Crack-like processes within frictional motion - Is slow frictional sliding really a slow process?

Shmuel M. Rubinstein, Gil Cohen, and *Jay Fineberg

Abstract

The dynamics of frictional motion have been studied for hundreds of years, yet many aspects of these important processes are not understood. First described by Coulomb and Amontons as the transition from static to dynamic friction, the onset of frictional motion is central to fields as diverse as physics, tribology, mechanics of earthquakes and fracture. By performing fast (real-time) visualization of the true contact area along a rough spatially extended interface separating two blocks of like material, we study the detailed dynamics of how this transition takes place The onset of motion is preceded by a discrete sequence of rapid crack-like precursors, which are initiated at shear levels that are well below the threshold for static friction. These precursors systematically increase in spatial extent with the applied shear force and leave in their wake a significant redistribution of the true contact area. Their cumulative effect is such that, just prior to overall sliding of the blocks, a highly inhomogeneous contact profile is established along the interface. At the transition to overall motion, these precursor cracks simultaneously excite, at their point of arrest, both slow propagation modes and intersonic ones. Overall frictional motion only takes place when either the slow propagation modes or additional shear cracks excited by these slow modes traverse the entire interface. Surprisingly, in the resulting stick-slip motion, the surface contact profile retains the profile built up prior to the first slipping event. These results suggest a fracture-based mechanism for stick-slip motion which is qualitatively different from other accepted descriptions.

Keywords: Friction, Tribology, Fracture, Slip.

Background

The dynamics of frictional motion are critical to fields ranging from nano-machines[1] to the study of earthquakes[2-7]. Frictional motion involves a huge range of time and length scales[8-11], coupling the elastic fields of two blocks under stress to the dynamics of the myriad interlocking microscopic contacts that form the interface defining their plane of separation. In spite of the immense practical and fundamental importance of friction, the basic physics of the dynamics of frictional motion along rough, spatially extended interfaces is far from complete.

The vast majority of the research performed on the dynamics of dry friction has focused on the macroscopic (overall) motion of two nominally rigid bodies in frictional contact. In this view, the bodies in motion are considered as nominally rigid and the relevant spatial degrees of freedom are the center of mass coordinates of each body. The sliding velocities considered in these studies are normally quite slow (μm/sec to mm/sec). Here we will describe recent work on

these processes in which the spatial degrees of freedom along a large contact plane (consisting of 10^7-10^8 discrete micro-contacts) separating the two bodies are taken into account. We will review recent results that demonstrate that families of rapid crack-like fronts that sweep through the interface, described by the contact plane, are the underlying mechanism that gives rise to the transition to macroscopic frictional motion. Although the sliding velocities are quite slow, the velocities of these fronts can be very rapid (some of these approach material sound speeds), so that short (μsec-msec) time scales must be considered. This work is quite new, and in many ways should be considered as "work in progress". We first present a brief overview of a number of key ideas and concepts. We then go on to a brief description of how the short-time behavior of the spatial degrees of freedom along the contact surface can be measured, and conclude with a description of the experimental results that show how the dynamics of these fronts drive the transition to frictional motion.

Macroscopic descriptions of friction and the importance of the net contact area

The dynamics of friction are generally considered to be slow processes, where the overall *macroscopic* motion of two bodies in frictional contact is mainly considered. In typical experiments where frictional properties are studied, both a normal force, F_N, and shear force, F_S, are applied to two blocks in contact. The sliding bodies are considered as rigid blocks which are parameterized by spatially uniform parameters such as their overall sliding velocity (V), normal and shear forces, and the friction coefficients (μ). In the traditional "Amontons-Coulomb" description of friction, friction coefficients are defined as the ratio F_S/F_N at the onset of motion ("static" friction) or during sustained sliding ("dynamic" friction). Although this law provides the basis for our phenomenal understanding of frictional dynamics, over the past two decades an impressive body of work has demonstrated that frictional dynamics are considerably more complex than this simple picture describes. The friction "coefficient" is a function of both the sliding velocity and the time of contact. A class of empirical models known as "rate and state" descriptions of friction[12-14] successfully describes many of the observed macroscopic phenomena that are observed in dry friction (friction without lubricants). These models provide a surprisingly good description of the frictional behavior of a broad range[12, 14-18] of materials.

