Eur. Phys. J. B **4**, 233–246 (1998)

THE EUROPEAN PHYSICAL JOURNAL B

Hysteresis and elastic interactions of microasperities in dry friction

C. Caroli[1] and Ph. Nozières[2,a]

[1] Groupe de Physique des Solides, Université Paris 6 et Paris 7 et CNRS, Tour 23, 2 place Jussieu, 75251 Paris Cedex 05, France
[2] Institut Laue-Langevin et Collège de France, BP 156, 38042 Grenoble Cedex 9, France

Received: 3 February 1998 / Accepted: 19 March 1998

Abstract. Velocity independent dry friction of a slider upon a base is due to an hysteretic response of relative displacement ρ to a tangential driving force F. We show that the purely elastic model for multistability considered in a previous publication is in no way essential: multistability arises just as well from adhesion. We emphasize the physical consequences of multistability for dynamic/static, a.c./d.c. friction. When the slider is moved from rest by an amount ρ the transition from the zero force static configuration to dynamic behaviour is progressive, spreading on a range equal to the width of the hysteresis cycle. When ρ is small, an elastic restoring force ensues, in agreement with observations. The competition of that elastic pinning with bulk elasticity generates a *screening length* which we believe is the natural size of Burridge Knopoff blocks. We then study the effect of elastic interactions between asperities: it is weak for dilute asperities, but its long range makes it important. In lowest order the interaction mediated displacement of a given asperity has logarithmically divergent fluctuations: they become comparable to the asperity radius when the slider size reaches another characteristic "Larkin length" λ_L, which for dilute micronic asperities is exponentially large. We give arguments suggesting that individually monostable asperities display *collective multistability* on scales larger than λ_L. For individually multistable sites we show that elastic interactions give rise to cascade processes in which the spinodal jump of a given asperity triggers the jump of others. We estimate the size of these cascades that should show up in the noise spectrum.

PACS. 46.30.Pa Friction, wear, adherence, hardness, mechanical contacts, and tribology – 81.40.Pq Friction, lubrication, and wear

Dry friction of a slider on a base is a very old problem, at crossroads between surface physics, mechanics and engineering. It has many facets, depending on the geometry of surfaces in contact, and also on the possible presence of contaminant (or lubricant) layers. Among the factors affecting friction, the *elastic* response to surface perturbations (crucial for long range interaction effects) plays an important role – but *adhesion* and *plastic flow* of the asperities are equally essential [1,2]. Disentangling these features is a difficult problem which we do not approach here: we are concerned only with the elastic part of the story.

Let us first set the qualitative stage. We consider a simple situation in which slider and base are in contact at *dilute* microcontacts between facing asperities (such a geometry is realistic in a wide range of normal loads and it was advocated by Greenwood and Williamson [3] in order to explain why friction does not depend on the apparent surface of contact). A given microcontact is characterized by a configurational coordinate ρ measuring the horizontal distance between its two constituent asperities. The resulting interaction $V(\rho)$, whether adhesive or elastic, produces a tangential force $F = -V'$; the elastic response of the bulk to that surface force acts to shift the coordinate ρ. A multistable regime may arise either from the potential $V(\rho)$ itself or from the elastic response: whatever its origin the corresponding hysteretic behaviour produces a macroscopic friction force independent of the pulling velocity [4] – just the feature we want for dry friction. Within that general frame we will address two questions:

(i) Can a single asperity display multistability? How does macroscopic friction ensue? That point was discussed in a previous paper putting all the emphasis on purely elastic effects. Here we show that the frictional behavior induced by contact hysteresis is quite general, *i.e.* independent of its physical origin.

(ii) What is the influence of *elastic* interactions between asperities? We will see that new physical concepts emerge, leading to the definition of several characteristic lengths.

[a] e-mail: `nozieres@ill.fr`

As it stands our paper does not claim to be "a" theory of dry friction: it is far too oversimplified for that purpose. Our aim is to provide simple guidelines in formulating the problem.

Let n_s and n_b be the densities of asperities per unit area respectively on the slider and base. A contact is "active" when two asperities face each other: the density of these asperities is consequently

$$n = n_s n_b a^2$$

where a is a typical asperity size (the probability that a given asperity on the slider finds one facing it on the base is $n_b a^2$). From $n = 1/d^2$ we infer the distance d between active asperities: "dilute" means $a \ll d$. Let the positions of the two facing asperities be

$$r_{is} = R + \frac{\rho}{2} \quad r_{ib} = R - \frac{\rho}{2}.$$

R is the *position* of the contact and ρ its *configuration coordinate* mentioned above. In a constant velocity drift ρ is time averaged for each asperity while R pertains to an ensemble average over all geometries. For a system at rest, both ρ and R are ensemble averaged. In what follows we shall consider only a naive one dimensional model for ρ: we do it for the sake of simplicity as it displays most of the important physics. It is clear that the real situation is 2D: the corresponding extension has been discussed by Tanguy and one of us [5]. It is non trivial since the asperities can escape each other sideways: the issue of tangential stability is essential. But the qualitative picture that emerges is not modified much.

Consider first a single contact (or equivalently ignore the elastic interaction with its neighbours). The interaction energy of slider and base is $V(\rho)$. In [4] we attributed that interaction to elasticity, as in a Hertz contact: the two facing asperities retract in order to leave way to each other. Such a contribution is definitely present, but it is not the only one: adhesion also contributes. Altogether we take V as a *phenomenological input*, assuming only that it is the non dissipative potential energy of a system in quasistationary equilibrium. Put another way we ignore the possibility of dissipative friction on a scale *smaller* than the asperity size a. That is by no means obvious – indeed it raises the whole issue of a hierarchy of friction forces, motion on a scale l depending on friction mechanisms at smaller scales l'. This is another story, albeit important[1]: we do not tackle it.

If the asperity responds without internal friction, a continuous drift with an infinitesimal velocity v cannot dissipate energy: the system remains in quasiequilibrium and no entropy is produced. Any viscous friction due to a retarded elastic response would automatically result in a friction force proportional to v, in contradiction to Coulomb's law. Finite dissipation at very low velocity can only occur if the contact is *multistable* in some range of values of its configuration coordinate, jumping discontinuously from one state to another at a spinodal limit: when the jump occurs, a bunch of phonons is radiated towards the bulk, carrying energy away. What matters is not dissipation during the jump, which is utterly negligible, but dissipation *after* the jump: the radiated phonons never return to the asperity and they are lost as far as the energy balance is concerned. We thus conclude that dry friction can only be explained in terms of *hysteresis*, a point which has been appreciated long ago [6,7]. Indeed the macroscopic force-displacement characteristic is obviously hysteretic: the only relevant questions are "where does the hysteresis come from?". Does it appear on the atomic level or on a (sub)micronic scale? Does it exist for a single contact or is it a collective effect due to interactions? In this paper we provide some answers to these questions. In Section 1, we complement our previous work for a single asperity. Section 2 is devoted to elastic interactions.

1 Multistability of a single contact

Assume for a moment that slider and base are rigid (no elastic shear response). The asperity is characterized by a potential energy $V(\rho) = V(r_s - r_b)$ the physical origin of which we do not specify for the moment. This potential gives rise to tangential forces $F_s = -F_b = -V'$ respectively on the slider and base. Due to the elastic response of the slider and base to these forces, acting in opposite directions, the two asperities recoil respectively to $(r_s + u_s)$ and $(r_b + u_b)$: the configuration coordinate ρ becomes $\eta = (\rho + u)$ where $u = u_s - u_b$ is the *relative* displacement[2]. (In addition to a displacement there is a local deformation u_{ij} which we ignore: we assume it is absorbed in the potential $V(\rho)$). Such an elastic response modifies F and V: the real energy cost of the contact is the *total* energy, pinning + elastic:

$$U = V(\rho + u) + \frac{1}{2}[\lambda_s u_s^2 + \lambda_b u_b^2]. \tag{1}$$

The second term in (1) is the elastic energy stored in the bulk. λ_s and λ_b are elastic stiffnesses which for a single scale asperity are of order Ea. We note that equilibrium at a given ρ corresponds to a minimal U:

$$\lambda_s u_s + V' = \lambda_b u_b - V' = 0. \tag{2}$$

[1] This issue is certainly essential to understand the so-called "direct effect" in the phenomenological friction laws *à la* Ruina-Rice, as well as the slow precursory sliding motion observed under loads close below the static threshold.

[2] Reducing the elastic response of the bulk to a displacement u is an oversimplification: in reality the asperity is subject to a displacement field $u(r)$, implying an elastic *deformation*. As a result the definition of u is ambiguous, depending on where it is measured. In our schematic model we choose to ignore such a complication: the local deformation is supposed to correct the pinning potential V and the stiffness λ. In order to display multistability, we retain only an average u, the precise definition of which is lacking (may be the maximum $u(r)$?).

