order, the processes we discussed are instantaneous: plastic yield, elastic response or brittle fracture all appear to happen very shortly after the appropriate constraints are applied. However, many time-dependent secondary processes interact with these main three, such as dislocation creep, desorption of protective films, formation of additional chemical bonds in the junction, cyclic fatigue, surface corrosion and wear of fresh surfaces, viscoelastic response in the bulk of asperities, elastic waves generated by ruptures, melting, etc.

The process usually recognized to be mainly responsible for variable behaviour over time is *plastic creep*. For crystalline materials, we may speak of dislocation creep [HBP+94, BHP94, PDW11]. Dislocations are produced through plastic events, and their slow thermally-activated displacement (this is what is called *creep*) may in return affect not only the plastic behaviour but also the condition for fracture. After a quick plastic yield occurs at the time of formation of a new contact, some new dislocations are formed (they are newborns, their age is zero). Since they have been freshly generated, they have not reach any equilibrium, nor even a stationary state. Thus they slowly diffuse and possibly trigger new (typically smaller) plastic events, which can in turn generate new dislocations. This dynamics progressively decelerates but never really reaches a stationary state, due to the infrequent bursts producing new dislocations: this is ageing [PDW11]. The mechanism for plastic creep in amorphous materials is a bit different (see [BL11] for a review and additional references). All in all, from our understanding of creep (or even the observation of macroscopic materials), one may expect asperities to age after contact and to slowly spread around the initial junction area.

Indeed, this effect has been observed directly in experiments. In 1994 [DK94], the diffraction of light through transparent samples allowed to directly observe the evolution of the true contact area over time.We reproduce these impressive results in Fig. 2.18. Conventional techniques of contact analysis prior to these works used to be *post-mortem*, i.e. after the surfaces had been in contact, one could analyse them to sort out the properties of the last contact zones. These post-mortem studies were of course unable to study the time evolution of contacts in such a way. From Fig. 2.18 it is clear that despite the constant load, the true contact area slowly increases over time, i.e. the contacts display some ageing.

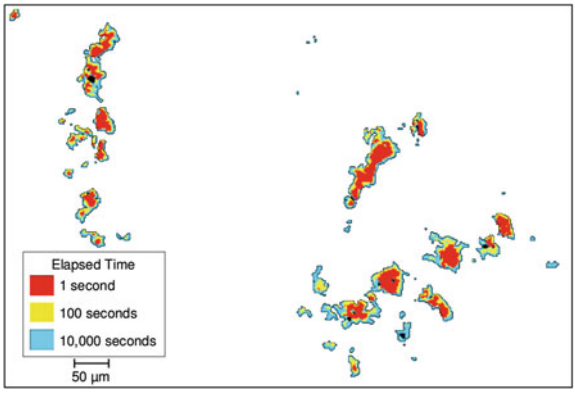

Fig. 2.18 From [DK94]. Using transparent samples, the contact area is made apparent thanks to the fact that the light diffracts everywhere but at the contacts. This allows to see the contacts directly, and most specifically to see their evolution over time

In the following paragraph we show how to incorporate this ageing into an effective law for friction as the RSF laws given above.

Interpretation of θ and D_c. Let's assume that θ represents the additional (or "bonus") contact strength due to ageing, and see how we can fit this idea into our observations.

On the one hand, at rest we expect a logarithmic increase of friction over time: if $\theta(v = 0) \sim t$, then the third term $b \ln(V^*\theta/D_c)$ goes like $\sim \ln(t) + \text{const}$. Still at rest, μ^* appears as the "instantaneous" friction, i.e. the friction obtained immediately after contact, due to the fast processes. The fact that $b \ll \mu^*$ corresponds well to the fact that creep is a secondary process, which only gives corrections to the main processes.

On the other hand at finite velocity the contacts "do not have time" to age: since the solid constantly slips, new contacts are constantly formed, and "old" ones broken. The crucial question is to estimate the contacts typical lifetime. Assuming a constant sliding velocity for the sake of simplicity, we may call D_c the "critical slip distance", i.e. the amount of slip (of the center of mass of the sliding block) necessary to break a newly formed junction. It takes a time $\theta_c = D_c/v$ for the bulk solid to slide over a distance D_c. Thus, the typical lifetime of a contact in the steady state is θ_c, so that the average or typical "bonus resistance" goes like $\sim \ln(\theta_c)$. This explains why in all RSF laws the evolution of θ must be chosen such that $\theta^{ss} = D_c/v$.