As Bowden and Tabor first noted[19], the *net* area of contact, A, between two nominally flat surfaces is a crucial quantity for understanding frictional dynamics. In general, even a nominally flat plane of contact between two materials consists of myriad interconnecting contacts, called asperities and A is generally a very small fraction of the *nominal* or apparent contact area. A is not a static quantity but actively evolves as a function of applied loads. When two materials are in initial contact, A is negligibly small. As a result, the overall pressure, F_N/A, on each of these contacts easily exceeds the material's yield stress and material flow initiates at all points of contact. This flow causes A to grow at every asperity until the local pressure is reduced below the yield stress. Thus, in the Bowden and Tabor picture, $F_N \propto A$. Upon the application of a shear force,

relative motion will not occur until the resulting contacts are "fractured". Thus, F_S is also proportional to A and the Amontons-Coulomb law can be understood.

Although the evolution of A plays a critical role frictional dynamics, direct measurement of A has proven to be experimentally challenging. Optical measurements of changes of A on microscopic single contact level, performed by Dieterich and Kilgore[20] for a number of materials, have indeed directly verified the Bowden and Tabor picture. In addition, recent optical measurements[21, 22] of macroscopic single contacts have provided direct tests of single-contact models in idealized situations. Optical visualization of A in large, spatially extended multi-contact interfaces[23] has revealed that, even without the application of shear, spatially dependent Poisson-driven dynamics can significantly affect the evolution of A.

Rapid fracture processes: shear cracks

As we noted above, frictional forces are governed by the net contact area, $A(\mathbf{x},t)$. As Bowden and Tabor suggested, the onset of friction takes place upon the "fracture" of the ensemble of micro-contacts that comprise the interface. In this sense, it is important to become acquainted with the fundamental concepts that are important in rapid fracture processes.

Rapid fracture occurs as a result of the propagation of cracks, which generally develop from the evolution of defects within a medium under stress. When stresses are applied to a material, cracks that are embedded within the material serve to concentrate stress at their tips. Fracture mechanics[24] tells us that these stress concentrations actually approach a stress singularity for mathematically sharp cracks. The mode of fracture that is most often considered is that of tensile fracture (known as "Mode I" fracture), where a crack is loaded under conditions of pure tension at its tip. In fact, a basic tenet of fracture mechanics (called the "principle of local symmetry") predicts that, in a homogeneous material, a propagating crack will *only* be subject to pure tensile stresses at its tip. If a homogeneous material is remotely loaded, for example, in pure shear conditions, any propagating cracks within the material will orient themselves to be locally under conditions of pure tension. Thus, in a homogeneous material "shear cracks" cannot exist.

An exception to this occurs in the case of an inhomogeneous material. One example of this is that of fracture along a weak interface; as predicted to occur in the Bowden-Tabor picture of friction. A frictional interface is composed of an ensemble of sparse, interlocking micro-contacts. Because of the sparseness of these contacts, the strength of a frictional interface is necessarily significantly weaker than in either of the surrounding bulk materials. In such instances, the principle of local symmetry does not necessarily apply and cracks that are locally under shear can propagate along this weak direction.

What is the fracture criterion for an interface crack? In tensile fracture there is a well-established criterion (the Griffith criterion) whereby the energy per unit area flowing into the tip of a crack must equal the fracture energy, the energy per unit area needed to create two new fracture surfaces. In the case of an interface crack, however, there is currently no generally accepted fracture criterion.

Although brittle fracture in the case of tensile fracture has been the object of considerable study[25], experimental study of dynamic shear fracture has been quite limited. In recent years, however, there has been a surge of interest, which was originally motivated by the need to both characterize and understand the failure modes of composite (multi-phase) materials. Studies of dynamic interface fracture driven by projectile impact[26] performed on a bimaterial interface, established that interface cracks can propagate at velocities up to and beyond the Rayleigh wave speed, V_R, (the limiting crack velocity in homogeneous materials) of the "softer" of the two materials surrounding the interface. Later studies[27] of interface fracture along an interface bounded by the same materials revealed that interface cracks are able to travel at speeds beyond the Rayleigh wave speed. In these experiments intersonic (velocities between the shear and longitudinal wave speeds) crack propagation along the interface was observed. These experimental results were later observed in molecular dynamic simulations[28]and finite-element calculations[29].