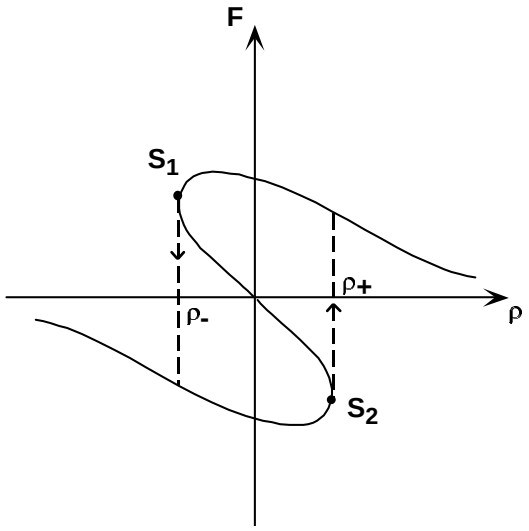

Fig. 1. A multistable $F(\rho)$: when the slider moves to the right (left), the contact remains on the lower (upper) branch until the spinodal limit $\rho_+(\rho_-)$ where it jumps to the other branch.

It follows that $\lambda_s u_s + \lambda_b u_b = 0$: we can therefore write everything in terms of the *relative* displacement u,

$$U = V(\eta) + \frac{1}{2}\lambda u^2 \tag{3a}$$

$$V'(\eta) + \lambda(\eta - \rho) = 0 \tag{3b}$$

where $\lambda = \lambda_s\lambda_b/(\lambda_s + \lambda_b)$ is an effective stiffness. Equation (3b) fixes u as a function of ρ: the total energy U depends on ρ both explicitly and implicitly *via* u.

The real control parameter is the nominal ρ *before* elastic response, not η (the force on the slider is applied far away from the contact): we thus express all quantities in terms of ρ. Since we minimized with respect to u:

$$\frac{dU}{d\rho} = \frac{\partial U}{\partial \rho} = V'(\eta) = -F. \tag{4}$$

From now on we consider $U(\rho)$ or, equivalently, $F(\rho)$ as the input of the model, forgetting for a while the question of where they come from (note that, due to elastic response, the bare V is replaced by the total U).

The simplest case is that in which U and F are single valued: then the contact remains everywhere in quasistatic equilibrium and no dissipation can occur. This is seen explicitly if we calculate the work of the pinning force when the slider is swept:

$$w = -\int_{-\infty}^{+\infty} F d\rho \tag{5}$$

(the external force is applied far from the impurity where the displacement is $d\rho$, not $d\eta$). Since $F = -dU/d\rho$ the integral vanishes identically, as expected.

The situation is different when F and U are multivalued, *i.e.* when the equilibrium of the contact is multistable. A typical sketch of the force $F(\rho)$ is shown in Figure 1. The spinodal limits ρ_- and ρ_+ mark the limits of metastability: in between, on the reentrant branch, equilibrium is unstable. The curve $F(\rho)$ is easily integrated into the energy curve $U(\rho)$ shown in Figure 2: the $(\rho - \rho_\pm)^{1/2}$ behaviour of F results in cusps of $U \approx (\rho - \rho_\pm)^{3/2}$. Such a multistable regime leads to hysteresis and to discontinuous jumps when the slider reaches the spinodal limit: *velocity independent dry friction is a direct consequence of that hysteresis*, whatever its origin. We analyze first the content of that statement: a brief discussion of why hysteresis may occur will come next. Most of the single asperity properties were described in [4]: here we add some comments and clarifications.

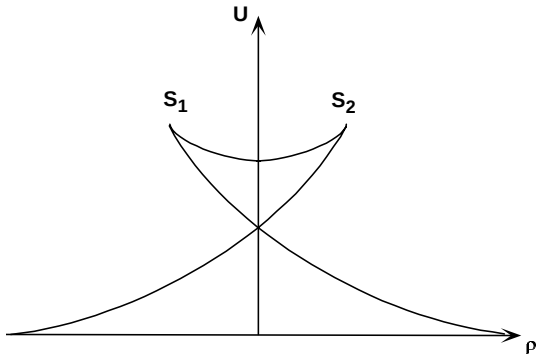

Fig. 2. The total (elastic + pinning) energy curve $U(\rho)$: the spinodal limits correspond to the cusps.

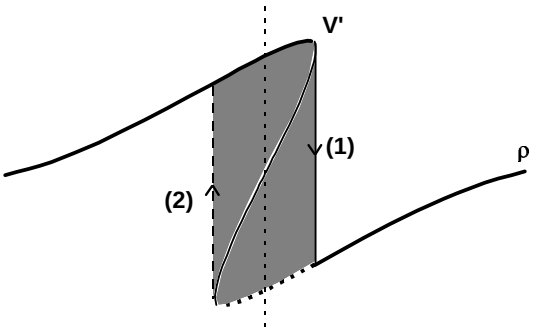

Fig. 3. Contact population (thick line) for sliding toward increasing ρ. The dynamic friction force is proportional to the striped (dotted) area for forward (backward) sliding.

1.1 Dynamic friction

The work of the pinning force when the slider is swept is given by (5): it is the hatched area of Figure 3 for forward drift (the dotted one for backward drift). For a round trip the dissipation is the area enclosed by the dissipation loop, as in magnetism. Let $dn = \nu d\rho$ be the density (per unit area) of asperities swept when the slider moves by $d\rho$: the dissipated energy is wdn and consequently the macroscopic friction force per unit area is $F = \nu w$ (its work is the dissipated energy). In order to estimate ν we note that

a given asperity sweeps an area $ad\rho$: hence

$$\nu \approx n_s n_b a \approx \frac{n}{a} \cdot \tag{6}$$

In practice the asperities are random and the work w is a stochastic quantity which must be averaged.

1.2 a.c. dissipation

When the slider oscillates with an amplitude A only those hysteresis cycles that jump forward *and* backward contribute to the steady dissipation. If an asperity jumps only forward, it will do it once and thereafter it will remain in quasistatic equilibrium on the same branch of the curve $F(\rho)$. The dependence of the a.c. dissipation on the oscillation amplitude thus provides information on the distribution of loop widths 2ξ When A is very small dissipation disappears (the relevant areas w goes to 0).

If an oscillation is superimposed upon a steady drift, two cases may occur depending on the amplitude A. For small amplitudes the instantaneous velocity is always forward: the friction force retains its value for steady drift. For large velocity modulation A, the slider has backward lapses: the friction force then goes to zero (except for backward jumps of very small cycles). The net dissipation is the same as for d.c. drift (each asperity jumps once), but the time dependence is an alternation of finite and zero forces.

Clearly, in real systems the analysis of such phenomena is complicated by the necessity to disentangle them from the contributions of dissipative processes occurring on the internal contact scale and, possibly, of slider viscoelasticity. For this purpose, it may be useful to note that, while the latter effects necessarily vanish in the low frequency/velocity limit, dissipation due to spinodal jumps is frequency-independent.

1.3 Force noise

All these results crucially depend on an ensemble average over asperity configurations and hysteresis cycles: fluctuations give rise to noise. For $N = n\ L^2$ statistically independent active asperities the noise amplitude is $\sqrt{N|\overline{F^2}|}$ (L is the width of the slider). The characteristic displacement scale of that noise is $\Delta\rho \approx a$: when the slider has moved by a all the contacts have been destroyed and replaced by a new set, and a fresh landscape ensues. If we ignore elastic interactions between asperities each of them lives its own life. When interactions are included it may happen that one jump triggers others, which reflects in the noise spectrum: we discuss these "cascade jumps" in the next section.

1.4 Static equilibrium and static friction

The ensemble average over active asperities is characterized by a uniform statistical distribution of their nominal configuration coordinates ρ. When the contact is multistable, one must also specify which of the two metastable branches is occupied: we thus define an occupation probability $P(\rho)$ for, say, the lower branch. Consider first a dynamical regime in which the slider is swept to the right: all contacts are drawn to the lower branch of the $F(\rho)$ curve, up to the spinodal limit S_+ (see Fig. 1): $P(\rho)$ is therefore 1 throughout the multistable region. The net force on the slider is the dynamic friction discussed before. If the external force is suddenly suppressed, the slider must *recoil* by an amount δ in order to achieve a zero net F. The new probability $P(\rho)$ in the resulting static state jumps from 1 to 0 at $\rho = (\rho_+ - \delta)$, the two branches partaking in the static configuration, as shown in Figure 4. (For narrow cycles $\xi < \delta$ the contact may even become monostable). If all asperities were the same (and symmetric) the discontinuity would occur at the Maxwell plateau of the multistable $F(\rho)$. In practice, as shown by Greenwood and Williamson [3], the width ξ is randomly distributed about a peak average value, and the force F is zero only on the average: broad cycles do not reach the Maxwell plateau, narrow ones overcome it.

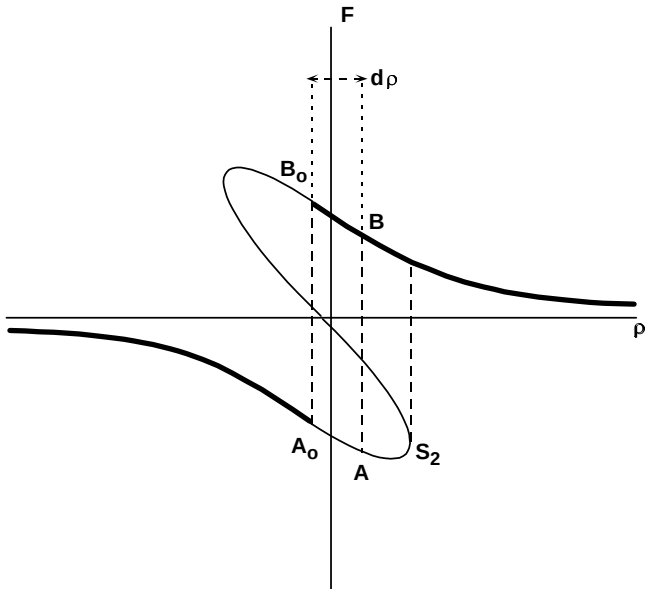

Fig. 4. The occupation probability $P_0(\rho)$ in the static regime under zero tangential force. $P_0 = 1$ on the heavy branches, 0 elsewhere. When the slider is moved by $d\rho$, the discontinuity of P_0 shifts from A_0B_0 to AB. The total restoring force is proportional to the area A_0B_0AB.