Similarly, the values of D_c can be interpreted straightforwardly. If asperities are sharp, in the sense that they resemble elongated needles, they may deform elastically and maintain contact over slip distances equal to several times the contact diameter D_a. On the contrary, if asperities are more like flat bumps with small heights, they will break contact as soon as the slip is a fraction of their contact diameter. In any case, for stronger bonds (larger contact diameter D_a), asperities will deform more before breaking, i.e. D_c increases with D_a. All in all, the contact-breaking slip distance is typically of the same order of magnitude as the asperities diameter, hence $D_c \sim$ 1–10 μm.

Now, in between $v = 0$ and $v = \text{const.} > 0$, there is a world of possibilities, and each RSF law (in particular the choice for the evolution of $\theta(t)$) will react differently to different experiments, as experiments with step-like variations of the driving velocity V_0, slip-hold-slip experiments, etc. The way in which each law reacts more or less realistically to each kind of input has been discussed in the reviews of reference, e.g. in [Mar98] where experiments are discussed, or more recently in [KHK+12] where the bibliography is abundant. However, we are not much interested in the details of each law's pros and cons: it is enough to note that no definitive consensus has been reached yet, and that a detailed microscopic analysis from which RSF laws would emerge is still missing. Thus, the problem in terms of fundamental physics is still largely open.

Interpretation of Other Variables The velocity V^* is merely a homogeneity constant: for any choice of units, it can be absorbed into μ^*. Thus, the value $V^* = 1\ \mu\text{m}\,\text{s}^{-1}$ is simply a convenient choice, since relevant velocities are usually of this order.

For μ^*, the interpretation seems quite simple: it is the default friction, corresponding to the fast processes we initially described (up to the absorption of constants as V^* and normalization expected at $t = 0$, depending on the exact form of the RSF law, (Eqs. 2.14 or 2.16)). In principle, μ^* can be estimated quantitatively: assuming a purely plastic formation of the true area of contact, we have $A_{\text{real}} = L/\sigma_c$, a number of bonds $N_{\text{bonds}} = A_{\text{real}}/A_{1\text{bond}}$, and a threshold breaking force per bond f_1. Denoting $F_\parallel$ the macroscopic shear force (tangential) and L the load (normal), this gives

$$\mu^* \approx \frac{F_\parallel}{L} = \frac{f_1}{\sigma_c A_{1\text{bond}}}, \tag{2.36}$$

where the yield strength σ_c is easy to measure, but the ratio $f_1/\sigma_c A_{1\text{bond}}$ is very hard to get.

The interpretation of a and b is usually directly related to creep [HBP+94, BHP94]. In a recent work [PDW11], the activation volume is defined in relation with the activation energy E^* ($\Omega^* = E^*/\sigma_c$) and the parameters a, b are predicted to be

$$a = \frac{k_B T}{\Omega^* \sigma_c}, \qquad b = \frac{k_B T}{\Omega' \sigma_c}, \tag{2.37}$$

where Ω' is some other activation volume. Unfortunately, direct access to these activation volumes and activation energies is difficult, so that these expressions for the RSF laws parameters are seldom used.[22]

Furthermore, the position of creep as dominant mechanism for ageing has been recently questioned in [LTGC11] where it was suggested that the strengthening of chemical bonds at junctions could be a more realistic explanation for the ageing of

[22] Furthermore, this interpretation of b is quite new and to be taken with caution. The interpretation of a is more commonly accepted [KHK+12], though it should still be taken with caution.

frictional contacts than creep. This casts doubts upon the trust we may put into old or current interpretations of the RSF laws in terms of plastic creep.

Stick-Slip Motion (with RSF Laws)

With a friction law that continuously depends on the sliding velocity v (velocity weakening), and possibly on some ageing "state" variable θ (increase of static friction over time), the dynamics of stick-slip becomes a bit more complex than what we forecasted in Sect. 2.1.1. However the main results are maintained: the existence of stick-slip in general and its disappearance at large velocity (V_0) or hard driving spring (k_0).

A thorough study of RSF laws applied to a single degree of freedom (a simple rigid block) was performed early in [GRRT84]. A more concise study of this problem was performed in [RT86], where the differences with the Amontons-Coulomb laws were emphasized. There, the main difference with the more simple law of friction is the emergence of two time scales or velocities (instead of one). For velocities below a first threshold, the motion is essentially described by the quasi-static picture (which neglects the velocity dependence). For velocities above this threshold but below the second one, the dynamical effects cannot be neglected. Above the second threshold, stick-slip disappears (similarly to what we found in our simpler model, Sect. 2.1.1).

