

Visualizing stick–slip: experimental observations of processes governing the nucleation of frictional sliding

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2009 J. Phys. D: Appl. Phys. 42 214016

(<http://iopscience.iop.org/0022-3727/42/21/214016>)

[The Table of Contents](#) and [more related content](#) is available

Download details:

IP Address: 132.64.14.247

The article was downloaded on 01/03/2010 at 13:21

Please note that [terms and conditions apply](#).

Visualizing stick–slip: experimental observations of processes governing the nucleation of frictional sliding

S M Rubinstein, G Cohen and J Fineberg

The Racah Institute of Physics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel

E-mail: jay@vms.huji.ac.il

Received 8 January 2009, in final form 13 April 2009

Published 22 October 2009

Online at stacks.iop.org/JPhysD/42/214016

Abstract

Understanding the dynamics of frictional motion is essential to fields ranging from nano-machines to the study of earthquakes. Frictional motion involves a huge range of time and length scales, coupling the elastic fields of two blocks under stress to the dynamics of the myriad interlocking microscopic contacts that form the interface at their plane of separation. In spite of the immense practical and fundamental importance of friction, many aspects of the basic physics of the problem are still not well understood. One such aspect is the nucleation of frictional motion commonly referred to as the transition from static to dynamic friction. Here we review experimental studies of dynamical aspects of frictional sliding. We focus mainly on recent advances in real-time visualization of the real area of contact along large spatially extended interfaces and the importance of rapid fracture-like processes that appear at the onset of frictional instability.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

1.1. The frictional motion of rigid bodies

Systematic studies of friction date as early as 1699 when Guillaume Amontons conducted an experimental study of sliding blocks and gave the first modern formulation of the static friction law ($F_S = \mu_S \cdot F_N$). In 1785 Charles-Augustin de Coulomb repeated the experiments adding the first formulation of dynamic friction laws and suggested a model that partially accounts for them. However, the accepted explanation as to why the friction coefficient exists and is independent of the (apparent) contact area was introduced only in the middle of the 20th century in the micro-contact interface model (MCI) of Bowden and Tabor [14] in the 1940s. These authors were the first to note that due to surface roughness the real area of contact is composed of a multitude of micro-contacts formed by interlocking asperities and is only a fraction of the apparent contact area (for most surfaces the real contact area is as low as 0.1%). Bowden and Tabor noted that the *true* contact area, which is proportional to the normal load, is the relevant parameter that controls the frictional force. They also realized

that its value is not strongly dependent either on the roughness or geometry of the bulk material.

When a normal force is applied to two surfaces in contact, initially the highest asperities of both surfaces will come into contact to form a micro-contact. The initial pressure at those contacting asperities tip is enormous. This pressure deforms each micro-contact plastically, effectively forming a ‘welded’ junction. The plastic deformation of the junction causes its contact area to increase thereby reducing the resultant pressure until the interface is able to support the normal load. The resistance of these ‘welded’ junctions to shear accounts for the dry friction.

Although the fully plastic description invoked by Bowden and Tabor can explain Amontons–Coulomb friction, it is not the only explanation. Greenwood and Williamson (1966) developed a statistical approach to treat realistic surfaces [15]. In their model, they assumed that asperities could be characterized as elastic half spheres and considered the behaviour of an ensemble of asperities that are in contact with a flat surface. Although a single such (Hertzian) contact does not produce a linear relation between normal load and contact area, when the asperity heights follow a Gaussian distribution

instead of a single height, Greenwood and Williamson showed that the linear relation could be obtained. It was later shown [13] that a Gaussian distribution is not necessary and any sufficiently *wide* distribution of asperity heights would reproduce the linear dependence of frictional strength on the normal load. It is now well accepted that for most materials neither fully plastic nor elastic approaches alone are sufficient to describe frictional interfaces and, even for moderate loads, an elastic–plastic approach is needed [16].

By what means does an interface first yield to coherent motion in response to an imposed shear? Typically, a body in frictional motion either slides steadily or can undergo unsteady start–stop motion, coined ‘stick–slip’. Many experimental studies of friction [17–23] have been geared to either measuring the statistics of slip–stick events or studying the stability of the system as a function of the driving parameters.

In 1966 Brace and Byerlee [24] suggested stick–slip friction as the mechanism for shallow earthquakes. Following this, a large emphasis in the geo-physical community was placed on studying the physics of friction in general and rock friction in particular. This effort led to generalizations of ‘classic’ Amontons–Coulomb friction laws that are known as ‘Rate and State’ laws [25–27]. Considering two half spaces in a state of steady sliding at rate V_0 with uniform shear, τ , and normal, σ , stresses the Rate and State framework models the variations of shear strength as due to dependence on both the sliding velocity and the evolving properties of the contact population along the interface [25, 26]. These evolving properties are typically represented by one or more phenomenological ‘state’ variables [21, 26–28], which can be considered as representing, for instance, the contact ‘age’. These models have been very successful in providing a quantitative generalization of Amontons–Coulomb friction that describes a broad range of materials ranging from PMMA [29] and granite [19] to granular materials [30]. These laws accurately describe memory effects in response to variations in slip rate and normal stress [31–36] and describe a gradual transition to motion that effectively replaces the ‘immediate’ jump from static to dynamic friction inherent in Amontons–Coulomb friction. In addition, rate and state laws can accurately reproduce the different modes of frictional sliding together with the bifurcations between them. In this picture, however, frictional motion is still considered as the relative motion of two infinitely rigid bodies, where the *spatial* degrees of freedom along the interface of separation are *not* considered. Naturally, since the interface is pinned by the micro-contacts, no relative motion can occur before these contacts are ruptured. Recent laboratory experiments have demonstrated the importance of rapid fracture-like modes that appear at the onset of frictional instability to understanding the contact rupturing mechanism and the transition to sliding.

1.2. Crack-like dynamics

As fracture processes are important in order to fully characterize frictional dynamics, we present a brief review of a number of basic principles of fracture mechanics. In the field of fracture, interface rupture has mainly been considered a

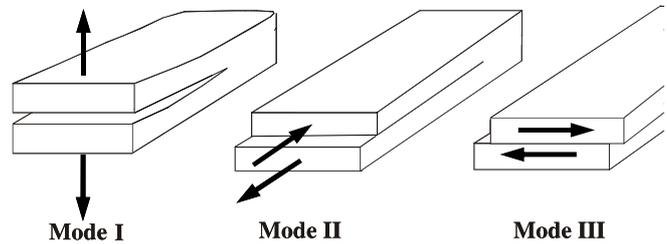


Figure 1. A schematic view of the three basic modes of fracture.

highly dynamic process, where the relevant velocities involved are in the range of the material sound speeds. Generally, the fracture energy is considered to be dominated by adhesion and deformation. If frictional contact between two surfaces is considered at all, it is thought to be a secondary effect.

Let us consider a crack within a brittle material. There are three basic modes of fracture that can occur [37]. These are shown schematically in figure 1. Mode I fracture occurs when the imposed stresses are purely tensile (i.e. normal to the crack faces), mode II fracture when shear is applied in the plane of the crack. Mode III fracture occurs when (out-of-plane) shear is applied in the direction within the fracture plane, but perpendicular to a crack’s propagation direction. For each mode of applied stress, a singularity of the stress field at the tip of a crack will result. The loading configuration of a mode II crack is analogous to that of a frictional system.

In a homogeneous material, it is generally believed that a crack will propagate only in the direction for which the mode II component of the stress at its tip is zero [38]. Thus, mode II fracture *cannot* occur in a homogeneous material. Along an interface, however, where the bonding strength is weaker than within the bulk material, mode II fracture can occur.

What is the fracture criterion for an interface crack, i.e. a crack that propagates along a weakened plane? If all dissipation takes place within the near vicinity of a crack’s tip, fracture mechanics provides a well-established criterion (the Griffith criterion [39]) whereby the energy per unit area, G (called the ‘energy release rate’), flowing into the tip of a crack must equal the energy per unit area, Γ (the fracture energy), needed to create two new fracture surfaces. This criterion forms the basis for the equation of motion of a single dynamic crack [37] that has been shown to be in excellent quantitative agreement with experiment [40]. In the case of an interface crack, however, dissipation is not confined to a finite region surrounding the crack tip but is rather distributed throughout the entire sliding part of the interface. Therefore formulation of a *Griffith like* criterion that accounts for the role of dry friction is much more complicated. Such a criterion, which includes the frictional characteristics of the interface, was suggested first by Palmer and Rice [41] and later by Andrews [42, 43]. By assuming that behind the tip of a singular mode II crack, on the fracture surface, shear stress is at its sliding friction level ($\mu_d \cdot \tau_N$) the losses due to frictional dissipation throughout the entire fracture surface could be added to the energy balance. This calculation resulted in a critical length, L_C , for a rupture to propagate dynamically. This model assumes that all of the nonlinear frictional behaviour is confined to a small region near the crack tip, where the transition to sliding occurs, and

that the shear distribution along the interface can be accurately represented by a simple superposition of two asymptotic fields. While this theory does not incorporate the complex memory and velocity dependence of the interface described by more realistic friction laws (e.g. rate and state friction), it does provide an important conceptual description of how frictional motion may initiate.

A more recent model was suggested by Brener and Marchenko [44, 45] to describe observations of localized rupture fronts observed within an interface formed by an aqueous gel and glass [46]. The authors incorporated friction by assuming that along sliding parts of the interface a linear viscous (Stokes friction) is valid. This model provided both a good qualitative description of the experiments and indicated that, for such a system, the Amontons–Coulomb threshold for static friction is equivalent to nothing but the Griffith threshold for dynamic crack propagation.

1.3. Shear fracture in experiments

Although brittle fracture in the case of mode I has been the object of considerable study [40, 47], experimental study of dynamic mode II fracture has been quite limited. In recent years, however, there has been a surge of interest in interface fracture. This interest has been motivated by three main factors (1) the need to both characterize and understand the failure modes of composite (multi-phase) materials [48] (2) experimental observation of interfacial intersonic fronts travelling at speeds beyond the Rayleigh wave speed (defined as the velocity of sound waves propagating along a free surface) [49] and (3) Brace and Byerlee's 1966 hypothesis [24] that stick–slip friction is the mechanism for shallow earthquakes.

A series of experimental studies aimed at investigating the rapid dynamics of stick slip were conducted by Johnson *et al* [50, 51]. A biaxial configuration was used to shear a quasi-two-dimensional interface of either granite–granite [50] or granite–dunite [51]. Relative displacement across the interface was measured by a single displacement sensor. In addition the authors also measured either the acoustic signature of rapid events by an array of piezoelectric acoustic sensors, or the shear stress field around the interface with solid state strain gauges that were coupled to the side of their samples. These authors reported that the transition to sliding is mediated by the passing of crack-like ruptures travelling at speeds comparable to and, in some cases, exceeding the shear wave velocity, V_S , of the bulk materials. They also observed that the sliding speed (particle velocity) is proportional to the overall stress drop and suggested that the frictional properties are well described by Coulomb friction. Interpreting the strain gauge data, the authors concluded that the transition from a static to a dynamic regime of friction occurs largely at the rupture front. This work was the first to show the crack-like nature of stick–slip in rocks; however, as their sampling rate was limited to 1 kHz, the authors were not able to investigate the details of the rapid transition to sliding.