In the above argument we implicitly assumed that the recoil δ was the same for all asperities: put another way, slider and base recoil *rigidly*, the distance between asperities remaining fixed. That is by no means obvious since elasticity is the basic ingredient of the model. We shall see shortly that recoil is indeed rigid on length scales smaller than a *screening length* λ_d. On larger scales the above discussion is oversimplified.

If a tangential force is applied again, the discontinuity of $P(\rho)$ moves back toward increasing ρ. Sliding starts

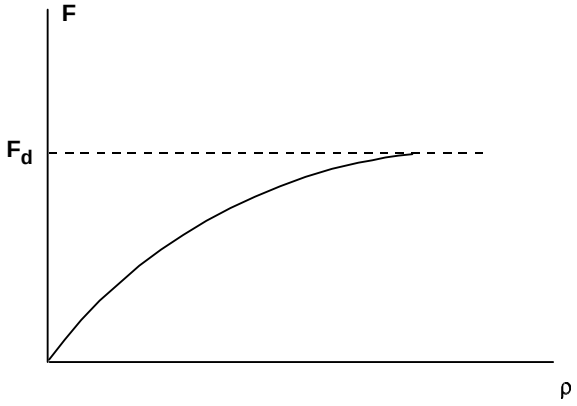

Fig. 5. Starting from the static configuration under zero external tangential force, the macroscopic force first displays an elastic regime when the slider is displaced by an amount ρ.

when it reaches the spinodal limit: it follows that the *static* threshold force is identical to the dynamic friction F, in contradiction with the well known Coulomb law. Indeed the dynamic F_d is the *maximum* force that the asperities can develop: our model has no way to produce a $F_s > F_d$. In order to account for such a behaviour, one has to introduce plastic effects into the framework described here, as discussed by Velicky and one of us [8].

One important point should be kept in mind. While there is only one way to achieve a steady drift of the slider, one can prepare a state with $F = 0$ in many ways. Stopping a drift is one of them, but one might think of bringing slider and base in contact keeping $F = 0$ all the way: the resulting probability $P(\rho)$ will be different (very likely the discontinuity will be close to the Maxwell plateau for all asperities, fixed by the "birth" of reentrant cycles). Stopping a driving motor is another, more commonly realized situation. The slider then remains at rest, after a decelerating transient, under a finite shear loading force $< F_s$ [9]. Another $P(\rho)$, different from the two above ones ensues.

In real experimental situations, the contacts strengthen slowly under the effect of plastic creep under load, which results in a slow increase of F_s with the time spent at rest before sliding. Different distributions $P(\rho)$ correspond to different distributions of pinning forces, hence to different rates of creep strengthening. Static friction thus depends on the past history, a feature which is confirmed experimentally [10].

1.5 Elasticity in the static regime, screening length

Starting from static equilibrium $F = 0$, let us move the slider by an infinitesimal amount $d\rho$: the discontinuity of $P(\rho)$ shifts from ρ_0 to $(\rho_0 + d\rho)$, as shown in Figure 4. The displacement creates a net force per unit area $\Delta F dn$ where ΔF is the discontinuity $(F_B - F_A)$ and dn the density of contacts which lie between ρ_0 and $(\rho_0 + d\rho)$. A typical cycle width is $\xi \approx a$ and thus $dn \approx n\, d\rho/a$: a slider at rest is subject to an *elastic restoring force* $dF = -\alpha d\rho$, the elastic constant (per unit area) being $\alpha \approx n\, \Delta F/a$. When all asperities have reached their spinodal limit the dynamic distribution is restored: F saturates at its dynamic value F_d as shown in Figure 5. The important result is the existence of an *elastic* region between the two leftward and rightward dynamic regimes [9].

Such a surface pinning force competes with the bulk elasticity. Assume that each interface point x is subject to an *extra* displacements $u_s(x)$ and $u_b(x)$ respectively for slider and base. What matters is the relative displacement: we set $u_b(x) = 0$, $u_s(x) = u$. The force per slider unit area $\Phi(x)$ needed in order to achieve such a displacement will have two components

- the above surface pinning part $-\alpha u(x)$ due to asperities;
- the usual bulk term due to the elastic stress field induced by u.

In the simplest case of a single Fourier component with wave vector k, for a semi-infinite slider:

$$\Phi_k = \left[\alpha + \frac{|k|E_s}{2(1-\sigma^2)}\right] u_k. \tag{7}$$

Note that the bulk term vanishes when $k \to 0$, as expected (a uniform translation of the surface with fixed center of mass creates a small shear $u_{zx} \approx u/H$ where H is the slider thickness: the corresponding σ_{zx} vanishes if $H \to \infty$). The factor $|k|$ is familiar in all problems in which the perturbation penetrates in the bulk over a wave length k^{-1}. We can invert (7) in order to obtain the displacement u induced by a given surface force Φ:

$$u_k = \frac{\Phi_k}{\alpha(1 + |k|\lambda_d)} \tag{8}$$

in which λ_d is a *screening length* given by

$$\lambda_d = \frac{E}{2\alpha(1-\sigma^2)} \approx \frac{Ead^2}{\Delta F}\,. \tag{9}$$

The response to the force Φ is dominated by surface pinning on long scales $k\lambda_d \ll 1$, while bulk elasticity is dominant on short scales, $k\lambda_d \gg 1$. In a purely elastic model we would expect $\Delta F \approx Ea^2$ and consequently $\lambda_d \approx d^2/a$: the screening length is much longer than the distance between asperities. Intrinsic multistability due to adhesion (see below) should lead to smaller ΔF and consequently to larger λ_d.

Note that the concept of a screening length is entirely due to the elastic restoring coefficient α, which in turn results from the hysteretic behaviour. *It entirely disappears in the dynamic regime*, in which the net force F is independent of the position ρ (the probability $P(\rho)$ extends to the spinodal limit for all active asperities). The average F is unaffected by the extra displacement u and only the bulk restoring force remains.

In the static case screening provides a clear answer to the question raised in Section 1.4 of a *rigid* recoil of the slider when F is suppressed. In order to make the argument as simple as possible, let us assume that the

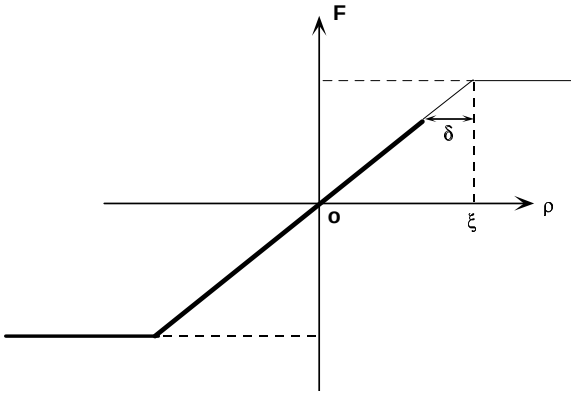

Fig. 6. Piecewise linear approximation of the force-displacement characteristic: the thick part corresponds to occupied configurations and δ is the recoil starting from the dynamic regime.

force-displacement characteristic of a given asperity has the piecewise linear shape of Figure 6: the cycle half width ξ is the spinodal limit beyond which dynamic friction is resumed (the slope discontinuity was washed out in Figure 5 because of the average over cycle widths). When that asperity recoils by δ the pinning force is

$$F = -\gamma(\xi - \delta).$$

The macroscopic restoring interfacial stiffness coefficient per unit area is $\alpha = n\gamma$. Assume now that α and ξ are modulated in space about average values α_0, ξ_0, with characteristic amplitudes α_1, ξ_1 and a wavelength λ. Rigid recoil would mean a residual pinning force of amplitude $F_1 \approx \alpha_0\xi_1$. Taking that as our starting point, we are back at our previous problem: when $\lambda \ll \lambda_d$ bulk elasticity precludes response to F_1 and recoil is rigid. If on the other hand $\lambda \gg \lambda_d$ the system prefers to adjust locally pinning to 0, which means that recoil adjusts locally by developing a modulation $\delta_1 \approx \xi_1$. The screening length λ_d marks the crossover between *rigid* behaviour at short scales and elastic response at large scales. (That issue of inhomogeneous recoil in static equilibrium was considered recently by de Gennes [11]).

1.6 Temporal effects

Up to now time never appears: we assume everywhere quasistatic macroscopic equilibrium. Dissipation occurs if we introduce viscous friction in the response to the pinning force: that occurs only, in our approximation of negligible small scale dissipation, at high velocities, comparable with those of elastic waves, that we do not consider. In the opposite limit of very low velocities, plasticity comes into play: that is another (essential!) story which we do not touch. In our simple model, time dependence may affect hysteresis in two ways.