Another complete, yet concise study of stick-slip motion was performed in [BCPR95]. They compare experimental results for paper on paper stick-slip with analytical computations (weakly non linear analysis around the Hopf bifurcation) and numerical integrations using the most common Rate-and-state friction law, (Eqs. 2.14 and 2.15).

Of course, different dynamics of stick-slip can be obtained when using various rate-and-state laws. However, the main features we are interested in remain the same: as sliding velocity increases, stick-slip motion shifts from very regular to rather chaotic, to non-existent. This rich behaviour has been the playground for intensive studies in Geophysics, as we will see in Chap. 3.

The RSF law is particularly useful in geophysics in order to study the dynamics of stick-slip, which involves the static friction coefficient and where the departure from zero to finite velocity is especially relevant. This is what we explain in the next section.

2.3 Conclusion: Friction Involves Randomness and Viscoelasticity

We have presented the basic phenomenology of Friction. The three historical laws have been amended to account correctly for the dependence on the sliding velocity, a crucial point in the study of the dynamical stability of frictional systems (stick-

slip instability). Intuition about the physical mechanisms behind these laws has been supported with direct observations (microscopic measurements of surface profiles, contact area and its time evolution). Simple models of (implicitly static) contact have been presented, outlining the role of elasticity, plasticity and some secondary mechanisms (fracture, plastic creep).

Microscopic models of friction should take into account the presence of randomness. A first source of randomness are the thermal fluctuations, responsible for the plastic creep [PDW11] which plays a key role in the ageing of contacts. A second source of randomness is the presence of "quenched disorder" induced by the heterogeneities and the roughness of the surfaces. The idea that surface self-affinity is crucial to the friction properties is now well-established [Per01, PAT+05], in particular it naturally explains the second friction law of Amontons. However most of the phenomenological models (e.g. [RB91, PAT+05]) deal with the average properties induced by the disorder and neglect the fluctuations of the dynamics. As we will see in Chap. 4, the validity of this assumption is a matter of scale [PT96, CN98]. At moderate scales (such as in laboratory experiments), the motion can be described by deterministic effective equations such as the Rate-and-State equations. At much larger scales, the motion is actually stochastic and displays a very complex avalanche dynamics. This is in particular the case for fault dynamics, which is characterized by random bursts of activity (earthquakes) that are random in magnitude, temporal and spatial location.

There have been a few tentative friction models including real randomness, but they have found rather limited echo until now: [RB91] is an example that received unfairly small attention. Excellent reviews on this topic are [KHK+12, VMU+13], but we will come back to this at length later. The problem with all other attempts is that they fail either at correctly account for randomness, or they overlook the role of microscopical ageing which is crucial in producing the RSF laws. All in all, no definitive consensus has been reached to this day on the foundations of the RSF law(s), even when resorting to such models: the search for a convincing yet simple microscopical model reproducing a realistic RSF law is still an open problem.

To summarize—crudely—there are two main issues that must be addressed in order to properly deal with friction. The first is the fluctuating, heterogeneous nature of the contacts involved: one must use a stochastic approach. The second is the ageing inherent to the microscopic mechanisms of contact. To deal with that, considering the natural field or degree of freedom (usually the stress field or the location of the current contacts) characterizing the instantaneous state of the system is not enough. One must include some additional degree of freedom atop the natural one, i.e. consider the dynamics of the displacement field to be non-Markovian.

References

[ABC+02] Ajdari, Armand, Jean-Philippe Bouchaud, Bernard Cabane, Michael Cates, Sergio Ciliberto, Letitia Cugliandolo, Alexei Finkelstein, Daniel Fisher, Walter Kob, Miguel Ocio, Zvi Ovadyahu, Giorgio Parisi, Zoltan Racz, and David Wales. 2002. *Slow relaxations and nonequilibrium dynamics in condensed matter*, Session Lxxvii, 1–26 July 2002. Berlin: Springer.

[AKBC+98] Alberici-Kious, F., J. Bouchaud, L. Cugliandolo, P. Doussineau, and A. Levelut. 1998. Aging in K1-xLixTaO3: a domain growth interpretation. *Physical Review Letters* 81(22): 4987–4990.

[Arc57] Archard, J.F. 1957. Elastic deformation and the laws of friction. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 243(1233): 190–205.

[AT06] Aranson, Igor, and Lev Tsimring. 2006. Patterns and collective behavior in granular media: theoretical concepts. *Reviews of Modern Physics* 78(2): 641–692.

[BCPR95] Baumberger, T., C. Caroli, B. Perrin, and O. Ronsin. 1995. Nonlinear analysis of the stick-slip bifurcation in the creep-controlled regime of dry friction. *Physical Review E* 51(5): 4005–4010.