Okubo and Dieterich [52] used a biaxial experimental system to study the dynamics of stick–slip on a $2 \times 0.4 \text{ m}^2$ granite interface. They measured both the differential

displacement and the shear stress at several locations along the interface at a 200 kHz sampling rate. Similarly to Johnson *et al* the authors found that slip is initiated by a large drop of shear stress. The stress drop nucleates within the interface and propagates outwards towards the edges. After the rapid stress drop, slip continues at a constant level of stress. Due to the high sampling rate, Okubu *et al* could investigate details of weakening by examining the finite slip that occurs during the stress drop. They found that the reduction in shear stress occurs over a finite slip distance, d_r . They noted that the *slip-weakening distance*, d_r , did not strongly vary with the normal load but increased significantly, from $\sim 5 \mu\text{m}$ to $\sim 30 \mu\text{m}$, when the smooth $0.2 \mu\text{m}$ (rms) surfaces were changed to rougher $80 \mu\text{m}$ (rms) ones. Roughness also had a significant effect on the rupture propagation speeds. While for the smooth interface, rupture appeared to rapidly accelerate towards speeds comparable to the shear wave speeds of granite, on the rougher interface rupture accelerated throughout the entire run, yet always remained significantly (30%) below the shear wave speed.

Ohnaka *et al* [53, 54] conducted a systematic experimental study of rupture dynamics during stick–slip in granite blocks. The authors mainly focused on the nucleation and transition to rapid events. Experiments were carried out on both a biaxial setup similar to the one used by Okubu and Dieterich (only smaller) and a double direct-shear failure apparatus in which two 290 mm long and 50 mm wide interfaces are formed when a granite block is sandwiched between two outer granite blocks. Tangential force on the frictional plane was applied by translation of the inner block. Using an array of strain gages and displacement sensors, they detected evidence of three stages of rupture: (1) a quasi-static phase where a weak nucleation patch is formed within the interface and continues to grow slowly as shear force is applied (They reported speeds of up to a few $\sim \text{cm s}^{-1}$). (2) an accelerating phase that is reached towards the transition to sliding with rupture speeds of up to $\sim 100 \text{ m s}^{-1}$ ($\sim 0.03 V_S$) and (3) a dynamic phase with high velocity ($\sim 0.4\text{--}0.8 V_S$) fronts propagating along the interface. Although the velocity measurements were rough (12 strain gages and six displacement sensors were placed at equal distances over a 290 mm long interface), this experiment provided solid evidence for the dynamic motion of a coherent fronts along a rough interface. This study showed clear evidence of slip-weakening behaviour and a critical rupture length where dynamic rupturing initiates. The existence of such a length scale was also suggested by Okubo *et al* [52].

The first real-time visualizations of dynamic interface rupture were performed by Tippur and Rosakis [55] on a bimaterial interface formed by bonding PMMA to steel or aluminium. These and consequent experiments [56–58] on bimaterial interfaces established that interface cracks can indeed propagate at velocities up to and beyond the Rayleigh wave speed (the limiting crack velocity for mode I cracks) of the 'softer' of the two materials surrounding the interface. Experiments [49] investigating interface fracture along an interface bounded by the *same* materials revealed that, here too, interface cracks are able to travel at speeds *beyond* the Rayleigh wave speed. Creating a weak interface separating

two sheets of Homalite-100 (a brittle polymer) by the use of a solvent, these experiments observed intersonic crack propagation along the interface (i.e. propagation at velocities between the shear and longitudinal wave speeds), where the crack's velocity was seen to approach the longitudinal wave speed of the material. More recently, the same group [59] reported visualization of a transition from sub-Rayleigh propagation to fronts moving beyond the shear wave velocity in a homalite/homalite interface under pure shear conditions. In all of these experiments, triggering of rupture was not spontaneous but rather caused by either projectile impact or an exploding wire placed in the interface.

These observed intersonic velocities appeared to be in contradiction to calculations, based on linear elastic theory [60], where intersonic speeds differing from $2^{1/2}c_s$ (the material's shear wave speed) would result in *negative* energy flux to the crack tip. In the above experiments it is interesting to note that the energy release rate at the crack tip, (measured by optical means) throughout the experiment appeared to *decrease* with increasing crack velocity. Molecular dynamic simulations [61, 62] and finite-element [63, 64] calculations of intersonic propagation indicate that these interesting effects can be understood as resulting from micro-crack nucleation *ahead* of the tip. Thus the total energy flux, which flows into both the main crack and micro-cracks, is much greater than that measured at the crack tip. An additional mode of interface rupture has been the subject of much recent interest. This mode of fracture is described by self-healing fracture fronts called slip-pulses. These pulses of spatially confined slip have been theoretically predicted both along bimaterial interfaces [65, 66] and in cases where strong velocity weakening friction and healing occur at the tip of an interface crack [5, 67, 68]. Indications of such rupture modes have been seen in seismic data [69] and experiments [23, 46] This intriguing topic is, however, beyond the scope of the current review.

2. Measurement of contact area

We have reviewed the importance of understanding the detailed processes within the frictional interface. In most experimental work, however, the evolution of the contact area composing the interface was not probed directly but rather deduced by studying stress or strain fields. Probing the interface *directly* is difficult. The reason for this is that, by definition, the real area of contact is *hidden* between two solid blocks and therefore, hard to directly access. Measurements of the average area of contact have been performed by studying the electrical conductivity through the interface [14] and its acoustic properties [70, 71]. However, these methods can only measure average quantities and cannot detect frictional variations throughout the interface.

Dieterich and Kilgore [72] were the first to directly visualize the actual points of contact. In this pioneering work, the authors used microscopy to directly observe the micro-contacts during the formation and sliding of a frictional interface formed between two transparent blocks of acrylic material. This study provided both the first direct visual confirmation of the Bowden and Tabor picture and offered

strong evidence relating the state parameter in the Rate and State laws to the average *age* of the real area of contact. In these experiments, the nominal area of the interface was relatively small ($15.8 \times 15.8 \text{ mm}^2$) and the field of view encompassed a few single contacts. This study, therefore, did not measure either collective modes or spatial variation in the contact population.

We now turn to recent work where the real area of contact was directly measured over a spatially extended interface [73–76]. Both the spatial and temporal resolutions in these studies enabled the investigation of such collective modes and how they couple to the complex evolution of the micro-contact population.

2.1. Real-time measurement of contact area: methods

A schematic diagram of the loading system used in the experiments reviewed in this section is presented in figure 2(a). The aim of the experimental system is to examine the short-time dynamics of long quasi-one-dimensional frictional interfaces by measurement of the net contact area. The interface is formed by two long and thin polymethylmethacrylate (PMMA) blocks, a slider of variable size ($(50\text{--}200) \times 80 \times 6 \text{ mm}^3$) and a base of constant size $270 \times 30 \times 30 \text{ mm}^3$ in the x (propagation), y (normal loading) and z (sample thickness) directions, respectively (see figure 2(a)). The slider contact face was diamond machined to a flatness of better than $0.1 \mu\text{m}$ (rms) and then roughened to a range of parameters $0.5\text{--}5 \mu\text{m}$ (rms). The base was also roughened to $1\text{--}5 \mu\text{m}$ (rms) thus fixing the overall roughness of the interface to be approximately square root average of both rms values. Roughness of the surfaces was characterized by measuring the height profile over a 15 mm length using a Taylor–Hobson (Talleysurf 3+) contact profilometer combined with a AFM surface scan over $5 \times 5 \text{ mm}^2$ area. These data were then used to calculate the RMS value of the height distribution we define as the roughness.

In the reviewed experiments normal loading, F_N , was applied to the slider via a spring array (with a $4 \times 10^5 \text{ N m}^{-1}$ stiffness) ensuring a uniform stress distribution. Shear forces, F_S were applied to the sample's trailing edge (see figure 2(a)) at a height, h ($2 \text{ mm} < h < 17.5 \text{ mm}$), above the interface plane. F_S was applied by pushing the slider edge via a rigid rod. The rod was sandwiched between the sample edge and one end of an S-beam load cell (stiffness $4 \times 10^6 \text{ N m}^{-1}$). The other end of the load cell was advanced at fixed rates ranging from 1 to 1 mm s^{-1} . The torque introduced by the rod was small, resulting in a maximal 1–10% variation of $F_N(x)$ over the length of the interface. Applied normal and shear stresses ranged from 0.2–6 and 0–3 MPa, respectively. Displacements of the slider edges were measured either by using a fibre optic displacement sensor (resolution better than $0.5 \mu\text{m}$) or by integrating the trailing edge load cell data (resolution of $2.5 \mu\text{m}$).

The experimental system was designed to perform simultaneous and rapid visualization of the net contact area, $A(x,t)$, over the entire spatial extent of the interface. A laser beam was expanded via a prism array into a $200 \times 5 \text{ mm}$ laser

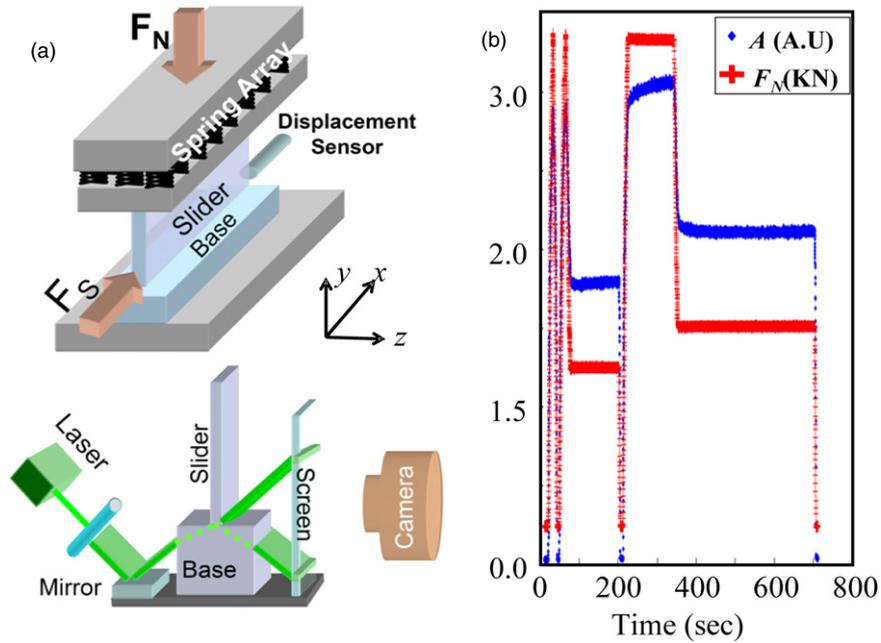


Figure 2. (a) A schematic view of the experimental system; (top) transparent slider and base blocks are loaded through a soft 40-spring array. F_N is applied at the trailing edge of the slider. Displacement is measured at both the leading and trailing edges of the slider. (bottom) A sheet of laser light illuminates the interface through the base. Both the transmitted and reflected sheets are imaged onto a fast camera. (b) F_N (red crosses) and the integrated intensity of light, hence the total net contact area, A (blue diamonds) are plotted versus time. A follows F_N , as predicted by the Bowden and Tabor theory of friction [14]. Note that under constant normal load A continues to evolve either increasing (ageing) or decreasing (de-ageing), depending on the loading history [74]. Both F_N and A are rescaled and presented in arbitrary units.