- Activated jumps above the potential barrier *before* the spinodal limit. That problem was considered in [4], where we concluded that thermal agitation was too small to be significant at a micronic level. We do not repeat the argument: they lead to a slow logarithmic growth of the friction coefficient (the area of the actual hysteresis cycle increases with velocity v).
- Delay in the spinodal bifurcation due to the finite sweeping velocity. That too was considered in [4]: it leads to a power law growth of the cycle area ($\approx v^{2/3}$) and it may be viewed as the smooth continuation of activation beyond the spinodal limit.

Both effects are negligible on the global scale of a micronic contact, but they must be kept in mind, as far as they may become significant at a smaller scale, *e.g.* that of a boundary lubrication layer at the interasperity interface.

1.7 Where does multistability comes from?

In [4] we ascribed multistability to the elastic response to vertical compression only. The bare potential $V(\rho)$ was assumed monostable, say repulsive: the solution to equation (3b) was obtained graphically as shown in Figure 7. If the elastic λ is large ("hard" materials) the problem only has one solution. If instead $(\lambda + V'')$ can become negative a multistable region develops, leading to hysteresis. While possible in principle such a mechanism may prove unlikely once numbers are put in. If the pinning V is due to elastic forces, both V'' and λ are of order Ea and everything is possible. Indeed a "brushlike" geometry in which asperities would be high and narrow would undoubtedly lead to discontinuous spinodal jumps! Unfortunately such a situation is artificial: asperities are rather in the opposite limit, *i.e.* low and broad. Put another way the relative radius of curvature R of the two facing asperities is much larger than the radius a of the contact area (the slope of the profile is small: see Fig. 8). The physics can be understood by first considering that contacts are Hertzian. Then, under the normal load F_n, the radius a and the vertical elastic retraction h are qualitatively related according to [12]

$$a \approx \sqrt{hR} \quad F_n \approx E\sqrt{Rh^3}. \tag{10}$$

The maximum pinning energy V_m is of order $F_n\, h$, and the corresponding curvature V'' is accordingly $\approx V_m/a^2 \approx F_n/R$. Altogether we find

$$\frac{V''}{\lambda} \approx \frac{h}{R} \approx \frac{a^2}{R^2} \ll 1 \tag{11}$$

It follows that a nearly flat Hertz contact, such that $a \ll R$, will always be monostable[3]. This conclusion holds even more strongly when compressive plasticity is taken into account, since it results in further smoothing of asperity profiles. In order to produce multistability at the single asperity level we need to introduce adhesion into the picture.

[3] We are grateful to Prof. K.R. Johnson for drawing our attention to that feature which we had not appreciated.

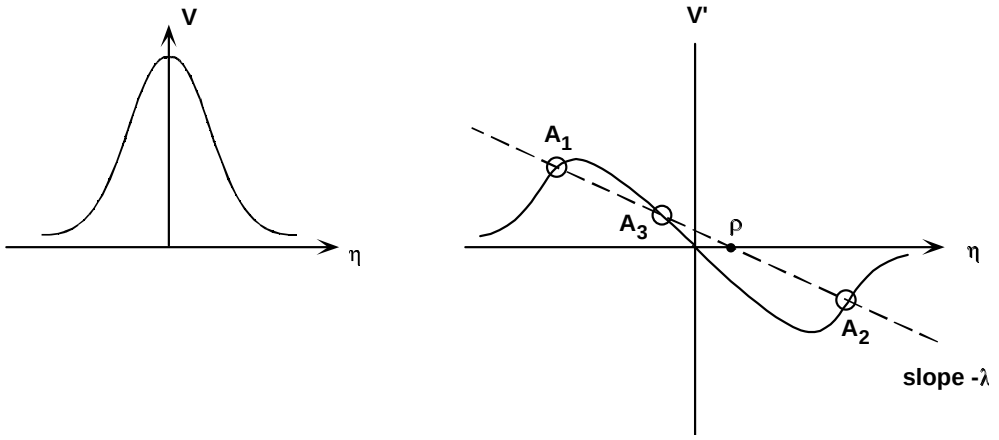

Fig. 7. A graphical solution of the contact equilibrium equation (3b) for the purely ealstic model. If the slope λ of the straight line is small, a range of multistability exists: A_1 and A_2 are locally stable, A_3 is unstable.

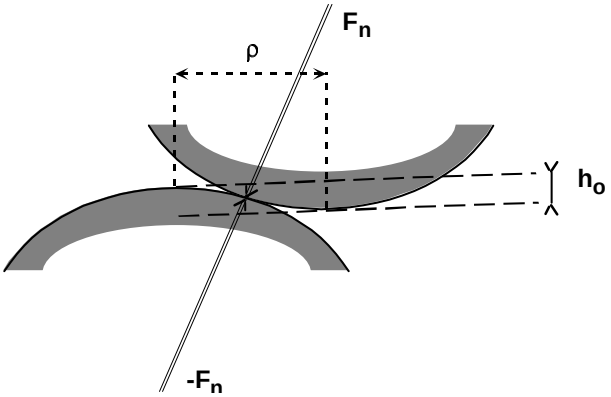

Fig. 8. A sliding Hertz contact, shown when the asperities just touch ($h = 0$). When $\rho = 0$, the elastic retraction is h_0.

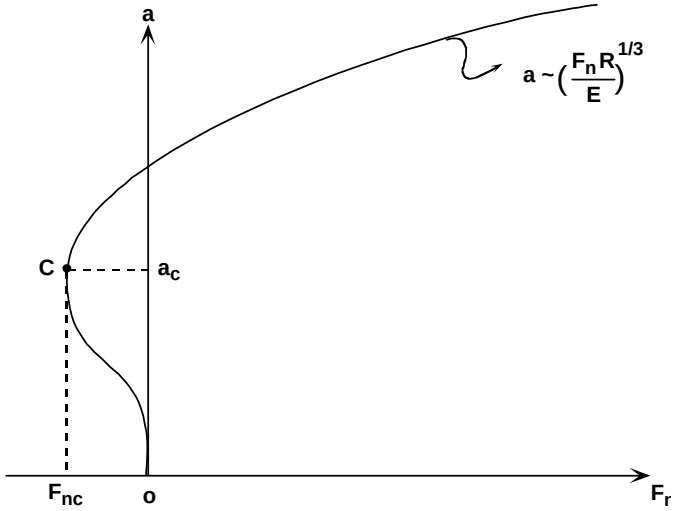

Fig. 9. The radius a of a JKR adhesive contact under a fixed normal load F_N. The critical point C corresponds to $a_c^3 \approx \gamma R^2/E$, $F_{Nc} \approx -\gamma R$.

In practice what matters is the global "mechanical characteristic of the asperity" $U(\rho)$, wherever it comes from. An intrinsic multistability of the bare $V(\rho)$ would work just as well. A simple example is that of a Jonhson-Kendall-Roberts (JKR) contact [12,13] in which the Hertz elastic response is corrected by a surface *adhesion energy* γ per unit area. The solution found in textbooks corresponds to a simple "normal" geometry in which the two asperities have a common vertical axis, with a normal load F_n. That problem does display hysteresis, as shown in Figure 9 in which the radius of contact a is plotted as a function of F_n. For negative loads (*i.e.* tractions), there exist two solutions for a given F_n: the slider can either be free or stuck by adhesion. Beyond a critical F_{nc} sticking breaks – conversely when the two asperities touch at $F = 0$, they suddenly stick together: in both cases the discontinuous evolution radiates a bunch of phonons leading to dissipation. As such, the above JKR model does not apply to our friction problem, in which one asperity is shifted sideways with respect to its partner by an amount ρ. If however we make the somewhat unrealistic assumption of a contact with no resistance to shear, then the net force is always normal to the common axis of revolution as shown in Figure 8. The JKR picture is simply rotated, the normal retraction h being a function of ρ fixed by geometry

$$h(\rho) = h_0 \left[1 - \frac{\rho^2}{2R^2}\right] . \tag{12}$$

It is a straightforward exercise to convert the curve of Figure 9 into a curve for the (elastic + adhesive) total energy U *versus* h (Fig. 10): we clearly see the usual spinodal cusp C at which the contact breaks. The other discontinuity occurs at O where sticking sets in (the mathematical singularity is different). When sweeping h one follows the trajectories marked by arrows. From $U(h)$ we infer $U(\rho)$, which has *two* hysteresis loops, corresponding to breaking the contact on either side. The $U(\rho)$ picture for one of them is shown in Figure 11: the critical points C and O

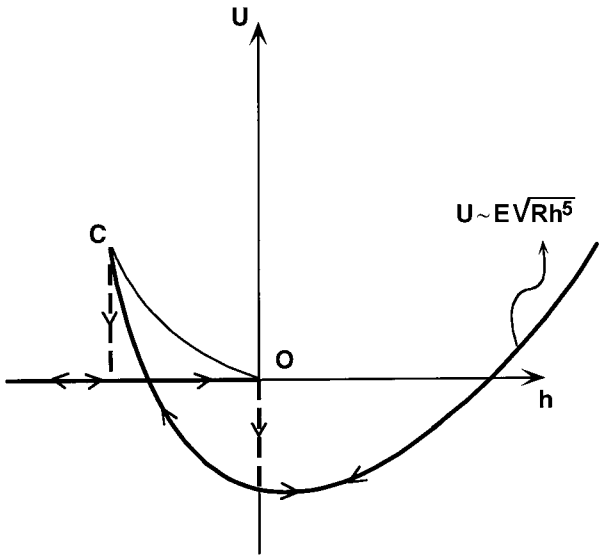

Fig. 10. Energy U of a JKR contact *versus* elastic retraction h. The spinodal cusp C corresponds to $h_c \approx -(\gamma^2 R/E^2)^{1/3}$. The OC branch is locally unstable. Actual backward and forward trajectories are shown as thick lines.