Experimental apparatus and methodology

We have seen above that knowledge of the detailed spatial-temporal dynamics of net contact area evolution as shear is applied, is important to an overall understanding of frictional processes. We have also seen that shear-induced fracture processes can indeed occur along a weak interface and that shear cracks can propagate at extremely high velocities. To determine whether such rapid spatial and temporal processes couple to frictional motion, it is therefore important to be able to measure the real area of contact at short (~μsec) times over long, spatially extended frictional interfaces.

Load application and sample geometry

A schematic diagram of the experimental system used in our experiments is presented in Figure 1a. A detailed description of our apparatus appears in Rubinstein et al.[30, 31]. To examine the short time dynamics of long quasi-one dimensional frictional interfaces, our group performed a series of experiments where the interface was formed by two long and thin PMMA (Polymethyl-Methacrylate) blocks. The lower block ("base") had (x,y,z) dimensions of (300,27,30) mm and upper blocks ("sliders") of sizes (140, 75, 6) mm and (200, 75, 6) mm, where the x, y, and z are, respectively, the direction of shear application, normal loading direction, and sample thickness. To ensure that the interface was both flat and consisted of a uniformly disperse ensemble of micro-contacts, the sliders' contact faces were first diamond machined to a flatness of

better than 0.1μm (rms). The sliders and base were then lapped to approximately a 1μm (rms) roughness.

In each experiment, F_N was fixed to a constant value which, to minimize interface aging effects, was applied over 100s prior to the application of F_S. To ensure spatial uniformity of the applied normal load, and to decouple the normal from shear loads, F_N was applied to the slider via a spring array coupled to a linear stage with a low static friction coefficient and negligible dynamic friction (0.002). F_S was applied (typically at rates of 0.5-1.5kN/min) by fixed rate displacement of a stepper-motor based actuator coupled to the "trailing edge" (see Figure 1a) of the sample via a load cell of stiffness 10^6 N/m. In the experiments described below, F_S was applied to the sample's trailing edge at a (center of mass) height, h, (2 mm < h< 20mm) above the interface plane.

Real-time visualization of the net contact area

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. The visualization system is shown schematically in Figure 1b. The roughened interface was illuminated via the transparent base block, by a 200 x 5mm Argon laser sheet, at an angle significantly larger than the critical angle for total internal reflection from the PMMA/Air boundary at the interface. As the interface was rough in comparison to the exponential decay length of evanescent light (~50nm), light passed into slider solely in the immediate vicinity of the contact points (where separation was less than 50nm). The intensity of light, $I(x,z,t)$, passing through the interface was therefore proportional[31] to the net contact area, $A(x,z,t)$ at each (x,z) point along the interface. In Figure 1c, the integrated intensity, I, is shown to be proportional to F_N, as expected from the Bowden and Tabor picture, when $A \propto I$. A fast CMOS sensor (VDS CMC1300 camera) was used to image $I(x,z,t)$. The sensor could be configured to frame sizes of 1280 x N pixels with frame rates of 500,000/(N+1) frames/sec. In our experiments, the entire interface was measured where each sensor pixel was generally mapped to Δx x Δz values of either =117μm x 1.5mm at 100,000 frames/sec (N=4) or 117μm x 1.0mm at 71,000 frames/sec (N=6). Each pixel encompasses numerous contact points (the mean distance between contact points was approximately 10μm).

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 takes place in the sub-msec range. Data was continually sampled at maximal frame rates into a circular buffer with a 2 second capacity. Acoustic signals, generated by rapid slip events, were used to trigger storage to disk of several thousand frames, which bracketed the acoustic signal. This enabled effectively continuous acquisition of $I(x,y,t)$ at maximal frame rates. The acoustic signals were detected by a sensitive broadband acoustic sensor (Valpey-Fisher pinducer), coupled to the leading edge of the slider. In parallel to this rapid triggered acquisition, a continuous record of the contact area

evolution was maintained by storage of single frames at fixed (20-100 msec) intervals throughout the duration of the experiment.