[BDF11] Ben-David, Oded, and Jay Fineberg. 2011. Static friction coefficient is not a material constant. *Physical Review Letters* 106(25): 254301.

[BDRF10] Ben-David, Oded, Shmuel M. Rubinstein, and Jay Fineberg. 2010. Slip-stick and the evolution of frictional strength. *Nature* 463(7277): 76–79.

[BGK06] Hans-Jürgen Butt, Karlheinz Graf, and Michael Kappl. 2006. *Physics and chemistry of interfaces*. Hoboken: Wiley.

[BHP94] Baumberger, T., F. Heslot, and B. Perrin. 1994. Crossover from creep to inertial motion in friction dynamics. *Nature* 367(6463): 544–546.

[Bir05] Biroli, Giulio. 2005. A crash course on ageing. *Journal of Statistical Mechanics: Theory and Experiment* 2005(05): P05014.

[BL11] Barrat, Jean-Louis, and Anael Lemaitre. 2011. Heterogeneities in amorphous systems under shear. In *Dynamical heterogeneities in glasses, colloids, and granular media*, ed. Ludovic Berthier, Giulio Biroli, Jean-Philippe Bouchaud, Luca Cipelletti, and Wim van Saarloos. Oxford: Oxford University Press.

[BSSBB14] Bar-Sinai, Yohai, Robert Spatschek, Efim A. Brener, and Eran Bouchbinder. 2014. On the velocity-strengthening behavior of dry friction. *Journal of Geophysical Research: Solid Earth* 119(3): 1738–1748.

[Bye67] Byerlee, James D. 1967. Theory of friction based on brittle fracture. *Journal of Applied Physics* 38(7): 2928.

[CBU13] Capozza, Rosario, Itay Barel, and Michael Urbakh. 2013. Probing and tuning frictional aging at the nanoscale. *Scientific Reports* 3: 1896.

[CN98] Caroli, C., and P. Nozières. 1998. Hysteresis and elastic interactions of microasperities in dry friction. *The European Physical Journal B* 4(2): 233–246.

[Cun93] Cundill R.T. 1993. *Ball Bearing Journal* 241: 26.

[Die79] Dieterich, James H. 1979. Modeling of rock friction: 1. Experimental results and constitutive equations. *Journal of Geophysical Research* 84(B5): 2161.

[DK94] Dieterich, James H., and Brian D. Kilgore. 1994. Direct observation of frictional contacts: new insights for state-dependent properties. *Pure and Applied Geophysics PAGEOPH* 143(1–3): 283–302.

[Dow79] Dowson, D. 1979. *History of tribology*. New York: Longman.

[FBK13] Ferrero, E., S. Bustingorry, and A. Kolton. 2013. Nonsteady relaxation and critical exponents at the depinning transition. *Physical Review E* 87(3): 32122.

[GRRT84] Gu, Ji-Cheng, James R. Rice, Andy L. Ruina, and Simon T. Tse. 1984. Slip motion and stability of a single degree of freedom elastic system with rate and state dependent friction. *Journal of the Mechanics and Physics of Solids* 32(3): 167–196.

[GTvHV12] Gómez, Leopoldo R., Ari M. Turner, Martin van Hecke, and Vincenzo Vitelli. 2012. Shocks near jamming. *Physical Review Letters* 108(5): 058001.

[HBP+94] Heslot, F., T. Baumberger, B. Perrin, B. Caroli, and C. Caroli. 1994. Creep, stick-slip, and dry-friction dynamics: experiments and a heuristic model. *Physical Review E* 49(6): 4973–4988.

[KBD93] Kilgore, Brian D., Michael L. Blanpied, and James H. Dieterich. 1993. Velocity dependent friction of granite over a wide range of conditions. *Geophysical Research Letters* 20(10): 903–906.

[KHK+12] Kawamura, Hikaru, Takahiro Hatano, Naoyuki Kato, Soumyajyoti Biswas, and Bikas K. Chakrabarti. 2012. Statistical physics of fracture, friction, and earthquakes. *Reviews of Modern Physics* 84(2): 839–884.

[KRGK09] Kolton, Alejandro, Alberto Rosso, Thierry Giamarchi, and Werner Krauth. 2009. Creep dynamics of elastic manifolds via exact transition pathways. *Physical Review B* 79(18): 184207.

[Kri02] Krim, J. 2002. Resource letter: FMMLS-1: friction at macroscopic and microscopic length scales. *American Journal of Physics* 70(9): 890.