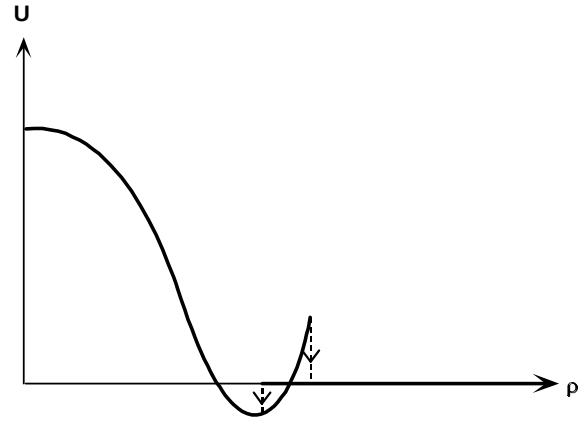

Fig. 11. Energy of the JKR contact U *versus* reference configuration coordinate ρ.

are still there. Due to the neglect of shear stresses such an estimate is caricatural[4] – but it clearly shows that hysteresis with a cusped $U(\rho)$ is common practice.

Whether one or two loops are present makes no difference: we assume one: $F(\rho)$ is the input. However, even if elasticity is not the source of hysteresis, we cannot forget it, as it necessarily comes in when interactions between distant asperities are included. Expressing $F(\rho)$ in terms of an underlying elastic picture allows us to use a *unified* language for single asperities and for their interactions. In a qualitative discussion we feel that such an economy of language is important: we therefore choose to pursue our elastic description, knowing that it is not realistic at the one asperity level – it just provides a simpler formulation.

2 Elastic interactions between asperities

We now consider a distribution of active asperities with random positions R_i (two dimensional) and configurations ρ_i (reduced to a one dimensional crude model). Their density per unit area is n. The pinning force acting on asperity "i" induces an elastic displacement at the position R_j of asperity "j", thereby shifting ρ_j and modifying the "j" pinning force – hence an elastic mediated long range interaction, described by the total energy:

$$U = \sum_i V(\rho_i + u_i) + \frac{1}{2}\sum_{ij} \lambda_{ij} u_i u_j. \tag{13}$$

Mechanical equilibrium is achieved when the energy is minimal:

$$V'(\rho_i + u_i) + \sum_j \lambda_{ij} u_j = 0. \tag{14}$$

λ_{ij} is an elastic stiffness matrix, which relates the displacements u_i of point asperities to the *local* forces F_i needed in order to achieve these u_i:

$$F_i = \lambda_{ij} u_j. \tag{15}$$

Note that the definition of λ is hybrid: we fix the *displacements* at discrete asperity positions, but, due to the very definition of the contact geometry, the surface *force* is zero everywhere between the asperities. A much more natural quantity is the inverse (compliance) matrix μ such that $u_i = \mu_{ij} F_j$: then the force is fixed everywhere, finite at discrete asperities and zero in between. The value of μ_{ij} is given in standard textbooks

$$\mu_{ii} \approx \frac{1}{Ea} \quad \mu_{ij} \approx \frac{1}{Ed_{ij}}\,. \tag{16a}$$

For a small system of dilute asperities one can easily invert the matrix, with the result

$$\lambda_{ii} \approx Ea \quad \lambda_{ij} \approx -\frac{Ea^2}{d_{ij}}\,. \tag{16b}$$

However extrapolation to large sizes is dangerous, due to the long range of elastic forces: the inversion of λ is non

[4] A much more sophisticated description, including shear effects, of the failure of a sheared JKR contact with an assumed Coulomb microscopic friction law has been elaborated by Savkoor [14], and discussed in [2]. Our spinodal discontinuity should then be viewed as a simplified representation of the fracture event.

trivial and whenever possible we will work with μ which is completely unambiguous[5].

The actual position of asperity i is $\eta_i = \rho_i + u_i$ and the corresponding pinning potential is $V(\rho_i)$. Force equilibrium at site i implies:

$$V'(\eta_i) + \lambda_{ij}(\eta_j - \rho_j) = 0. \tag{17}$$

Solving (17) means diagonalizing of a non linear random $N \times N$ problem, a fearful task that implies expansions and a configuration average on the position R_i of asperities – or equivalently on their mechanical characteristic $F = -V'$. In what follows we address a few salient questions at a very naive level of sophistication.

2.1 Total elastic displacement at a given asperity

It results from the response to all pinning forces:

$$u_i = \sum_j \mu_{ij} F_j = -\sum_j \mu_{ij} V'(\eta_j). \tag{18}$$

Strictly speaking a given u_k reacts on all the η_j and the solution of (18) is by no means trivial: let us try an expansion in powers of site interactions. In zeroth order we retain only the term $i = j$ in (18), thereby defining a net coordinate η_{i0} such that

$$\frac{d\eta_{i0}}{d\rho_i} = \frac{\lambda_{ii}}{\lambda_{ii} + V''(\eta_{i0})}\,. \tag{19}$$

In first order the force on other sites j would create an extra displacement $\delta u_{i0} = \mu_{ij} F_j$ at site i if there were no pinning. But the pinning potential $V(\eta_i)$ opposes that displacement: δu_{i0} is equivalent to a shift $\delta\rho_i$ of the *nominal* position, distinct from the real shift $\delta\eta_i$. The latter is obtained from (19), and consequently the net elastic displacement due to other sites is *in first order*

$$\delta u_i = \left[\frac{V''(\eta_{i0})}{\lambda_{ii} + V''(\eta_{i0})}\right] \sum_{j\neq i} \mu_{ij} V'(\eta_{j0}). \tag{20}$$

Since ρ is our control parameter it is simpler to describe the effect of interactions in terms of its shift at site i:

$$\delta\rho_i = -\sum_{j\neq i} \mu_{ij} V'(\eta_{j0}). \tag{21}$$

The summation in (21) involves two features:

- an integration over the *position* R_i of the active asperities, multiplied by their density n: that takes care of counting;
- an average over their *nominal* configurational coordinates ρ_j, which are statistically independent while the η_j are correlated by elastic interactions.

The two steps can be lumped together in the form

$$\sum_j = \nu \int d^2R_j \int d\rho_j \tag{22}$$

where $\nu \approx n/a$ is the sweeping rate of asperities introduced in dynamic friction. Equations (21–22) yield the average correction $\delta\rho_i$ which is of course supplemented by fluctuations that must be calculated separately.

Let us first assume that all sites are *individually monostable*: then the integral

$$\int_{-\infty}^{+\infty} V'(\eta_{j0}) d\rho_j = \int_{-\infty}^{+\infty} V'(\eta_j) \frac{\lambda_{jj} + V''(\eta_{j0})}{\lambda_{jj}} d\eta_{j0} \tag{23}$$

is identically zero: there is no single site average friction force and consequently no elastic deformations induced by asperity interactions. That was to be expected! But it is only true on the average: what about statistical fluctuations? In order to answer that question we calculate:

$$\delta\rho_i^2 = \sum_{j,k\neq i} \mu_{ij} V'(\eta_{j0}) \mu_{ik} V'(\eta_{k0}). \tag{24}$$

Since the η_0 are uncorrelated only the $i = k$ terms contribute: we thus find

$$\delta\rho_i^2 = \nu \int d^2R \int_{-\infty}^{+\infty} d\rho \left[\frac{V'(\eta)}{ER}\right]^2. \tag{25}$$

The average over ρ is non zero, of order F^2/E^2, and altogether ($\nu \approx d^{-2}$):

$$\delta\rho_i^2 \approx \int d^2R \left[\frac{F}{ERd}\right]^2. \tag{26}$$

The integral diverges logarithmically, a direct consequence of long range elastic interactions. It is cut off at $R \approx d$ at

[5] It could be tempting to argue that the elastic energy should be translationally invariant, depending only on *differences* $(u_i - u_j)^2$. That such an argument is wrong is obvious if we consider two asperities: according to it, the elastic energy would be $\approx Ea^2(u_i - u_j)^2/d_{ij}$. Actually the energy is $\approx Eau^2$ even if they move by the same amount u, because they move relative to a fixed center of mass (or reference surface at infinity: see Landau). If $u_i = -u_j$ the center of mass does not matter – still the estimate is wrong as it yields an energy too small by a factor $\approx a/d_{ij}$ (the energy is dominated by the near field of impurities). The reason for the paradox may be traced to the hybrid boundary conditions mentioned earlier. For simplicity assume a scalar dimensionless relation where the force is controlled *everywhere*:

$$u(r) = \int d^2r' \frac{f(r')}{|r - r'|}\,.$$

The 2d Fourier transform is straightforward, $u_k = 2\pi\, f_k/k$, yielding at once $f_k = ku_k/2\pi$ (translational invariance is reflected in the fact that $f_0 = 0$). Inverting the Fourier transform yields $f(r)$ when the displacement u is controlled *everywhere*:

$$f(r) = \frac{1}{(2\pi)^2} \int d^2r' \frac{(u(r) - u(r'))}{|r - r'|^3}\,.$$

Then it is true that only the difference $(u - u')$ enters. But this does not hold for our system of discrete contacts in which f is zero in between.

short distances, and at a typical slider dimension L at large distances (often L will be the *thickness* of the slider beyond which the R^{-1} elastic law is cut off). In the end the variance $\delta\rho_i^2$ is $b^2\mathrm{Log}(L/d)$, with $b \approx [F/Ed]$; for a purely elastic pinning model the force F is $\approx Ea^2$ and $b \approx a^2/d$.