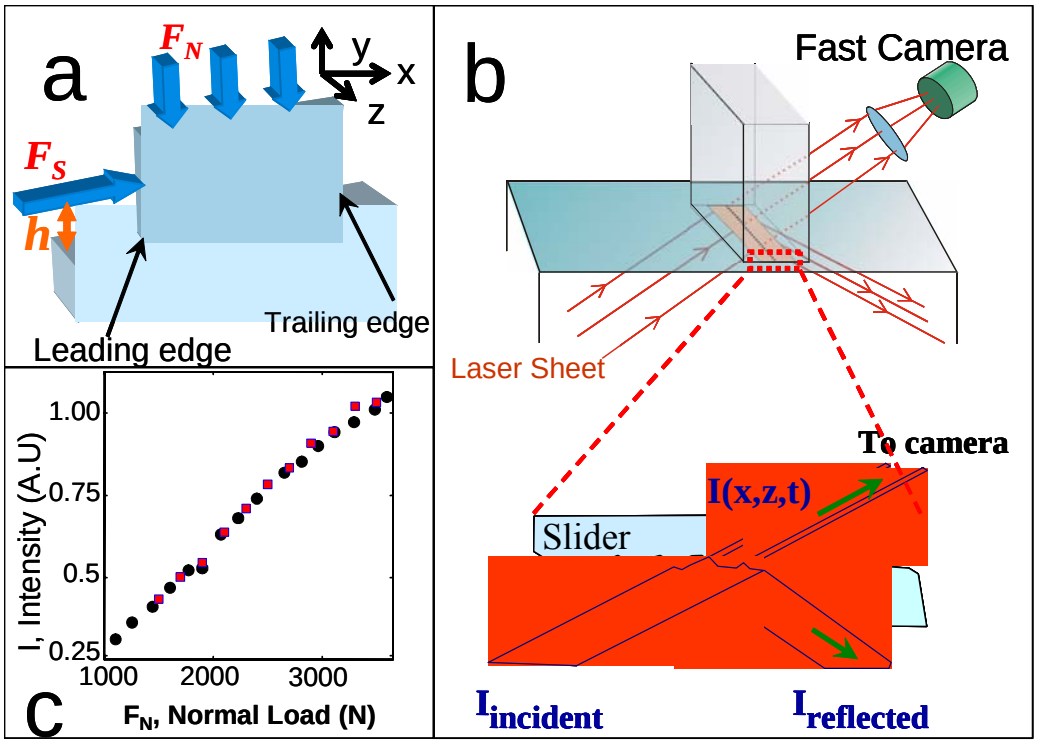

Figure 1. A schematic view of the experimental system. (a) The loading configuration used in the experiments, where an initially uniform normal force is imposed (in the *y* direction) and shear (in the *x* direction) is applied at an edge (the "trailing edge") of the system. (b) An illustration of how the real contact area is imaged in real-time. As the rough surface of the interface is illuminated with a laser sheet, whose incident angle is far beyond that for total internal reflection, the only intensity, $I(x,y,t)$, transmitted through the interface is proportional to the real contact area at each point along the interface. This is imaged onto a fast camera, capable of imaging the entire interface at frame rates exceeding 100,000 frames/sec. (c) The integrated intensity across the entire interface as a function of applied normal load, F_N demonstrates that I is indeed proportional to the net contact area, A. The figure was adapted from (30).

Rapid processes at the transition from static to dynamic friction

The Onset of Frictional Motion

A loading curve, $F_S(t)$, in a typical experiment is presented in Figure 2. The applied value of F_S arrives at its peak ("static peak") value (circle in Figure 2) at the Amontons-Coulomb threshold, when overall frictional sliding commences. Our system is in the stick-slip regime of frictional sliding. Slip-stick is indicated by the saw-tooth profile in the loading curve. The sharp drops in the F_S (t) curve are due to the finite compliance of the load cell that is used to measure F_S. At

each slip of the trailing edge, the load cell becomes slightly decompressed and the applied values of F_S drop accordingly.