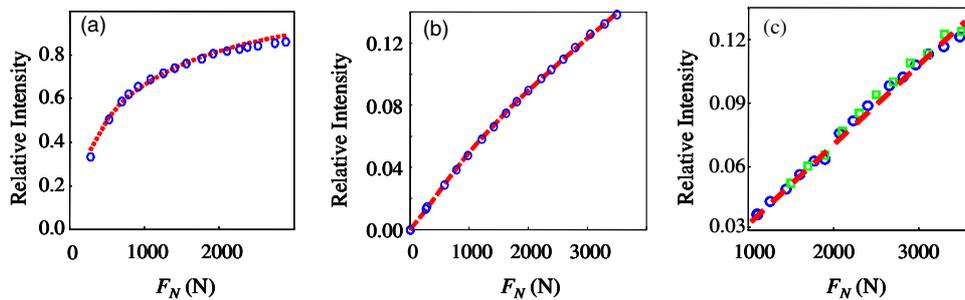


Figure 3. Relative light intensity transmitted through the interface as a function of the applied load for samples with an interface roughness of (a) 50 nm rms, (b) 500 nm rms and (c) $2\mu\text{m}$ rms. Measured values (circles) are compared with numerical values calculated for a fully plastically deformed statistically random surface. Note that for a sufficiently rough interface the transmitted intensity is proportional to A . Adapted from [75].

sheet. The entire interface was illuminated by the laser sheet through the base block at an angle significantly larger (angle of incidence of 70°) than the critical angle for total internal reflection from a PMMA/Air boundary at the interface. The laser light could transverse the slider–base interface either directly, at the micro-contact points between the two blocks, or by evanescent propagation, where incident light tunnels across the air gap between the two blocks. For large ratios of the roughness of the interface to the exponential decay length (~ 50 nm) of the evanescent light, one can (for the range of normal forces applied) neglect evanescent contributions to the intensity passing through the interface. In such cases, the light passing through the interface is proportional to the real area of contact (see figure 3(c)), where we consider ‘contact’ as separations of less than 50 nm. For roughness below $\sim 0.1\mu\text{m}$ rms, or higher values of F_N the evanescent field must be taken

into consideration [75] (see figure 3(a)). The transmitted light was directly focused onto a fast CMOS sensor (VDS CMC1300 camera), enabling rapid imaging of the transmitted intensity, $I(x, z, t)$. The sensor could be configured to frame sizes of $1280 \times N$ pixels with frame rates of $500\,000/(N+1)$ frames s^{-1} .

The data acquisition apparatus was designed to capture both slow processes at the quasi-static time scales governed by the loading rate and rapid, crack-like, processes whose entire duration falls in the sub-millisecond range. Acoustic signals, generated by rapid slip events, were used to trigger storage to disc of several thousand frames, which bracketed the acoustic signal. This enabled effectively continuous acquisition of light intensity data at maximal frame rates. In parallel, the evolution of the contact area was monitored by storage of single frames at fixed (20–100 ms) intervals throughout the duration of an experiment.

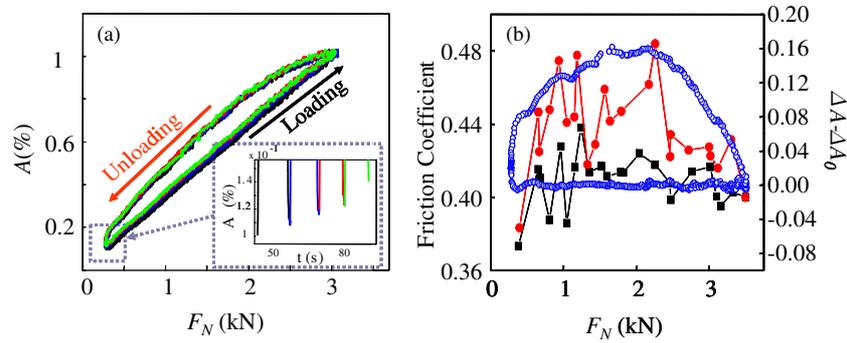


Figure 4. The real area of contact, A (as a per cent of the nominal contact area), as a function of F_N for four successive loading and unloading cycles. The samples were not separated between cycles and a minimum load of 270 N was retained. (inset) A , as a function of time at the minimal values of F_N , shows a very small ($<0.15\%$) accumulation of contact area between consecutive cycles. (b) The static friction coefficient, μ_S , is plotted versus F_N . μ_S is directly measured by shearing the system upon loading (black squares) and unloading (red circles), as in (a). The blue circles indicate the difference in contact area, ΔA , between loading and unloading that is derived from (a). The interface in these experiment has a nominal contact area of $150 \times 6 \text{ mm}^2$. Adapted from [74].

Determination of the propagation velocities of the different fronts was performed as follows. The intensity measurements were first averaged in the y direction to create a single array of intensities along the x direction for each frame. Using this data, a time series of intensities measured at each spatial point was obtained (see figures 2(a)). To each time series, 1D wavelet transforms (Haar and db4 kernels) were applied to determine the arrival time of each of the different fronts at each spatial location along the interface (e.g. figure 8(b)). Differentiation of this data then yielded the velocity of each front.

In figure 2(b) we plot measurements where values of F_N were varied and measured in parallel to $A(x)$. $A(x)$ was then integrated over x to give the total value of the contact area across the entire interface, A . A precisely followed the (fast) temporal variations in F_N . This is in accordance with the Bowden and Tabor theory for friction [14]. For $1 < F_N < 4 \text{ kN}$ the true contact area varied from 0.35–1.35% of the interface’s nominal contact area.

In figure 3(a) we present measurements of the light intensity passing through the interface as a function of the applied normal load for well polished surfaces (50 nm rms). The dotted lines show calculated transmitted intensities for the experimental conditions obtained by using the best fit values for the fully plastic deformation of a numerically generated rough surface. A fully plastic deformation model will yield a linear relation between A and F_N . However, for smooth surfaces and over large ranges of F_N the light intensity will not be proportional to A due to evanescently transmitted light. The transmission of evanescent light will, therefore, induce some curvature in $I(F_N)$. $I(F_N)$ best corresponds to the Bowden–Tabor picture. When the yield stress of the material is used as a fitting parameter, the best fit was obtained for an average stress of 150–400 MPa at micro-contact tips. This number is comparable to the penetration hardness strength measured by both Briscoe *et al* [77] and Dietrich *et al* [72] in similar materials. For rougher surfaces, where the evanescent decay length is small compared with the mean separation of the surfaces, the curvature in $I(F_N)$ diminishes (figure 3(b)) and the relation between transmitted light through the interface and F_N is linear, as shown in figure 3(c).

2.2. Real-time measurement of contact area: results

2.2.1. The dynamics of static friction.

The measurements in figure 3 show that the formation of true contact area during normal loading is well described by a fully plastic deformation model. As the contact area is proportional to the normal load, these results are consistent with both the Bowden and Tabor and Williams and Greenwood pictures, the former, however, was shown [75] to be more compatible with the material parameters describing our system. When the normal load is reduced, however, the results are somewhat surprising [74]. In a typical loading/unloading cycle, significant hysteresis of A is observed. Figure 4(a) shows four consecutive loading/unloading sequences in which F_N was cycled between the same minimum (270 N) and maximum loads. The loading curves retraced each other nearly perfectly, with less than a 1% residual memory between cycles (figure 4(a), inset). This suggests that a mechanism exists that essentially renews the contact area for each cycle. As figure 4(b) indicates, the change in A throughout each cycle is echoed by the static coefficient of friction, $\mu = F_S/F_N$; for each value of F_N , μ is measurably lower during the loading stage (squares) than in the unloading stage (circles).

The hysteresis curves exhibit scaling behaviour. Figure 5(a) presents five different loading/unloading cycles, where the maximum load was increased in each successive cycle. Between cycles, significant slip was imposed on the system, in order to erase all (small) memory of past loading. When F_N is normalized by the overall range, ΔF_N , applied in each cycle, all of the response curves superimpose nearly perfectly (inset to figure 5(a)). This data collapse suggests a scale-free mechanism, which is possibly due to the self-affine character of the rough interface.

When no slip is imposed on the system, long-term memory is retained. This is demonstrated in figure 5(b), where four successive loading cycles are shown in which the maximum load was constant but the minimum load points were successively decreased. This time, no slip was imposed between cycles and the system retained a memory of the maximum contact area reached. This memory effect is clearly

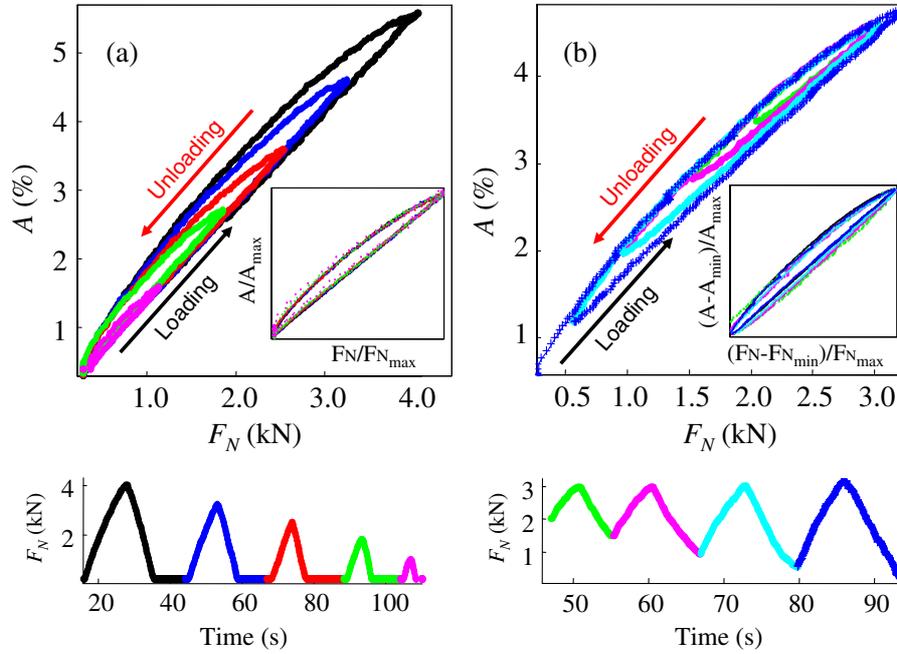


Figure 5. A (in per cent of nominal contact area) as a function of F_N (top) and F_N as a function of time (bottom) in successive loading or unloading cycles. (a) F_N is cycled from 270 N to decreasing maximal loads. Between cycles, 1 mm of slip was generated. (inset of (a)) When F_N is normalized by the overall range of the normal load applied in each cycle, ΔF_N , all curves collapse onto a single curve. (b) Similar to (a), but here the maximal value F_N remains the same in all cycles and successively lower minimal loads are applied. Here, no slip was imposed between cycles and long-term memory is retained. This is seen by the orientation of the reloading curve axis towards the maximal point achieved in the initial loading cycle. Here (inset of (b)) data collapse upon normalization by ΔF_N occurs only after rotation is performed to offset the axis orientation induced by long-term memory. The interface in these experiment has a nominal contact area of $30 \times 6 \text{ mm}^2$. Adapted from [74].

evident when the curves are, again, normalized by ΔF_N . In order for the data to collapse to a single curve, the hysteresis contours must first be rotated to a common axis because, upon reloading, the system *targets* the point of maximal load.