We will return to these fluctuations in a moment. Before that, we consider the opposite case of individually *multistable* asperities, in a dynamic situation where the metastable branch is occupied up to the spinodal limit. Then the average force $F_j = -V'(\eta_{j0})$ is non zero (it is responsible for friction!). When carried into (21) it yields a *linearly divergent* displacement $\delta\rho_i$ – an expected result as the macroscopic friction force shears slider and base. Friction fixes their shear deformation u_{zx} and consequently the relative displacement at the interface is inversely proportional to the thickness H, as signaled by the linear divergence. Keeping track of such a singularity (*i.e.* working with a large but finite H) is of no interest for our present purpose: we choose instead to refer *the nominal position of asperities to its average value, as modified by macroscopic stresses*. The average value of $\delta\rho_i$ is thus eliminated from the outset, shear being built in the model. We are only left with fluctuations, which are measured from that average value. The calculation of the variance is then unchanged, the only difference being the replacement

$$\overline{F^2} \Rightarrow \overline{F^2} - \bar{F}^2.$$

Fluctuations retain their logarithmic divergence.

At that stage we should ask two questions:

(a) How good is the first order calculation? Higher terms would describe processes in which site j perturbs site k, which in turn affects site i: does that change the physics? Since we decided to ignore the average $\delta\rho_i$ by an appropriate choice of origin, the issue if of no importance here: we postpone it to the next paragraph.

(b) What is the physical meaning of the logarithmic divergence of fluctuations? We could be tempted to claim that a shift of ρ_i disappears in the configurational average! In fact such a claim is incorrect. One should indeed remember that the contact landscape is entirely renewed when the slider moves by an amount $\delta\rho \approx a$ (the old set is replaced by a new one, uncorrelated with its predecessor). Consequently $\delta\rho_i$ can move in such a small displacement all the way from one side of the statistical distribution to the other side. That does not matter much if $\delta\rho_i \ll a$, but it has drastic effects if $\delta\rho_i$ is comparable to the asperity size a, *i.e.* if the fluctuations can sweep the contacts across the spinodal discontinuity. That happens if the slider and base lateral sizes reach a characteristic value λ_L:

$$\lambda_L \approx d \exp\left[\frac{a^2}{b^2}\right]. \tag{27}$$

λ_L is the "Larkin length" [15] which marks the onset of a collective fluctuation regime. On scales $\ll \lambda_L$ interactions play a minor role while they are dominant on scales $\gg \lambda_L$. Note that λ_L depends exponentially on $[Ead/F]^2$: it is astronomically high for dilute asperities and usually completely irrelevant for multicontact interfaces between macroscopic conforming bodies[6]. It becomes essential when asperities are concentrated, $d \approx a$, as may be the case for contact between non conforming bodies (*e.g.* ball on plane) under high enough normal loads – a situation for which our present approach is not appropriate.

Comparing $\delta\rho_i$ to a is a handwaving argument, hardly acceptable when it controls a large exponent. We must therefore turn to more precise questions, accurately posed: the first of these is the influence of interactions on otherwise monostable asperities.

2.2 Interaction induced multistability

An equilibrium configuration is stable if its energy is a local minimum: that implies a positive definite second order variation in (13), *i.e.* a positive definite matrix:

$$A_{ij} = \delta_{ij} V_i'' + \lambda_{ij}. \tag{28}$$

The spinodal limit is marked by a zero eigenvalue of $\mathbf{A}$, *i.e.* by the condition $\mathrm{Det}[\mathbf{A}] = 0$. We can rewrite $\mathbf{A}$ as $\boldsymbol{\lambda}[\mathbf{1} + \boldsymbol{\mu}\mathbf{V}'']$: $\mathrm{Det}[\mathbf{A}]$ is a product of two determinants. $\mathrm{Det}[\boldsymbol{\lambda}]$ is uninteresting as it does not involve pinning (it has a trivial zero eigenvalue when translational invariance holds. Here, since the center of mass is held fixed, its lowest eigenvalue $\approx 1/L$, where L is the slider lateral size). All the physics lies in $\mathrm{Det}[\mathbf{1} + \boldsymbol{\mu}\mathbf{V}'']$, which involves the well defined matrix $\boldsymbol{\mu}$.

We can look at the problem differently. Assume that the slider moves by $d\rho$: all the ρ_i increase by the same amount. Let us calculate the resulting increase in one of the actual coordinates η_i. From (14) we infer at once

$$\eta_i + \sum_J \mu_{ij} \mathbf{V}'(\eta_j) = \rho_i. \tag{29}$$

Upon derivation we obtain

$$\frac{d\eta_i}{d\rho} = \sum_j [\mathbf{1} + \boldsymbol{\mu}\mathbf{V}'']^{-1}_{ij}. \tag{30}$$

The same matrix $\mathbf{B} = [\mathbf{1} + \boldsymbol{\mu}\mathbf{V}'']$ enters: here we must invert it. A typical matrix element of $\mathbf{B}^{-1}$ involves a minor divided by $\mathrm{Det}[\mathbf{B}]$. If the minor vanishes η_i goes through an extremum but it remains a single valued function of ρ. If on the other hand $\mathrm{Det}[\mathbf{B}]$ vanishes $d\eta_i/d\rho$ goes through infinity. Let us consider these two possibilities separately. When backward returns of η_i are possible, the "deterministic noise" due to other asperities may drive site "i" across its stability limit several times. If such is the case, it will jump "prematurely" the first time it can. Thereafter it remains on the stable branch, except if the cycle is so narrow that noise can also induce backward jumps (that would act

[6] In this situation, $(d/b)^2$ is at least on the order of the ratio of the apparent to the real areas of contact, *i.e.*, for typical normal load levels, at least 10^3.

to increase friction, but it is unlikely for dilute asperities). Early jumps are absorbed in the configurational average and therefore do not yield any interesting effect. In contrast a zero in Det[**B**] signals a spinodal limit: we recover the same criterion for multistability as was obtained from the local stability condition.

So, let us consider the case of N asperities which are individually monostable, and ask the question of whether interactions can lead to collective multistability. Unfortunately the problem cannot be solved exactly: we cannot calculate Det[**B**] without a configurational average, and the averaged determinant by itself has no physical meaning (as will be shown later). We thus must reduce our ambition: we will once more expand in powers of the interaction. Let us assume that $\mu_{ii} = Ea$ is the same for all asperities, and set $\mu_{ij}/\mu_{ii} = a/d_{ij} = \varepsilon_{ij}\,(i \neq j)$. We likewise define a dimensionless pinning $v = V/Ea$. The matrix **B** is given by:

$$B_{ii} = 1 + v_i'', \quad B_{ij} = \varepsilon_{ij} v_j''. \tag{31}$$

Since the diagonal terms of **B** dominate we separate them out by writing $\mathbf{B} = [\mathbf{1} + \mathbf{D}]\mathbf{C}$, with:

$$C_{ij} = [1 + v_i'']\delta_{ij} \quad D_{ij} = \varepsilon_{ij}\frac{v_j''}{1 + v_j''} \tag{32}$$

(**D** is off diagonal). The inverse matrix is

$$\mathbf{B}^{-1} = \mathbf{C}^{-1}[\mathbf{1} - \mathbf{D} + \mathbf{D}^2 + ...]. \tag{33}$$

In first order we find

$$\frac{d\eta_i}{d\rho} = \frac{1}{1 + v_i''}\left[1 - \sum_{j\neq i}\varepsilon_{ij}\frac{v_j''}{1 + v_j''}\right]. \tag{34}$$

The second term in the bracket describes the effect of interactions: we want to compare it to the first one. As before v_j'' is calculated at the *real* (zeroth order) position η_{j0} while the average involves an integration over the *nominal* positions ρ_i. It follows that

$$\int_{-\infty}^{+\infty}\frac{v_j''}{1 + v_j''}d\rho_j = \int_{-\infty}^{+\infty} v''(\eta_{j0})d\eta_{j0}. \tag{35}$$

The average effect of interactions on multistability vanishes in first order, just as the average correction $\delta\rho_i$ does. The advantage of (33) is that we can push the expansion systematically to higher orders: a typical term of the expansion will correspond to a "walk" from asperity to asperity, starting at site i and ending at site j. In order to achieve a non zero average every visited site (excluding the site under scrutiny i) must be visited more than once (actually an even number of times if the average restores the parity of $V(\rho)$). The simplest graph of that type is shown in Figure 12, yielding a contribution

$$-\frac{v''_i}{(1 + v_i'')^2}\sum_{j\neq i}\varepsilon_{ij}^3\frac{v''^2_j}{(1 + v_j'')^3}\,. \tag{36}$$

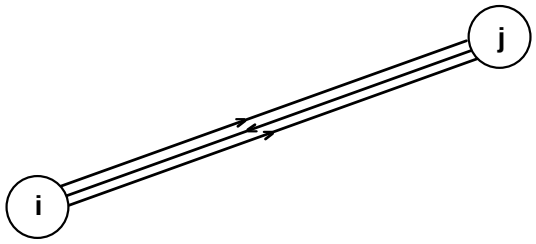

Fig. 12. The simplest non zero term in the expansion (19).