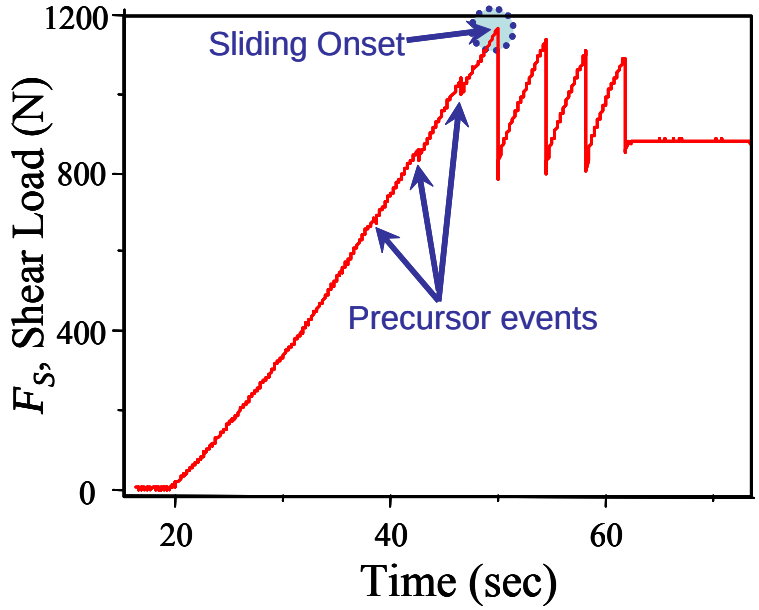

Figure 2. A loading curve in a typical experiment. After imposing a given value of F_N, shear, F_S, is imposed starting from zero until sliding occurs in the region denoted by the shaded circle. Beyond this point, stick-slip motion of the system occurs. Arrows denote small stress drops that occur as a result of precursor events that are excited at values of F_S that are significantly below the shear level needed to excite overall sliding.

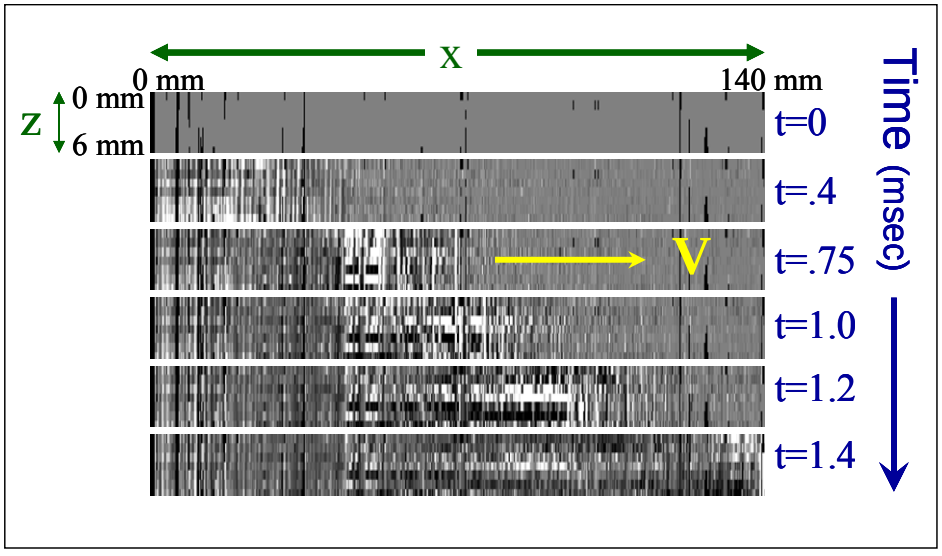

Figure 3. The onset of slip occurs via coherent fronts[30] that propagate across the interface (from left to right). Behind these fronts the net contact area is reduced. Shown are six photographs, of spatial resolution 1,280 × 16 pixels, of the net contact area of the interface between two blocks, taken at times (in msec) noted on the right. All photographs are normalized by the net contact area at t = 0. Dark (light) shades correspond to a decrease (increase) in contact area. The figure was adapted from (30).