Figure 6 describes another series of loading cycles in which F_N was held at a constant value for about 60 s at two specific loads (1540 kN going up and 1040 kN going down) to enable the system to mature with time. When F_N is held during loading, strengthening of the interface with time occurs. Such strengthening with time is well known [78–80] and is commonly referred to as ‘ageing’. Surprisingly, the contact area *decreases* with time when the system is allowed to mature during the unloading phase. This was first observed in [74] and was coined ‘de-ageing’. Both ageing and de-ageing evolve quasi-logarithmically in time. Unlike ageing, which has been shown to persist over very long times (10^5 s) [26], de-ageing only persists for a finite time that is dependant on the prior hold time, after which strengthening ensues.

The observations of the hysteresis and memory effects in the contact area themselves are not surprising. In fact, there are a number of possible mechanisms that could produce hysteresis. These include the effects of humidity, adhesive effects or possibly visco-elastic effects. Surprisingly, the aforementioned experiments demonstrated [74] that the behaviour of A as a function of F_N , as shown in figure 4 and 5, is virtually unaffected by changes of the ambient humidity, loading rate or the use of different materials (bases composed of soda-lime and BK7 glass either bare or with different chemical coatings) for the two blocks.

Elasto-plastic descriptions [14, 16, 81, 82] of contact mechanics predict that, following an initial load, the contact surface will deform and cause hysteresis in A during unloading. This is, however, an irreversible scenario and is predicted to occur solely within the first loading cycle. As the contact deformation should not be significantly changed so long as the maximum value of F_N is not increased, upon subsequent loading cycles it was expected that A should *retrace* the unloading curve. Such models, therefore, can not account for the repetitive hysteretic cycles demonstrated in figure 4. If, however, renewal of micro-contacts were to occur between cycles, such repetitive behaviour would occur. For any rough interface with multiple contacts, such effective contact renewal necessitates slip throughout the interface whose value is larger than the size of a single mean contact. Since only the normal load was changed in these experiments, some mechanism that transforms normal load into tangential displacement is therefore required.

One simple and very general mechanism that can give rise to such surface renewal under purely normal external loading is due to the Poisson effect in elastic materials. The Poisson coefficient is the ratio of how strains are coupled in orthogonal directions. Thus, external application of pure stress in one direction immediately generates stresses in the orthogonal ones. Therefore, under certain conditions Poisson expansion/contraction will give rise to differential motion and displacements that, although microscopic in size, are sufficiently large to cause complete contact renewal. This will trivially occur when considering a dissimilar interface,

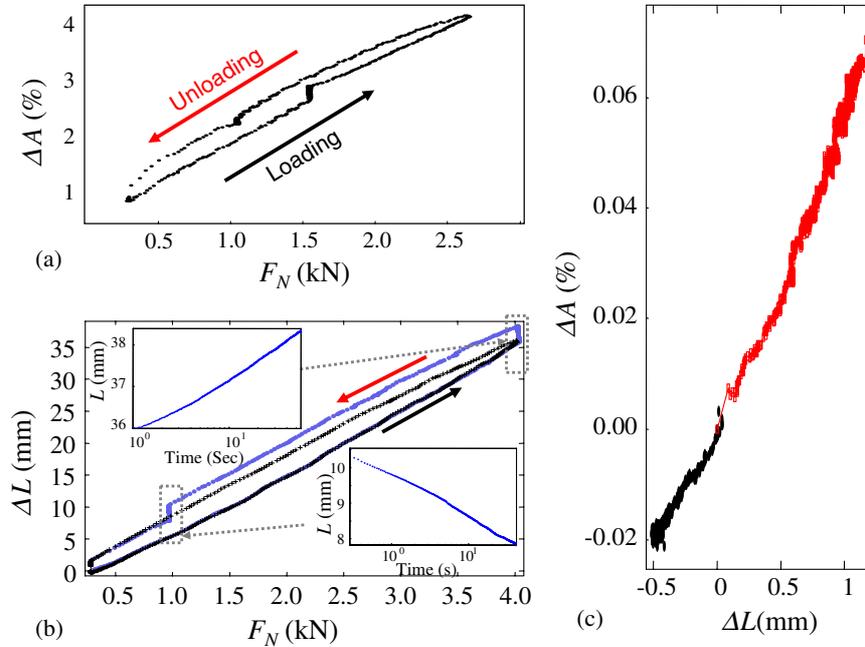


Figure 6. (a) The real area of contact, A (in per cent of nominal contact area), as a function of F_N for a single loading cycle, where the load was held at 1.54 kN upon loading and 1.04 kN upon unloading. (b) Measurement of slider expansion, ΔL , for two consecutive loading cycles, where in the second cycle (circles) the load was held for about a 60 s at 4 kN upon loading (black arrow), and 1 kN upon unloading (red arrow). Increase at 4 kN (upper inset) and decrease at 1 kN (lower inset) of slider length with the logarithm of time. The decrease is observed subsequent to a rapid reduction of F_N and continues until arriving at the steady-state (cross) curve. (c) Change in A (per cent of nominal contact area) as a function of ΔL where the normal load was held at $F_N = 2000$ N during loading (light line) and at $F_N = 1500$ N during unloading (dark line). The nearly linear dependence demonstrates both the strong coupling between ΔL and A and the equivalence of the increase in ΔL with ageing and decrease with de-ageing. Adapted from [74].

composed of two materials with different Poisson coefficients. This will also occur within interfaces composed of the same bulk material where, as in the experimental system described above, the plates have different geometries and/or boundary conditions; e.g. a thin slider and much wider base that are, respectively, under effectively plane stress and plane strain conditions. Moreover, due to the multi-contact nature of the interface, Poisson induced motion could play an important role even in the case of the contact of ostensibly identical bodies. Although, on average, two identical bodies should have no relative motion, contact renewal is not determined by average relative motion but, rather, by microscopic motion of *individual* contacting asperities. In this case, small motions of the contacts around a zero mean will have the same effect as the renewal of surface contacts by a large relative displacement. For Poisson expansion to be negligible, one would have to assume that the expansion/contraction of both inter-connecting surfaces is identical at all scales on both sides of the interface, with each contact perfectly following its counterpart. This assumption is unrealistic as even for moderate loads the overall expansion of the sample can be one to two orders of magnitude larger than the average micro-contact size.

Direct experimental support for Poisson-driven surface renewal was obtained by the direct measurements of the expansion of the sliding block, ΔL , during normal loading and unloading [74]. Measurements of a PMMA/PMMA interface (with no ostensible material mismatch) showed (figure 6(a)) that ΔL closely echoes the hysteretic cycle of the A - F_N measurements.

Notably, ΔL was less than half of the total expansion that would occur for a frictionless interface. Residual shear stresses are, therefore, generated and are held back by the micro-contacts along the interface. These residual stresses account for the ‘de-ageing’ behaviour (figure 6(a)) that occurs immediately after decrease of F_N . At a constant value of F_N , ΔL will, in general, increase with time (figure 6(b), upper inset). This continuous creep allows the system to slowly relax the Poisson-driven shear stress along the interface, which is opposed by the effective friction caused by interactions between the micro-contacts across the surface. This ‘glassy’ behaviour of the contacts also occurs when F_N is suddenly reduced (figure 6(b), lower inset). The corresponding decrease in ΔL is not immediate. Sliding, with a rate that continuously decreases with time, occurs until the shear stresses are, once again, balanced by micro-contact interactions. Surprisingly, as shown in figure 6(b), ΔA is nearly proportional to ΔL . This is true when ΔL both increases and decreases in time, in both cases with the same constant of proportionality.

The hysteretic effects described here are not accounted for in current theories of friction. They demonstrate that micrometre-scale motion has a very large *macroscopic* influence on the frictional properties of an interface. These effects should have a significant influence on the near critical behaviour of frictional systems that are close to their threshold for motion. For instance, as mentioned above, a pre-condition for the generation of stick–slip instability is that some initial motion exists. In this context, Poisson-driven slip could suggest a new mechanism for triggering of new earthquakes via the stress changes generated by previous ones.

The strength of this mechanism for contact surface renewal is in its generality. The micro-slip generated by Poisson contraction or expansion will cause contact renewal in any spatially extended system in which the interaction is restricted to discrete localized regions. Such regions can be micro-contacts, capillary bridges, or any other strongly localized junctions.

2.2.2. *The transition from static to dynamic friction.* In the previous subsection we considered interface motion when only normal loading was externally imposed on the system. We showed that microscopic motion, induced by Poisson contraction or expansion has a large corresponding influence on the system's frictional strength. We now turn our attention to examining the detailed response of a frictional interface to externally imposed shear. When only the overall motion of sliding blocks is considered, the dynamics of frictional sliding are usually thought to be a slow process. However, as noted in the introduction, the detailed dynamics of stick slip are much more complicated and involve both slow quasi-static time scales of nucleation as well as very rapid crack-like processes of contact rupture.