The average is finite – but the integral of $1/R_j^3$ coming from the ε_{ij}^3 term converges and it yields a small correction for dilute asperities (small ε limit). We conclude that *on the average* the interaction of dilute monostable asperities does not give rise to any induced multistability.

The situation is different when we take statistical fluctuations into account. We return to the second term X in the bracket of (34), which describes first order interactions, and we calculate its variance

$$\overline{X^2} = \sum_{j\neq i}\left[\varepsilon\frac{v_j''}{1 + v_j''}\right]^2 \tag{37}$$

(as before the cross terms $k \neq j$ do not contribute). The configuration average is now non zero

$$\beta = \int_{-\infty}^{+\infty}\left[\frac{v''(\eta_j)}{1 + v''(\eta_j)}\right]^2 d\rho_j = \int_{-\infty}^{+\infty}\frac{(v''(\eta_j))^2}{1 + v''(\eta_j)}d\eta_j \neq 0. \tag{38}$$

The space integration is again divergent and it yields a Larkin logarithm, $\overline{X^2} = b'^2\mathrm{Log}(L/d)$, where $b' = (2\pi v a^2\beta)^{1/2}$ is here calculated exactly. If $\overline{X^2}$ reaches 1, the first term of the series expansion of $d\eta/d\rho$ becomes comparable to the leading one. Although this does not *prove* that the expansion diverges, it can be considered as a strong hint that tails of the distribution may drive the asperity i unstable – put another way, that long range elastic interactions between asperities produce hysteresis on a scale larger than the Larkin length λ_L.

This can be made clearer by considering the exactly solvable case of two individually stable asperities 1 and 2, a distance d apart. Let us write the matrix μ as

$$\mu = \frac{1}{Ea}\begin{bmatrix}1 & \varepsilon\\ \varepsilon & 1\end{bmatrix} \tag{39}$$

where $\varepsilon \approx a/d$ is a small quantity. The condition for stability is

$$\begin{aligned} &1 + v_i'' > 0,\\ &\Delta = [1 + v_1''][1 + v_2''] - \varepsilon^2 v_1'' v_2'' > 0. \end{aligned} \tag{40}$$

Interaction is destabilizing when v_1'' and v_2'' have the same sign, in practice when both contacts are close to being individually multistable, so that there exists a range of ρ for

which $v_i'' \approx -1$. It follows that *interaction helps multistability*[7]. We expect the same to occur for the real N contact system – the smallness of the effect being compensated by the divergence of the integrals for large system sizes.

The main lessons learnt from the above discussion are

(a) Multistability arises from *fluctuations* and not from an average effect.
(b) The relevant length scale is the Larkin length:

$$\lambda_L \approx d \exp\left(\frac{d^2}{a^2}\right) \cdot \tag{41}$$

The harvest is modest, but reasonably under control. We would like to estimate the area of the resulting collective hysteresis cycle – *i.e.* the dynamic friction: it remains out of reach.

We emphasized the need for averaging well defined physical quantities. As an illustration let us try to average the determinant Det[**B**] directly. Up to second order in interactions it is diagonal except for a single 2×2 square: it follows that

$$\mathrm{Det}|\mathbf{B}| = \sum_{ij} \left[(1+v_i'')(1+v_j'') - \varepsilon_{ij}^2 v_i'' v_j''\right] \prod_{k\neq i,j} (1+v_k''). \tag{42}$$

We can carry the averages as before, and we do find logarithmic divergences: a Larkin length emerges similar to (22), but with a different coefficient β. One might argue that inside an exponent that is a significant difference. But it must be kept in mind that such estimates must not be taken literally, but only as signaling the order of magnitude of the length scale on which interactions become relevant.

2.3 Interaction of individually multistable asperities

The most naive effect is a change in the single site hysteresis cycle due to the other asperities. The effect does exist, but it implies a round trip from i to j and back: it is second order in ε_{ij} while we are primarily concerned with first order effects. Let us instead consider the response to a global translation $\delta\rho$. Due to the spinodal discontinuity the derivation that led to (34) is no longer valid. We must return one step earlier, writing in first order

$$\frac{d\eta_i}{d\rho} = \frac{1}{1+v_i''}\left[1 + \sum_{j\neq i} \frac{1}{ER_{ij}} \frac{dF_j}{d\rho} + ...\right] \cdot \tag{43}$$

[7] One can also calculate $d\eta/d\rho$

$$\frac{d\eta_1}{d\rho} = \frac{1+v_2'' - \varepsilon v_1''}{\Delta}.$$

A soft site 2 (close to multistability) can drive a stable site 1 ($v_1'' > 0$) backwards, but this does not matter as 1 is far from jumping.

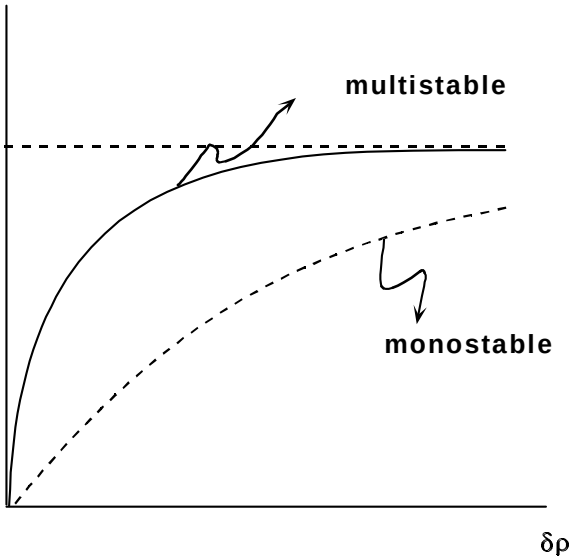

Fig. 13. The fluctuation amplitude of the force increment δF when a given asperity is moved by $\delta\rho$, respectively for a monostable and a multistable regime.

$dF_j/d\rho$ now has a discrete δ-function. That δ-function does not matter much when we average: whether F_j is continuous or not, its average is independent of ρ (translational invariance), and the average $dF_j/d\rho$ vanishes. We conclude that the *average* effect of elastic interactions on the shape $\eta(\rho)$ is identically zero, as it was in the absence of hysteresis. Technically, when the displacement $d\rho$ leads some sites j to jump, others climb the ascending part of the cycle and the net average distribution remains the same.

The situation is different when we consider the fluctuations: we cannot average the square of a δ-function. We must then calculate the statistical distribution of the force increment δF_j for a *finite* increment $\delta\rho$. Most contacts j are far from their spinodal limit: for them δF_j is proportional to $\delta\rho$, leading to a contribution $\approx \delta\rho^2$ to δF_j^2. A small fraction $\approx \delta\rho/a$ jumps upon the displacement $\delta\rho$, experiencing a finite *discontinuity* ΔF: their contribution to δF_j^2 is $(\delta\rho/a)\Delta F^2$. The resulting variance of δF as a function of $\delta\rho$ is sketched in Figure 13: the force and displacement scales are the same as in the monostable case (the force looses memory when $\delta\rho \approx a$), but the behaviour when $\delta\rho \to 0$ is different. That will result in the "cascade jumps" discussed in the next paragraph. Note that spinodal jumps are always *forward*, leading to a positive discontinuity ΔF.

For a single contact we ascribed dissipation (and consequently friction) to the energy radiated as phonons during the discontinuous jump. Starting from its original position u_a the asperity oscillates around its new equilibrium position u_b: that oscillation is eventually damped, mostly through radiation in the bulk (genuine viscous damping is negligible compared to "radiation damping" familiar in electrodynamics). The energy loss is the discontinuity in

the curve of Figure 2:

$$\Delta U = U(\rho_+, u_a) - U(\rho_+, u_b). \tag{44}$$

The situation is more complicated in the presence of interactions. When asperity i jumps it shifts the equilibrium elastic displacements of all other asperities j, even if they are far from their spinodal limit. Again u_j must go from u_a to u_b, hence an oscillation, radiation and dissipation. That must be summed over all sites j and the issue of convergence arises again, since $(u_b - u_a) \approx 1/R_{ij}$ decays very slowly. Fortunately u_b corresponds to a minimum of the energy as a function of u_j and consequently the energy radiated is of order $(u_b - u_a)^2$, decaying as $1/R_{ij}^2$ at large distances. Still a logarithmic divergence remains: how can we cope with it? We did not settle the issue but we feel the answer must be found in the finite response time of the spinodal jump. The period of the oscillation is typically c_s/a where c_s is the sound velocity; radiation damping of a point source is quick and the oscillation amplitude decays on the same time scale. That must be compared to the rate of arrival of signals from various jumping asperities i to a given j: the further away they come from, the more numerous they are. When the rate of arrival is comparable to the response time, one cannot treat the resulting elastic relaxations as independent processes. We should instead consider site j as subject to a random noise to which it responds globally. That should provide a cutoff in the logarithmic integration over R_{ij} and guarantee a well defined finite energy dissipation. We did not pursue the analysis in detail, mostly because a Log is never a dramatic singularity. But the problem is there and it deserves thought. We only draw one conclusion: the interaction between asperities *increases* dissipation since it creates secondary discontinuities at all sites.