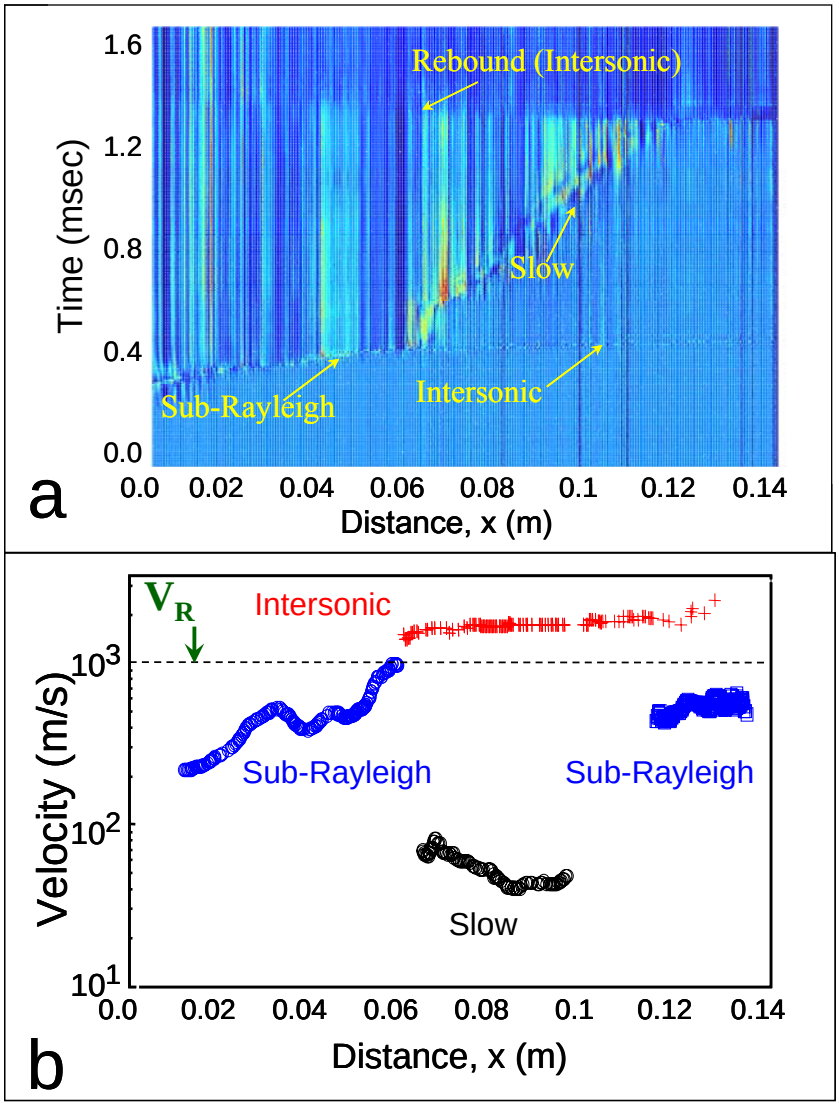

Figure 4. The dynamics of slip, at the sub-msec time scales surrounding the transition from static to dynamic friction. The dynamics take place via the interplay between different types of coherent crack-like fronts. (a) An *(x,t)* plot of relative intensity measurements, averaged in the z direction, of a typical experiment at the onset of slip (here a 100,000 frames/sec sampling rate was used). Each horizontal line in the figure is the proportional to the *A(x,t)* over the entire interface at time *t*. The intensity measurements, normalized by their initial values at each point, are color-coded to reflect the change in the contact area at each spatial point as a function of time. Hot (cold) colors reflect increased (decreased) net contact area. The four different observed fronts are labeled within the graph. (b) Velocity records in space corresponding to the slow, sub-Rayleigh, and intersonic fronts seen in (a). The figure was adapted from (30).

We first consider the dynamics of interface that occur at short times in the immediate vicinity of the static peak, when the transition from static to dynamic friction is just taking place. In Figure 3 we present a series of photographs of the net contact area *A(x,z,t)*, taken at (relatively "slow") 0.4 msec intervals, which bracket the static peak. *A(x,z,t)* was normalized by *A(x,z,t=0)*, where *t*=0 is a time

immediately preceding the sliding events that occur at the static peak. The figure shows that a quasi-1D front sweeps across the interface. The front initiates at the sample's trailing edge (the edge where F_S is applied) and traverses across the entire interface, leaving in its wake a region of significantly reduced contact area (darker grey-scales in the figure).We take a closer look at the overall dynamics surrounding the transition from static to dynamic friction in Figure 4, at a much more rapid data rate. Each horizontal line in the figure is a measurement of the real contact area, $A(x,t)= \int A(x,z,t)dz$, integrated over the sample width. The frames were acquired at 10μsec intervals. We emphasize the short-time changes in A by normalizing each line by $A(x,t=0)$, as in Figure 3. The typical transition sequence demonstrated in the figure shows that the transition to frictional motion is governed by the interplay of a number of different *types* of fronts. The first of these is a rapid crack-like, "sub-Rayleigh", front which initiates at the trailing edge ($x=0$) of the sample and propagates into the interface. Sub-Rayleigh fronts initiate at high (400-500m/s = 0.4-0.5V_R) velocities and accelerate up to the Rayleigh wave speed, V_R, which is the limiting velocity for both tensile and *most* shear cracks. Instead of traversing the entire interface, once this front approaches the limiting velocity (V_R ~ 1000m/s in PMMA) of a "garden-variety" shear crack, it undergoes a sudden transition to two additional types of fronts, intersonic and "slow" fronts. In Figure 4a this transition takes place at approximately t=0.4msec, when (to our resolution) both types of fronts are simultaneously emitted.