Let us now consider a series of experiments [73, 75, 76] that focus on the detailed dynamics of contact rupture. In these experiments, slip was initiated by increasing F_S at a constant rate after applying a constant value of F_N . The results demonstrate that the dynamics of interface rupture are governed by well-defined crack-like fronts, which extend in the z direction and propagate in the x direction along the interface. Dynamic, real-time, measurements of $A(x, z, t)$, are used to study these fronts. Figure 7 shows six successive snapshots of the net contact area of the interface between two blocks. In these photographs, dark (light) shades correspond to a decrease (increase) in contact area. The short-time changes in contact are emphasized by normalizing each frame by the frame taken at $t = 0$ (immediately prior to the transition to sliding). As a result of the normalization the first frame appears completely uniform. The onset of slip occurs via coherent fronts that propagate across the interface (from left to right). Behind a front the net contact area is reduced, whereas ahead the contact area is unchanged.

The characteristics of the different fronts in a typical experiment are demonstrated in figure 8, where $A(x, t)$ was sampled at a rate of 100 Kframes/sec ($A(x, t)$ is $A(x, z, t)$ after integrating over z). A number of distinct propagating fronts are evident in the plot presented in figure 8. Three different types of crack-like fronts precede sliding [73, 75], each with its own characteristic signature and velocity range. Initially (at $t = 0$) there is no motion along the interface. In this experimental setup shear stress is applied at the trailing edge ($x = 0$) of the slider and rupture is always initiated at this point. The initial fronts that emanate from $x = 0$ are 'sub-Rayleigh' fronts. Sub-Rayleigh fronts propagate at rapid subsonic velocities up to the Rayleigh wave speed, V_R (930 m s^{-1} for PMMA). Once initiated, these fronts rapidly accelerate to approximately V_R (see figure 8(b)). Surprisingly, they then arrest and immediately trigger two new types of detachment fronts, 'slow detachment' fronts and intersonic

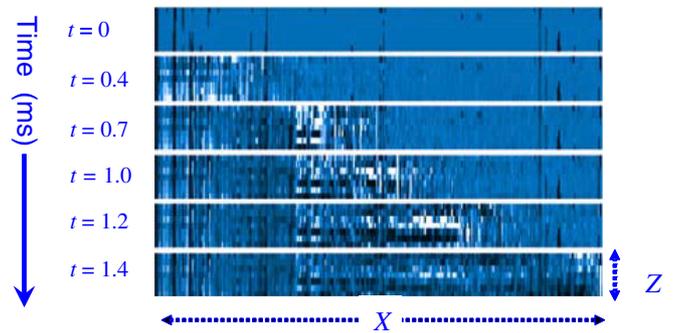


Figure 7. Six photographs of the net contact area of the interface between two blocks normalized by the net contact area at time $t = 0$. The photographs, having spatial resolution of 1280×16 pixels, were taken at (from top to bottom) times of 0, 0.4, 0.75, 1.0, 1.2 and 1.4 ms from the onset of the rupture process. The ($x \times z$) scales of each photograph are $140 \times 6 \text{ mm}^2$. Dark (light) shades correspond to a decrease (increase) in contact area. Adapted from [73].

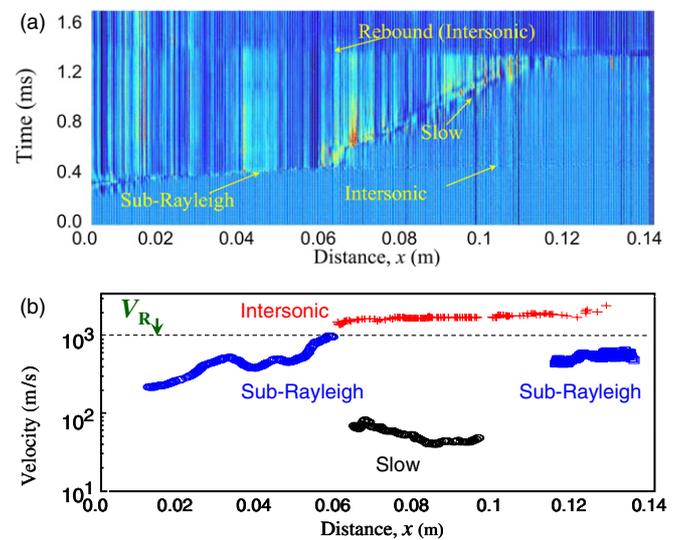


Figure 8. Rapid, sub-millisecond, dynamics of slip at the transition from static to dynamic friction. The dynamics take place through the interplay between different types of coherent crack-like fronts. (a) Measurements of the contact area, $A(x, t)$, as a function of time and position, acquired at a 100 kHz frame rate, averaged in the z direction, of a typical experiment at the onset of slip. Each horizontal line in the figure represents the value of the contact area, $A(x, t = \text{const})$, over the entire interface. The intensity measurements, normalized by their values at $t=0$ reflect the change in the contact area at each spatial point as a function of time. Data is colour coded so that bright shades reflect increased net contact area, while decreased contact area is reflected by dark shades. The dynamics take place through the interplay between four different types of coherent crack-like fronts. The four different observed fronts are labeled within the figure. (b) Velocity values in space corresponding to the slow, sub-Rayleigh, and intersonic fronts shown in (a). The Rayleigh wave speed, V_R , is denoted by the dashed line. Adapted from [73].

[49, 59, 62] fronts (i.e. fronts propagating at intersonic speeds). Both of these detachment fronts are emitted simultaneously. The range of intersonic velocities measured is between 1200 and 1700 m s^{-1} , with a $\pm 200 \text{ m s}^{-1}$ accuracy, placing them

between the shear wave (1000 m s^{-1}) and longitudinal wave (2000 m s^{-1}) velocities of PMMA. The velocities of the slow fronts range from 20 to 100 m s^{-1} , which are one to two orders of magnitude lower than those of the other fronts.

The intersonic fronts are barely visible in figure 8, as they give rise to a minimal (1–2%) reduction of A . In fact, although they nearly immediately traverse the entire interface, they give rise to negligible slip. The slow fronts, however, are much more effective in reducing the contact area and produce an approximate 20% reduction upon their passage. Like the intersonic fronts, once excited a slow front can traverse the entire system. They are also capable of undergoing a reverse transition back to sub-Rayleigh fronts (as shown in figure 8). The sub-Rayleigh fronts are a bit less efficient in contact area reduction than the slow fronts, reducing the contact area by about 10%.

Figure 8(a) also indicates an additional feature accompanying the transition to sliding. ‘Rebound’ waves that initiate at the system’s leading edge upon the arrival of the last of the forward-moving fronts propagate backward across the interface at intersonic speeds of $1300 \pm 150 \text{ m s}^{-1}$. Unlike the forward-moving intersonic fronts that initiate erratic changes in A within the interface, the rebound front appears to suppress this jitter and initiates a uniform evolution of the interface.

Only upon the arrival of the last of the fronts (either a slow or sub-Rayleigh front), does any measurable slip occur at the leading edge of the system. The trailing edge undergoes slip during the passage time of either of these two fronts through the system, but the leading edge stays entirely pinned. (This causes the system to undergo a slight compression—more on this in the next section.) Only when either of these fronts traverses the entire system does any macroscopic motion of the blocks take place.

2.2.3. Precursors to frictional motion. Figure 9 presents a measurement of $F_S(t)$ for a typical experimental run in which F_N is held constant and shear is applied to the edge of the slider at a constant rate (as described in section 2.1). F_S increases in a linear fashion until reaching a peak force (dashed circle in the figure). At this point, F_S drops abruptly as the transition from static to dynamic friction takes place. The abrupt drop is a result of the system’s motion away from the point where shear is applied. For these system parameters, the system undergoes stick–slip motion, as is evident by the saw-tooth shape of the force curve after the peak value of F_S . The dynamics of the transition from static to dynamic friction were described in the preceding section. As we saw, this transition is characterized by the interplay of the three different types (intersonic, sub-Rayleigh, and slow) of fronts. The typical temporal duration of this transition is governed by the average velocity of the slowest of the three fronts and is roughly in the millisecond range (for a 10–20 cm long interface). A close look at the loading curve, however, reveals that well before the Coulomb threshold small (1–2% of the mean value of F_S) drops in the measured shear force are evident.

As discussed in section 1, long-time nucleation processes have been observed during frictional loading. Precursory acoustic signals have also been observed well before sliding

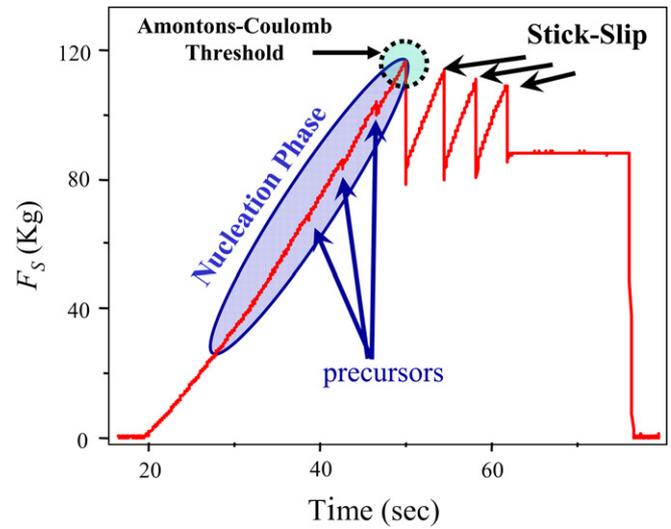


Figure 9. Loading curve in a typical experiment. F_N is held constant and shear is applied by advancing a compliant support coupled to the trailing edge of the slider. Under these conditions, F_S increases nearly linearly until reaching a peak force (dashed circle) followed by an abrupt force drop. Beyond this point, stick–slip motion of the system occurs as indicated by the saw-tooth shape of the force curve. Blue arrows denote small (1–2% of the mean value of F_S) decreases in F_S that occur during the nucleation (shaded ellipse) processes that result from precursor events (described in text) that are excited at values of F_S that are significantly below the shear level needed to excite overall sliding. Stick–slip sliding that occurs after the peak values of F_S are denoted by black arrows.

in studies of stick–slip on an interface composed of smooth Westerly granite [8]. Recently, Bennewitz *et al* [83] conducted a set of experiments where an engineered PDMS tip array was sheared over polished glass surface. These authors found that sliding is preceded by crack-like precursors that cause compressive strain to form along the interface. When shear load is applied to the edge of the sliding block it is natural to assume the existence of some mechanism that transfers the stresses across the frictionally bound interface during loading. Below we will review recent work [76] that has demonstrated how small precursory stress drops in F_S correspond to localized slip that is mediated by a discrete sequence of crack-like fronts that traverse a partial segment of the interface, before arresting. The precursors are triggered at shear forces that are well below those usually associated with the static friction coefficient. They increase systematically in length and significantly redistribute the real area of contact. Thus, by the time that the critical shear force for sliding is reached, the initially uniform contact area along the interface has already evolved to one that is highly non-uniform in space.