[LBI13] Lee, Dong Woog, Xavier Banquy, and Jacob N Israelachvili. 2013. Stick-slip friction and wear of articular joints. *Proceedings of the National Academy of Sciences of the United States of America* 110(7): E567–E74.

[LTGC11] Li, Qunyang, Terry E Tullis, David Goldsby, and Robert W Carpick. 2011. Frictional ageing from interfacial bonding and the origins of rate and state friction. *Nature* 480(7376): 233–236.

[Mar98] Marone, Chris. 1998. Laboratory-derived friction laws and their applicaton to seismic faulting. *Annual Review of Earth and Planetary Sciences* 26(1): 643–696.

[MlyHHR92] Må løy, Knut, Alex Hansen, Einar Hinrichsen, and Stéphane Roux. 1992. Experimental measurements of the roughness of brittle cracks. *Physical Review Letters* 68(2): 213–215.

[MTS09] Mo, Yifei, Kevin T. Turner, and Izabela Szlufarska. 2009. Friction laws at the nanoscale. *Nature* 457(7233): 1116–1119. February.

[PAT+05] Persson, B.N.J., O. Albohr, U. Tartaglino, A.I. Volokitin, and E. Tosatti. 2005. On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. *Journal of Physics. Condensed Matter : An Institute of Physics Journal* 17(1): R1–R62.

[PDW11] Putelat, Thibaut, Jonathan H.P.P. Dawes, and John R. Willis. 2011. On the microphysical foundations of rate-and-state friction. *Journal of the Mechanics and Physics of Solids* 59(5): 1062–1075.

[Per00] Persson, Bo N.J. 2000. *Sliding friction: physical principles and applications*. NanoScience and technology, 2nd edn. Berlin: Springer.

[Per01] Persson, B.N.J. 2001. Theory of rubber friction and contact mechanics. *The Journal of Chemical Physics* 115(8): 3840.

[PSS+08] Persson, B.N.J., I.M. Sivebaek, V.N. Samoilov, K.E. Zhao, A.I. Volokitin, and Z. Zhang. 2008. On the origin of Amonton's friction law. *Journal of Physics: Condensed Matter* 20(39): 395006.

[PT96] Persson, B.N.J., and E. Tosatti (eds.). 1996. *Physics of sliding friction*. Dordrecht: Springer.

[RB91] Rundle, John B., and Stephen R. Brown. 1991. Origin of rate dependence in frictional sliding. *Journal of Statistical Physics* 65(1–2): 403–412. October.

[RCF04] Rubinstein, Shmuel M., Gil Cohen, and Jay Fineberg. 2004. Detachment fronts and the onset of dynamic friction. *Nature* 430(7003): 1005–1009.

[RCF06] Rubinstein, S., G. Cohen, and J. Fineberg. 2006. Contact area measurements reveal loading-history dependence of static friction. *Physical Review Letters* 96(25): 256103.

[RCF07] Rubinstein, S., G. Cohen, and J. Fineberg. 2007. Dynamics of precursors to frictional sliding. *Physical Review Letters* 98(22): 226103.

[RT86] Rice, James R., and Simon T. Tse. 1986. Dynamic motion of a single degree of freedom system following a rate and state dependent friction law. *Journal of Geophysical Research* 91(B1): 521.

[Rui83] Ruina, A. 1983. Slip instability and state variable friction laws. *Journal of Geophysical Research* 88: 10459 10.

[SF14] Svetlizky, Ilya, and Jay Fineberg. 2014. Classical shear cracks drive the onset of dry frictional motion. *Nature* 509(7499): 205–208.

[SLPD09] Scheibert, J., S. Leurent, A. Prevost, and G. Debrégeas. 2009. The role of fingerprints in the coding of tactile information probed with a biomimetic sensor. *Science (New York, N.Y.)* 323(5920): 1503–1506.

[VMU+13] Vanossi, Andrea, Nicola Manini, Michael Urbakh, Stefano Zapperi, and Erio Tosatti. 2013. Colloquium: modeling friction: from nanoscale to mesoscale. *Reviews of Modern Physics* 85(2): 529–552.

[WCDP11] Wandersman, E., R. Candelier, G. Debrégeas, and A. Prevost. 2011. Texture-induced modulations of friction force: the fingerprint effect. *Physical Review Letters* 107(16): 164301.

http://www.springer.com/978-3-319-20021-7

Viscoelastic Interfaces Driven in Disordered Media
Applications to Friction
Landes, F.P.
2016, XIII, 206 p. 79 illus., 37 illus. in color., Hardcover
ISBN: 978-3-319-20021-7