Intersonic fronts

A glance at the velocities corresponding to these two new types of fronts (Figure 4b) indicates that they have entirely different characters. The intersonic fronts propagate at velocities that are significantly faster than V_R. The range of intersonic velocities that we have measured[30] are between 1200-1700m/s (1.3-1.8V_R), with a ±200 m/s accuracy. These fronts propagate at velocities similar to those measured for intersonic cracks[32].

We differentiate between pure "cracks" and the fronts that we observe propagating along contact interfaces in the following sense. In the case of shear (Mode II) cracks, fracture of bonds occurs solely at the crack tip and the crack faces that are formed in the wake of the crack tip's passage are considered to be "frictionless" free surfaces. In the case of the frictional fronts that we observe, the surfaces behind the leading edge of the front are *always* in very substantial frictional contact. It is far from obvious that either intersonic or sub-Rayleigh cracks (in fracture) and intersonic or sub-Rayleigh *fronts* along a frictional interface are the same object.

While bearing this qualitative difference in mind, the close resemblance between the limiting velocities and dynamics of both types of fronts to the corresponding types of cracks is certainly suggestive that the phenomena are related. The transition between sub-Rayleigh and intersonic fronts is reminiscent of recently measured crack-like fronts which were observed to exist in projectile-driven fracture[27, 33] and in explosively triggered frictional motion[27, 34]. In these

experiments, which measured the stress field in the bulk material surrounding the interface, sub-Rayleigh cracks were observed to trigger intersonic cracks that propagated ahead of the sub-Rayleigh cracks that formed them. Once this transition occurred, the intersonic cracks appeared to generate the dominant stress fields in the surrounding material. In our experiments the loading is not explosive but very slow. In addition, we did not measure the stress fields in the bulk, but rather the contact area generated by the different fronts. In these experiments, intersonic fronts were not observed to have a large apparent quantitative effect on the contact area along the interface[30, 31]. Once excited, intersonic fronts left a very faint `footprint', reducing $A(x,t)$ by a minimal (1-4%) amount upon their passage. We understand this slight reduction in $A(x,t)$ as being due to the fracture of a small sub-population of micro-contacts along the interface as a result of the passage of intersonic fronts. This idea is supported by the fact that, upon the passage of intersonic fronts, displacement sensors positioned at the leading edge of the slider detect (to sub-μm resolution) *no* overall frictional motion of the two blocks in our experiments. The direct effect of these fronts on the contact area is, therefore, minimal in our experiments, although they may well be important in other contexts and/or loading conditions. Upon their passage, intersonic fronts do, however, initiate erratic motion of the contact points along the interface. It is not yet clear how this "noisy" behaviour couples to the ensuing dynamics along the interface.

Slow fronts

In contrast to the sub-Rayleigh and intersonic fronts, which may be intuitively related to known types of shear cracks, the slow fronts observed in Figure 4 do not appear to have any known analogs in fracture. Observed propagation velocities of slow fronts range from 20-80m/s. These velocities are, on the one hand, over an order of magnitude slower than either the sub-Rayleigh or intersonic fronts. On the other hand, slow front velocities are 6-7 orders of magnitude larger than the typical loading rates imposed on the system.

While the passage of intersonic fronts through the interface produced little overall changes in the system, slow fronts produce a much greater overall effect. Upon their passage, these fronts precipitate an approximate 20% reduction in contact area. This is large, when compared to both the 10% reduction in $A(x,t)$ generated by sub-Rayleigh and the 1-4% reduction in $A(x,t)$ caused by the intersonic fronts. Once generated, slow fronts will either traverse the entire interface, or undergo sharp `reverse' transitions to sub-Rayleigh fronts. Overall relative motion of the two blocks will not occur until either the slow or sub-Rayleigh fronts reach the leading edge of the system. Once this occurs, frictional sliding is initiated. Sliding is generally accompanied by the generation of an intersonic `rebound' front (see Figure 4a) from the leading edge of the interface.