A typical experiment is presented in figure 10. Figure 10(b) describes the evolution of $A(x, t)$ throughout the duration of a entire experimental run. In contrast to the work described in the previous section, these experiments occur over times (minutes) that are orders of magnitude longer than the millisecond time scales over which the transition to sustained sliding takes place. The propagation of these precursory fronts and the effect that they have on the real area of contact can be observed in the figure as sharp discrete ridges in the x direction within the $x-t$ plane. Each ridge

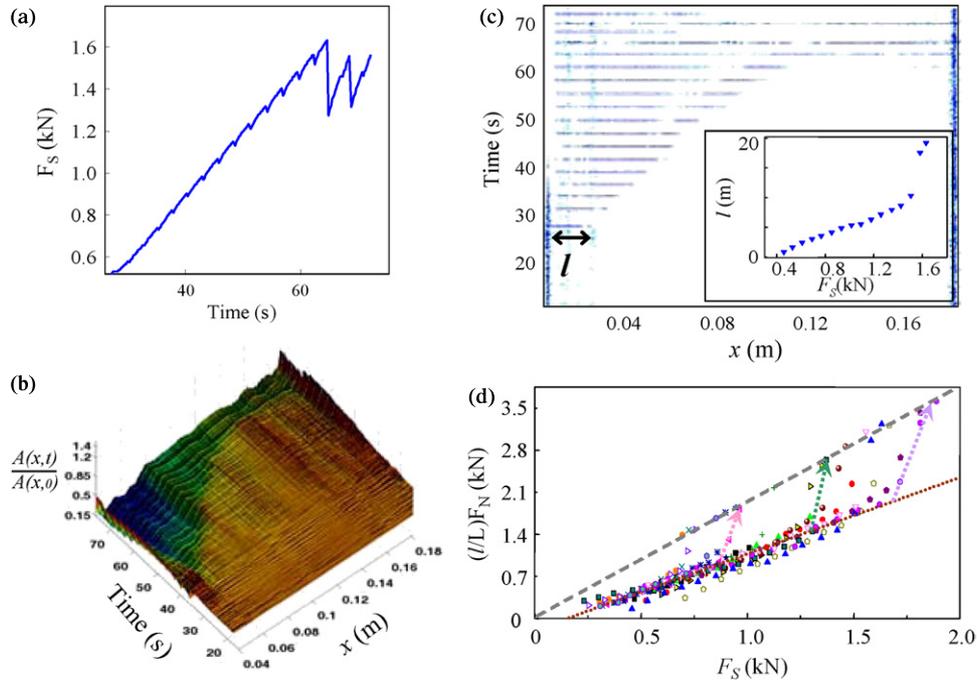


Figure 10. Sliding is preceded by a sequence of frustrated crack-like precursors. (a) The applied shear force, F_S , as a function of time, for a 200 mm long interface. Prior to the first stick–slip type sliding (at $t = 62$ s), a sequence of small stress drops occurs. These are caused by crack-like precursors that initiate at the trailing edge and arrest before reaching the leading edge. (b) $x-t$ plot of the contact area, $A(x, t)$, as a function of time and position for the experiment described in (a). As in figures 7 and 8, $A(x, t)$ is normalized with respect to its spatially uniform value, $A(x, t = 0)$. The discrete precursors in (a) correspond to the steep ridges in the figure. Note that each successive precursor significantly changes the contact area as a weakened region is formed along the path of the precursors while, at the same time, the edges of the interface are strengthened. These discrete precursors are highlighted in (c) where we present the temporal derivative, $|dA(x, t)/dt|$, at each x location along the interface for the data shown in (b). The value of $|dA(x, t)/dt|$ is proportional to the colour intensity. Large changes (blue stripes) in contact area occur solely during the precursor’s short passage time, and along its entire length, l . (inset) l as a function of increasing F_S . For the first 12 precursors l depends approximately linearly on F_S . Above $l \sim L/2$ (where $L = 0.2$ m is the sample length), the linear dependence of l with F_S is broken and l rapidly approaches L with increasing F_S . (d) When F_N is scaled by l/L , data from 23 different experiments collapse on a single curve in the regime prior to the transition to sliding (bottom line). Arrows mark the transition to sliding and the breakdown of scaling in three typical cases. Beyond this transition, the system reaches the Coulomb threshold denoted by the dashed line. Adapted from [76].

corresponds to an event associated with a small stress drop observed in the shear force measurements of figure 10(a). The time derivative of $A(x, t)$ at each spatial point of these data (figure 10(c)) highlights the changes precipitated in $A(x)$ by each event. Large, discrete changes in $A(x)$ correspond to the large and rapid changes in the contact population caused by the propagation of arrested crack-like fronts. Each front is initiated at the trailing edge and traverses a segment of length l (figure 10(c)) before arresting abruptly. In each successive event, l increases by a constant length of Δl (figure 10(c), inset), obeying a generalized friction law $F_S = l/L \cdot F_N$ [76]. This is demonstrated in figure 10(d) where $l/L \cdot F_N$ versus F_S is plotted for 23 different experiments conducted at various values of F_N (using nominal normal pressures ranging from 1.4 to 3.5 MPa) and L (140 and 200 mm). When scaled in this way, all of the data points collapse to a single linear curve. The scaling breaks down as l approaches L , when each $l-F_S$ curve diverges from linearity (arrows in figure 10(d)).

This breakdown of scaling signals the transition from precursor-generated slip to overall sliding motion of the two blocks and usually occurs when l is roughly half of L . The transition at a relatively constant l/L value is strongly indicative of the threshold not being influenced by waves

returning from boundaries normal to the sliding direction, as this is governed solely by the value of L . The upper line in figure 10(d) corresponds to the points where sliding occurs at $l = L$ and F_N is proportional F_S at the onset of overall sliding. This line therefore simply sketches the Amontons–Coulomb law of friction.

A closer look at figure 10(b) reveals that the contact area is significantly altered in the regions where a front has passed. A high concentration of contact area is formed near the trailing edge whereas, beyond this region, the contact area is significantly decreased over the length, l , of each successive event. Towards the leading edge the contact area strengthens. The result of this process is such that, following the passage of the first front, the next one traverses an, already, altered (inhomogeneous) interface. This evolution of contact surface inhomogeneity is highlighted in figure 11, where individual profiles of A over the entire interface taken are shown. This sequence of profiles reveals that the initially uniform contact area may change by as much as 60% before any macroscopic motion takes place. Once such sliding of the blocks occurs, one would naively expect that these inhomogeneities would be erased by the ensuing surface renewal. Surprisingly (figure 11(b)), the interface retains its

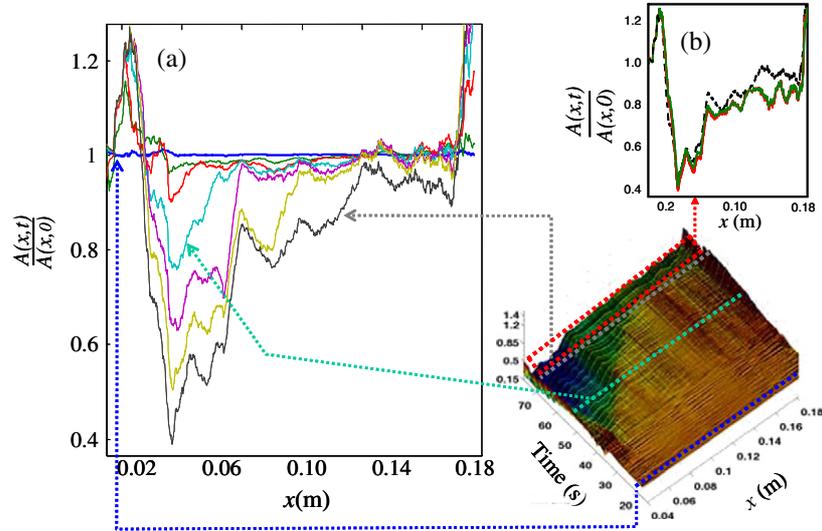


Figure 11. During the nucleation sequence, the contact area is significantly altered. (a) $A(x, t)$ profiles formed by the precursor sequence described in figure 10, where contact profiles formed following every third precursor are shown. For clarity, only three profiles (blue, gray and cyan) are traced back to the full 3D $A(x, t)$ plot. Note how the dynamics are creating a strong barrier adjacent to the trailing edge and a weakened region beyond it (b) Three $A(x, t)$ are shown, taken immediately prior to the first sliding event (dotted black curve) and following the next two consecutive stick–slip event (green and red curves). Note that, perhaps surprisingly, the memory is not erased during the sliding events of approximately $50 \mu\text{m}$ in length and the interface remains deformed. Adapted from [76].

highly deformed configuration throughout the following stick–slip events. In fact, non-uniformity of the contact interface appears to be the system’s preferred configuration for stick–slip sliding.

These data indicate that long-term slip nucleation processes are dominated by the propagation of crack-like precursors, which condition the entire interface towards sliding. In the previous section we showed that there are three distinctly different types of crack-like ruptures associated with frictional sliding; Sub-Rayleigh, intersonic and slow fronts. To investigate what role each of these fronts play in the precursor dynamics, we now turn to the dynamics of the precursor events at micro-second time scales. Figure 12 presents the short-time dynamics of four of the precursory events described in figures 10 and 11. The first two precursors (I and II), which are both well within the linear scaling regime, are qualitatively similar. Both of these events correspond to sub-Rayleigh fronts that initiate at the trailing edge, rapidly accelerate to $\sim 80\%$ of V_R and abruptly arrest within the interface. This behaviour is typical of all of the precursors that are excited within the linear scaling region described by the generalized friction law (lower line in figure 10(d)). In contrast, as seen in the event denoted by III, the breakdown of scaling in the l versus F_S curves is accompanied by a wholly different scenario. Here, the transition region to sliding is accompanied by a slow detachment front that is triggered by the arrest of the sub-Rayleigh front. In the transition region (vertical arrows in figure 10(d)) the slow fronts do not traverse the entire interface, but arrest. Overall sliding of the entire system occurs, as in figure 8, when the slow front, triggered at the same spatial location as in the preceding events within the transition region, propagates either all the way through the interface or triggers an additional sub-Rayleigh front that reaches the leading edge.

2.2.4. An intuitive explanation for the precursory events.

The experiments reviewed in this section suggest an intuitive picture for the sequence of events described in the preceding section, which lead to overall frictional sliding. In these experiments the shear load is applied to the edge of the slider. Initially, the interface is pinned by frictional forces and the applied shear is localized over an area of length l_o from the edge (see figure 13). As shear is increased, it remains concentrated within this region. The size of this region scales with the height, h , above the interface where mean load is applied. Thus, while most of the interface is relatively unaffected by the loading, a region of spatial extent h from the trailing edge is placed under locally high shear stresses ($\sigma_S \approx F_S/h$) (note the significant increase in A near the trailing edge in figure 11(b)). As F_S is increased, the stresses in this localized region increase accordingly until, at a sufficiently high level, a shear crack starts to propagate across this area. This crack both relieves the stresses along this section and transfers the shear stress inward across the interface. Once the crack extends beyond the high stress region, it arrests when the stresses are not sufficient to support fracture. Because a stress singularity exists at the tip of even an arrested crack, the area surrounding the tip of the arrested crack remains under high shear stress. This shear stress cannot diffuse away, because it is trapped by the frictional contacts that are rapidly re-established once the crack has arrested.