2.4 The problem of cascade jumps

Consider a slider with lateral size L *smaller* than the Larkin length. Assume a primary site i jumps: its pinning force changes by ΔF and it creates on every other site j an extra forward displacement

$$\delta u_j = \frac{\Delta F_i}{E R_{ij}} \cdot \tag{45}$$

If site j is within δu_j from its spinodal limit, it jumps also – hence a cascade in which the primary jump induces the jump of P other sites. The total friction force changes by $P\Delta F$, a feature which is reflected in the noise of the total friction force F. One immediate question is whether that process is cumulative: site i triggers the jump of site k which would in turn trigger site j. In such a case one would have to solve the problem self consistently. We believe that it is not the case. The jump of intermediate sites k does produce an extra forward displacement on site j – but the effect is countered by the backward recoil of all the sites k which do not jump. *On the average* the effect of sites k cancels out, in much the same way as the average value of $d\eta_i/d\rho$ vanished (the average system is translationally invariant). Only direct cascades remain, in which the jump of j is directly triggered by i.

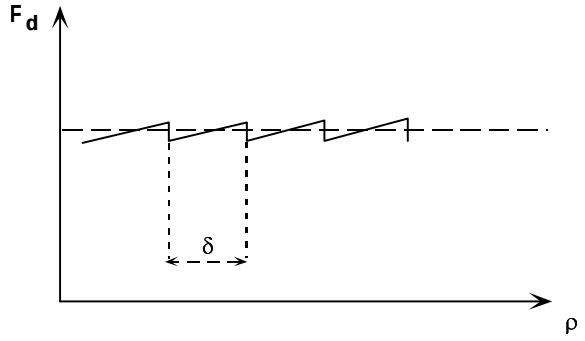

Fig. 14. A sketch of the noise in the macroscopic friction force.

The probability that a site j lies within δu_j from its spinodal limit is of order $\delta u_j/a$: the number of induced jumps is therefore

$$P = \int d^2 R_j n \frac{\delta u_j}{a} \approx \frac{\Delta F}{E d^2} \int \frac{2\pi R dR}{aR} \cdot \tag{46}$$

The integral diverges linearly: if L is a typical slider size we find that the cascade has an amplitude $P \approx L\,\Delta F/Ead^2$. If we remember that the static screening length is $\lambda_d \approx Ead^2/\Delta F$, we see that $P \approx L/\lambda_d$, a very simple result. (Of course the argument makes sense only if $P \gg 1$: otherwise averages are meaningless).

If P asperities jump together, that must occur P times less often. Let us first assume that all jumps are individual. When the slider proceeds by $d\rho$, $L^2\nu d\rho \approx L^2 d\rho/ad^2$ asperities are swept: the spacing between consecutive jumps is $\delta_0 \approx ad^2/L^2$. If jumps are collective that spacing becomes

$$P\delta_0 \approx \frac{ad^2}{L\lambda_d} \tag{47}$$

equal to a^2/L in a purely elastic model. The resulting behaviour of the friction force F has the sawtooth shape of Figure 14, the vertical amplitude of discontinuities being of order $L\Delta F^2/Ead^2$ (Ea^3L/d^2 for elastic pinning).

Cascade jumps are a spectacular consequence of elastic interactions between asperities. If that noise could be resolved it would yield precious information on the underlying pinning parameters.

3 Conclusion

The model of multicontact solid friction studied here is clearly schematic, as far as it neglects effects of paramount importance to real experimental systems – namely dissipative processes taking place on the smaller subcontact scale(s), related to plastic and/or viscoelastic slow relaxation. In particular, the question of (non linear) viscous intra-contact dissipation under shear forces on the order of the static threshold, which remains up to now elusive, is central for modelling stick-slip dynamics.

However, this schematic analysis permits to identify what we believe to be the basic physical ingredients of multicontact solid friction, thus providing a framework for more realistic future models.

Namely, solid friction, as defined by the existence of a static threshold and by finite dissipation at vanishingly small sliding velocities, implies *contact hysteresis*. The sudden asperity jumps occurring at the spinodal limits of the ranges of multistability of sliding contacts are responsible for the finite level of low velocity dissipation.

When the external force driving a sliding system is suddenly suppressed, the slider comes to static equilibrium by *recoiling* with respect to the track, the contact population is then shared between the two branches of the force-displacement hysteresis loop.

In this static regime, the interface exhibits a finite elastic stiffness. This reversible elastic range, which terminates at the static threshold (identical, for our simple model, to the dynamic friction force), connects smoothly the leftward and rightward sliding regimes[8]. In the static regime, contact hysteresis gives rise to the new phenomenon of *interface elastic screening*: when submitted to space dependent forces, the interface responds (*e.g.* recoils) *rigidly* on scales up to the screening length λ_d. This length, typically $\approx d^2/a$, is much larger than the average intercontact distance.

An interesting consequence of this unusual behavior concerns earthquake models of the Burridge-Knopoff type [16,17]. These models describe sliding faults in terms of a set of discrete internally rigid elastically coupled blocks; moreover, each interfacial block experiences a solid friction force described by a non linear phenomenological law. One of the questions about some of these descriptions is concerned with the fact that, as they contain no explicit length scale, their continuum limit is not well behaved [18]. This signals the existence of a physical cut-off, which has not been identified yet. We claim that this cut-off is the above mentioned elastic screening length. Indeed, interface rigidity on small scales precludes the possibility for regions of dimensions $< \lambda_d$ to start sliding *independently* while their neighbours would remain stuck.

On the other hand, as interface recoil adapts on length scales $> \lambda_d$, this length appears also as the maximum length of a rigid block. From this we conclude that, in principle, λ_d *is* the size of a Burridge-Knopoff block. Note, however, that for practical purposes, *e.g.* numerical investigations of the dynamics, it is legitimate to choose a larger size L, provided that L remains smaller than the space scale of the stick-slip instability. As argued by Rice *et al.* [17], this sets an upper limit L_R to L. One easily checks that L_R/λ_d is on the order of $(d\mu_d/d\mathrm{Log}V)^{-1}$, *i.e.* typically of order 10^2, thus confirming the validity of the Rice criterion for the choice of block sizes.

Finally, we have shown that elastic interactions, in spite of their long range, only play a minor role in solid friction of multicontact interfaces. Indeed, although they can in principle induce collective multistability of individually monostable contacts, this can only happen on scales larger than a Larkin length which is, for this problem, so huge as to be irrelevant in practice, at least under usual conditions. The contrast between this and the case of vortices [15] or charge density waves [18] stems from dimensionality: here we deal with point pinning centers coupled *via* 3D elasticity of the underlying bulk solids. The main non trivial effect of interactions – force noise induced by jump cascades – is expected to have a modest amplitude, and will probably remain difficult to unravel from other noise sources in real experiments. This justifies the tacit assumption implied by block models that noise on scales smaller than the block size itself is irrelevant.

These results lead us to conclude that the simplest version of our model – a large random set of independent centers experiencing a pinning and a shear elastic force and exhibiting individual multistability – provides an appropriate framework for the development of more sophisticated descriptions of real solid friction.

[8] Such a behavior can be paralleled with that of an ideally plastic solid – perfect elasticity below the yield stress, followed, for larger deformation, by plastic flow at constant stress.

References

1. F.P. Bowden, D. Tabor, *The Friction and Lubrication of Solids* (Clarendon Press, Oxford, 1950).
2. K.L. Johnson, *Fundamentals of Friction: Macroscopic and Microscopic Processes*, edited by I.L. Singer, H.M. Pollock, NATO ASI Series, Series E: Applied Sciences, Vol. 220, (Kluwer Acad. Publ., Dordrecht, 1992).
3. J.A. Greenwood, J.B.P. Williamson, Proc. Roy. Soc A **295**, 300 (1966).
4. C. Caroli, Ph. Nozières, *Physics of Sliding Friction*, edited by B.N.J. Persson, E. Tosatti, NATO ASI Series, Series E: Applied Sciences, Vol 311 (Kluwer Acad. Publ., Dordrecht, 1996).
5. A. Tanguy, P. Nozières, J. Phys. I France **6**, 1251 (1996).
6. M. Brillouin, *Notice sur les Travaux Scientifiques* (Gauthier-Villars, Paris, 1909).
7. G.A. Tomlinson, Philos. Mag. **7**, 905 (1929).
8. C. Caroli, B. Velicky, J. Phys. I France **7**, 1391 (1997).
9. T. Baumberger, Sol. State Comm. **102**, 175 (1997).
10. P. Berthoud, T. Baumberger, private communication.
11. P.G. de Gennes, C.R. Acad. Sci. Paris, Series IIb **325**, 7 (1997).
12. K.L. Johnson, *Contact Mechanics* (Cambridge Univ. Press, Cambridge, 1985).
13. K.L. Johnson, K. Kendall, A.D. Roberts, Proc. Roy. Soc. A **324**, 301 (1971).
14. A.R. Savkoor, Ph.D. thesis, Delft Univ. of Tech, Delft, 1987.
15. A.I. Larkin, Yu.N. Ovchinnikov, J. Low Temp. Phys. **34**, 409 (1979).
16. J.M. Carlson, J.S. Langer, B.E. Shaw, Rev. Mod. Phys. **66**, 657 (1994).
17. J.R. Rice, J. Geophys. Res. **98**, 9885 (1993).
18. P.A. Lee, T.M. Rice, Phys. Rev. B **19**, 3970 (1979).