As a result of this process, the trailing edge undergoes a slight frictional motion and is displaced from its original position, while the remainder of the interface remains pinned beyond the point of crack arrest. This results in a net compression over the length of the crack. The slider, over this scale, must then undergo a slight elastic deformation. This deformation generates a region of increased contact area of scale h that is adjacent to the trailing edge together with a

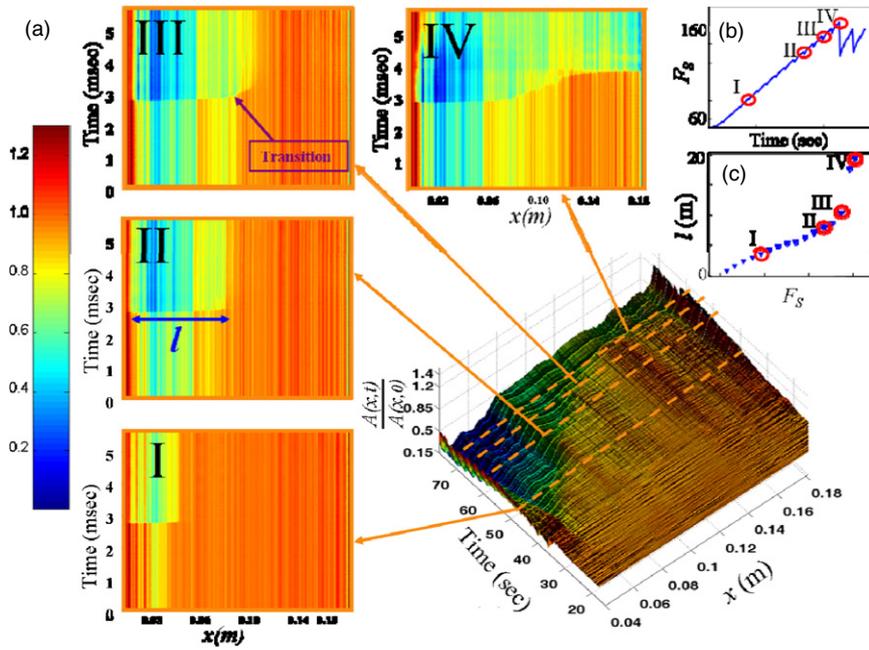


Figure 12. Short-time dynamics of the precursory events (a). Evolution of the relative contact area, $A(x, t)$ normalized by $A(x, t = 0)$ at $F_S = 0$. The four plots (I–IV) correspond to the 6 ms bracketing each precursor’s propagation and are highlighted in the force versus time (b) and precursor length versus force (c) plots to the right of (a). I and II show the details of precursory events within the linear regime in (c), which are characterized by arrested sub-Rayleigh fronts. Event III belongs to the transitional regime that bridges the linear scaling of l with F_S and overall sliding at the Coulomb threshold. The arrow in III marks the transition from a sub-Rayleigh front to a slow front. Event IV corresponds to the Coulomb threshold and is similar to the transition event shown in figure 8(a). Note that the transition to the slow front occurs at the *same* location in x as in III. The colour bar indicates the size of the change in A , when normalized by the value of A at $t = 0$. Adapted from [76].

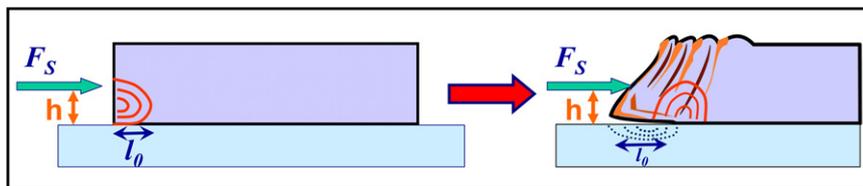


Figure 13. A schematic drawing of how shear stress is transported and surface inhomogeneity is formed during the precursor sequence. Solid lines signify high shear stress concentrations, while dotted curves denote regions in which shear stresses have been relieved. Shown are schematic stress concentrations prior (a), and immediately after (b) the first precursor.

region of reduced contact area in its wake. This region of increased contact area acts as a heightened barrier to slip at the trailing edge.

As F_S is increased further, the process repeats itself for sufficiently high values of F_S . This time, when the new crack traverses the barrier formed near the trailing edge, it propagates into a region that has been both weakened by the previous crack and, at the location where the previous crack arrested, it encounters a highly stressed region that is just nominally below the threshold for fracture. As a result, the new crack easily fractures the pre-stressed micro-contacts in this region and further extends itself. As before, the displacement that the crack leaves in its wake causes additional deformation of the slider, thereby both strengthening the barrier adjacent to the trailing edge and further weakening the region beyond. As F_S is increased still further, this process repeats itself and successive cracks extend further into the interface. Because only regions of limited length $l < L$ are fractured, the threshold values of F_S are all significantly lower than the value

needed to precipitate motion along the entire interface (at the Coulomb threshold).

This simple picture essentially describes the mechanism by which stress is transported and distributed across the interface. The mechanism is as general as the loading configuration that is applied. Note that the strains imposed during the precursory sequence can only be fully relieved if the normal forces are entirely relaxed. As a result, the deformation imposed by the precursor sequence remains intact in successive events (as in e.g. figure 11(b)).

The model described above suggests that the periodicity and size of the precursor events should scale with h . As we show in figure 14, this is indeed the case. The figure presents results from five different experiments that differ only by the mean height where shear load was applied. Shown are derivatives of $A(x, t)$ over the entire interface as a function of applied shear force. Precursor cracks appear as bright horizontal lines, which initiate at the left edge of the frame. Notably, while most of the precursors seem to conform to one

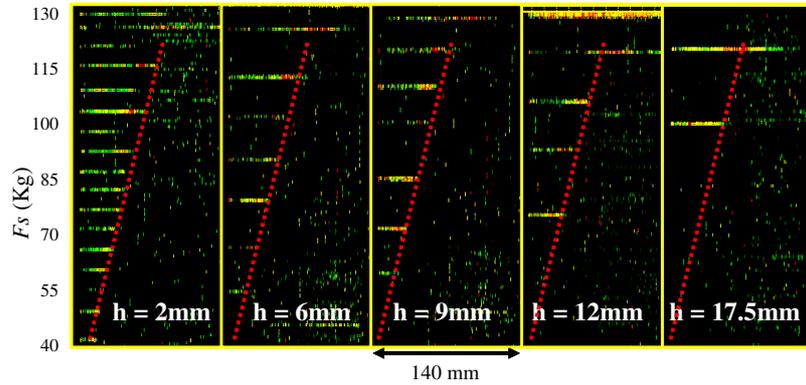


Figure 14. $|dA(x, t)/dt|$ from five different experiments that differ only by the mean height, h , above the interface plane where the shear load was applied. The colour code is chosen so that precursor cracks appear as bright horizontal lines, which initiate at the left edge of the frame. Note that, while most the precursors seem to conform to a single master curve (denoted by the dotted line in the figure), the intervals of F_S between consecutive events become larger with increasing h .

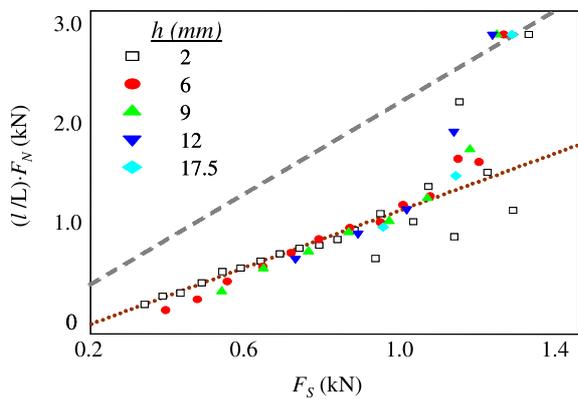


Figure 15. Analysis of the scaling behaviour of the lengths l of each successive slip event. (a). As in figure 10, $l/L \cdot F_N$ is plotted as a function of F_S . The different symbols correspond to experiments where F_S was applied at different heights, h , above the interface. All runs shown here were performed at $F_N = 3$ kN and $L = 140$ mm. Note that while the size of the intervals between successive events is dependent on h , the overall scaling behaviour has no dependence on h . Adapted from [76].

master curve, the intervals of F_S between consecutive events become larger with increasing h .

The observation that all of these precursor trace out nearly the same locus (the dotted red line in the figure) shows that while the force intervals between precursors are highly dependent on h , the scaling regime is h -independent. This is demonstrated in figure 15, where the precursor lengths, l , (scaled as in figure 10(d)) are plotted as a function of F_S for experiments using different values of h . Both the linear regime and point of transition towards sliding (breakdown of scaling) are not affected by the change in h . This suggests that the details of nucleation towards sliding are not sensitive to changes in h .

3. Summary

If we are to understand frictional phenomena we must understand the physics of the interface that is formed between two rough solids forced into contact. Specifically, how different are the physical properties of the interface from that

of the bulk material? The solid blocks that form the interface remain under relatively low stresses during any reasonable loading and sliding configuration. This ‘bulk’ material will, therefore, retain its physical characteristics. This is not the case for the frictional contacts which, by the Bowden and Tabor picture, are *always* close to their yield stress when *any* normal load is applied. Naively, one could assume that the interface should undergo plastic flow that is similar to the evolution of a bulk material, when stressed beyond its yield point. There is, however, a distinct difference between the two cases. Bulk material is continuous while a rough interface is discrete, where the average distance between contacts can be over an order of magnitude larger than their size. As a result, within the interface the coupling between the contacts points is not direct, but solely occurs via the (slightly stressed) elastic bulk. Understanding the behaviour of the interface by analogy to a solid undergoing yield would only be valid if this coupling were so weak that individual micro-contacts could not *affect* each other. In this limiting case, the behaviour of the interface would be simply governed by the uncorrelated actions of its discrete parts. The fact that the observed interface rupture is analogous to mode II fracture strongly suggests that the elastic coupling between the micro-contacts plays a key role in frictional processes—and that this limiting case is unrealistic.

Can we view the rupture process of an interface as a pure mode II fracture process? Here we find the opposite problem. While elastic coupling is too strong to be neglected for a fully plastic description of the interface to be valid, the effects of plasticity within the interface are too robust for an elasticity-based theory to be valid as is. The strength and universality of fracture mechanics lies in the assumption that all of the *complicated*, nonlinear, dissipative processes in which bonds are actually broken can be confined to a small region surrounding a crack’s tip. This is valid for *tensile* fracture, where the resulting free surfaces (and any complex relaxational processes that may occur on them) in the wake of the crack tip have no effect on a crack’s evolution. This is clearly not the case in frictional dynamics, where there are no free surfaces, as bonds are repeatedly broken and reformed over the entire interface. This rehealing both contributes to the overall dissipation (or energy balance) during even partial slip and

contributes nontrivial stresses on the crack faces that should certainly have an effect on the evolution of the rupture.

It is not entirely clear to what extent an approach of separation of time scales to individually describe each stage (i.e. nucleation, rupture propagation, and rehealing) that contributes to the rupture process would be valid. Beyond their immediate influence on rupture dynamics, the complex processes that take place during the reformation of the interface are also important in both preparing and weakening the system. These processes thus set the stage for future rupture events. The details of the surface renewal and strengthening, which occur after the passage of a rupture front, also strongly influence future events. These, however, are still far from understood.

We have reviewed how detailed studies of changes in frictional strength within the interface under various loading conditions reveal rich and complex dynamics. These include the interplay between quasi-static creeping modes, crack-like modes and an anomalous slow dynamic mode. Direct investigation of the spatial and temporal evolution of the real area of contact offers a new and exciting experimental approach to studying frictional phenomena. However, for now, finding a unifying framework that is based on physical principles and accounts for the rich and universal behaviour of frictional dynamics remains an open challenge.

Acknowledgments

The authors acknowledge the support of the Israel Science Foundation (Grant 57/07) as well as the US–Israel Binational Science Foundation (Grant 2006288). JF acknowledges the Max Born chair for Natural Philosophy for its support.

References

- [1] Carpick R W, Agrait N, Ogletree D F and Salmeron M 1996 *J. Vac. Sci. Technol. B* **14** 1289–95
- [2] Gnecco E, Bennewitz R, Gyalog T and Meyer E 2001 *J. Phys.: Condens. Matter* **13** R619–42
- [3] Dieterich J H 1992 *Tectonophysics* **211** 115–34
- [4] Lapusta N and Rice J R 2003 *J. Geophys. Res.—Solid Earth* **108** 2205
- [5] Ben-Zion Y 2001 *J. Mech. Phys. Solids* **49** 2209–44
- [6] Ohnaka M 2004 *Earth Planets Space* **56** 773–93
- [7] Scholz C H 1998 *Nature* **391** 37–42
- [8] Thompson B D, Young R P and Lockner D A 2005 *Geophys. Res. Lett.* **32** L10304
- [9] Luan B Q and Robbins M O 2005 *Nature* **435** 929–32
- [10] Vanossi A, Manini N, Caruso F, Santoro G E and Tosatti E 2007 *Phys. Rev. Lett.* **99** 206101
- [11] Gerde E and Marder M 2001 *Nature* **413** 285–8
- [12] Urbakh M, Klafter J, Gourdon D and Israelachvili J 2004 *Nature* **430** 525–8
- [13] Persson B N J 2000 *Sliding Friction Physical Principles and Applications* (New York: Springer)
- [14] Bowden F P and Tabor D 2001 *The Friction and Lubrication of Solids* (New York: Oxford University Press)
- [15] Greenwood J A and Williams J B P 1966 *Proc. R. Soc. Lond. Ser. A (Math. Phys. Sci.)* **295** 300–19
- [16] Kogut L and Etsion I 2002 *J. Appl. Mech.—Trans. ASME* **69** 657–62
- [17] Byerlee J 1978 *Pure Appl. Geophys.* **116** 615–26
- [18] Scholz C, Johnson T and Molnar P 1972 *Trans. Am. Geophys. Union* **53** 513
- [19] Dieterich J H 1978 *Pure Appl. Geophys.* **116** 790–806
- [20] Vallette D P and Gollub J P 1993 *Phys. Rev. E* **47** 820–7
- [21] Heslot F, Baumberger T, Perrin B, Caroli B and Caroli C 1994 *Phys. Rev. E* **49** 4973–88
- [22] Baumberger T, Caroli C, Perrin B and Ronsin O 1995 *Phys. Rev. E* **51** 4005–10
- [23] Bodin P, Brown S and Matheson D 1998 *J. Geophys. Res.—Solid Earth* **103** 29931–44
- [24] Brace W F and Byerlee J D 1966 *Science* **153** 990–2
- [25] Ruina A 1983 *J. Geophys. Res.* **88** 359–70
- [26] Dieterich J 1979 *J. Geophys. Res.* **84** 2161–8
- [27] Rice J R and Ruina A L 1983 *J. Appl. Mech.* **50** 343–9
- [28] Tullis T E and Weeks J D 1986 *Pure Appl. Geophys.* **124** 383–414
- [29] Baumberger T, Berthoud P and Caroli C 1999 *Phys. Rev. B* **60** 3928–39
- [30] Mair K and Marone C 1999 *J. Geophys. Res.—Solid Earth* **104** 28899–914
- [31] Marone C 1998 *Nature* **391** 69–72
- [32] Linker M F and Dieterich J H 1992 *J. Geophys. Res.—Solid Earth* **97** 4923–40
- [33] Prakash V 1998 *J. Tribol.—Trans. ASME* **120** 97–102
- [34] Marone C 1998 *Annu. Rev. Earth Planet. Sci.* **26** 643–96
- [35] Boettcher M S and Marone C 2004 *J. Geophys. Res.—Solid Earth* **109**
- [36] Bureau L, Baumberger T and Caroli C 2000 *Phys. Rev. E* **62** 6810–20
- [37] Freund L B 1990 *Dynamic Fracture Mechanics* (Cambridge: Cambridge University Press)
- [38] Cotterell B and Rice J R 1980 *Int. J. Fract.* **16** 155–69
- [39] Griffith A A 1920 *Phil. Trans. R. Soc. Lond. Ser. A* **221** 163–98
- [40] Sharon E and Fineberg J 1999 *Nature* **397** 333–5
- [41] Palmer A C and Rice J R 1973 *Proc. R. Soc. Lond. Ser. A—Math. Phys. Eng. Sci.* **332** 527–48
- [42] Andrews D J 1976 *J. Geophys. Res.—Solid Earth* **81** 5679
- [43] Andrews D J 1976 *J. Geophys. Res.* **81** 3575–82
- [44] Brener E A and Marchenko V I 2002 *JETP Lett.* **76** 211–4
- [45] Brener E A, Malinin S V and Marchenko V I 2005 *Eur. Phys. J. E* **17** 101–13
- [46] Baumberger T, Caroli C and Ronsin O 2002 *Phys. Rev. Lett.* **88**
- [47] Sharon E and Fineberg J 1996 *Phys. Rev. B: Condens. Matter* **54** 7128–39
- [48] Ravi-Chandar K, Lu J, Yang B and Zhu Z 2000 *Int. J. Fract.* **101** 33–72
- [49] Rosakis A J, Samudrala O and Coker D 1999 *Science* **284** 1337–40
- [50] Johnson T, Wu F T and Scholz C H 1973 *Science* **179** 278–80
- [51] Johnson T L and Scholz C H 1976 *J. Geophys. Res.* **81** 881–8
- [52] Okubo P G and Dieterich J H 1984 *J. Geophys. Res.* **89** 5817–27
- [53] Ohnaka M and Kuwahara Y 1990 *Tectonophysics* **175** 197–220
- [54] Ohnaka M and Shen L F 1999 *J. Geophys. Res.—Solid Earth* **104** 817–44
- [55] Tippur H V and Rosakis A J 1991 *Exp. Mech.* **31** 243–51
- [56] Lambros J and Rosakis A J 1995 *Int. J. Solids Struct.* **32** 2677–702
- [57] Lambros J and Rosakis A J 1995 *J. Mech. Phys. Solids* **43** 169–88
- [58] Singh R P and Shukla A 1996 *J. Appl. Mech.—Trans. ASME* **63** 919–24
- [59] Xia K W, Rosakis A J and Kanamori H 2004 *Science* **303** 1859–61
- [60] Freund L B 1979 *J. Geophys. Res.* **84** 2199–209
- [61] Gao H, Huang Y, Gumbsch P and Rosakis A J 1999 *J. Mech. Phys. Solids* **47** 1941–61
- [62] Gao H J, Huang Y G and Abraham F F 2001 *J. Mech. Phys. Solids* **49** 2113–32

- [63] Xu X P and Needleman A 1996 *Int. J. Fract.* **74** 289–324
- [64] Needleman A 1999 *J. Appl. Mech.—Trans. ASME* **66** 847–57
- [65] Andrews D J and BenZion Y 1997 *J. Geophys. Res.—Solid Earth* **102** 553–71
- [66] Cochard A and Rice J R 2000 *J. Geophys. Res.—Solid Earth* **105** 25891–907
- [67] Perrin G, Rice J R and Zheng G 1995 *J. Mech. Phys. Solids* **43** 1461–95
- [68] Zheng G and Rice J R 1998 *Bull. Seismol. Soc. Am.* **88** 1466–83
- [69] Heaton T H 1990 *Phys. Earth Planet. Interiors* **64** 1–20
- [70] Dwyer-Joyce R S, Drinkwater B W and Quinn A M 2001 *J. Tribol.—Trans. ASME* **123** 8–16
- [71] Baltazar A, Rokhlin S I and Pecorari C 2002 *J. Mech. Phys. Solids* **50** 1397–16
- [72] Dieterich J H and Kilgore B D 1994 *Pure Appl. Geophys.* **143** 283–302
- [73] Rubinstein S M, Cohen G and Fineberg J 2004 *Nature* **430** 1005–9
- [74] Rubinstein S, Cohen G and Fineberg J 2006 *Phys. Rev. Lett.* **96**
- [75] Rubinstein S M, Shay M, Cohen G and Fineberg J 2006 *Int. J. Fract.* **140** 201–12
- [76] Rubinstein S M, Cohen G and Fineberg J 2007 *Phys. Rev. Lett.* **98**
- [77] Briscoe B J, Fiori L and Pelillo E 1998 *J. Phys. D: Appl. Phys.* **31** 2395–405
- [78] Rabinowicz E 1995 *Friction and Wear of Materials* (New York: Wiley)
- [79] Berthoud P, Baumberger T, G'Sell C and Hiver J M 1999 *Phys. Rev. B* **59** 14313–27
- [80] Crassous J, Bocquet L, Ciliberto S and Laroche C 1999 *Europhys. Lett.* **47** 562–7
- [81] Pei L, Hyun S, Molinari J F and Robbins M O 2005 *J. Mech. Phys. Solids* **53** 2385–409
- [82] Kim J Y, Baltazar A and Rokhlin S I 2004 *J. Mech. Phys. Solids* **52** 1911–34
- [83] Bennewitz R, David J, de Lannoy C F, Drevniok B, Hubbard-Davis P, Miura T and Trichtchenko O 2008 *J. Phys.: Condens. Matter* **20** 015004