While the leading edge of the interface remains pinned until the arrival of either a slow or sub-Rayleigh front, the trailing edge of the interface is in motion when either of these fronts is propagating through the system. The measured `slip' velocity of the trailing edge is in the range of 1-5 cm/s and, since the leading edge

of the interface is pinned during this time, the resulting displacement at the trailing edge serves to compress the system. As shear is applied to the system at the trailing edge, we see that the interplay of these three types of fronts is the vehicle which transfers the applied shear across the interface. In the next section we will demonstrate that this shear transfer is the culmination of a process that is initiated at shear forces that are far below the level at which the transition from "static" to "dynamic" friction actually takes place..

Precursors to frictional motion

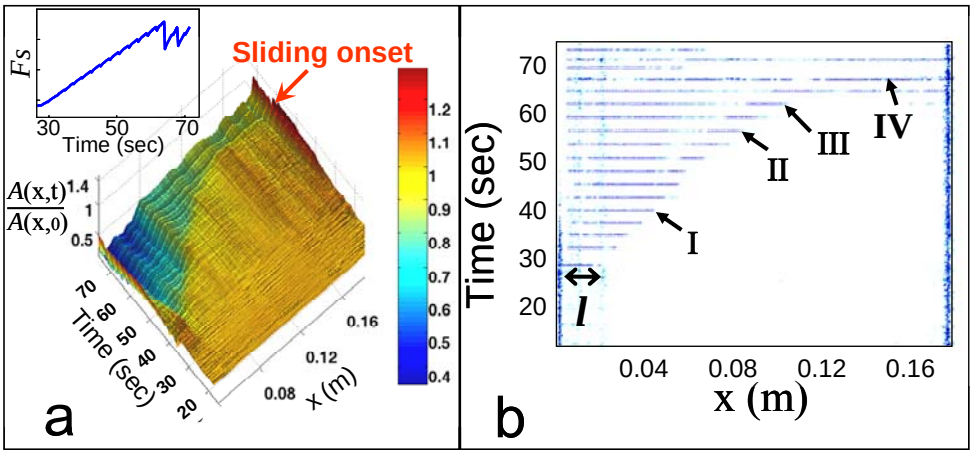

Figure 5. (a) The contact area *A(x,t)* as a function of time and position. *A(x,t)* is normalized with respect to its spatially uniform value *A(x,0)* at the start of the experiment. The discrete slip events noticeable in the loading curve (inset) are manifested as bright ridges in the *A(x,t)* plot. Sliding is preceded by a sequence of precursors; discrete crack-like events that initiate at the sample's trailing edge. (b) The temporal derivative ∂A/∂t at each *x* location whose intensity is proportional to ∂A/∂t. The major changes in *A* are precipitated by the precursors' passage. Each value of *A(x,t)* was smoothed over a 10 mm scale. The figure was adapted from (44).

While it is clear from the above discussion that the observed fronts are the vehicle for the transition to frictional sliding, it is currently unclear how their dynamics couple to both the threshold and predictability of frictional failure. Failure predictability is related to frictional precursors. Acoustic precursors have been observed in laboratory studies of fracture nucleation[35] and stick-slip[7] in granite blocks under triaxial loading. Rapid precursor events have also been observed in models[36, 37] of amorphous materials under shear, where shear localization along slip planes yields frictional sliding along spontaneously formed frictional interfaces. We now demonstrate that precursory dynamics can take place along the interface at stresses that are well below those needed to initiate macroscopic frictional motion. We will also show that the non-spatially uniform loading applied in our system (i.e. shear is applied to the system at the trailing edge of the slider) leads to precursor evolution that dynamically generates a highly non-uniform contact-area profile. As non-uniformity locally increases (decreases) fracture toughness along the interface, this process is directly relevant to both the nucleation[3, 38, 39] and arrest[40, 41] of mechanisms that promote interface

weakening and eventual failure (to sliding) . These are intrinsically important fundamental questions with ramifications to both material science and earthquake dynamics.

While the spatial profile of frictional contact surfaces is often considered to be statistically uniform [8, 10, 11, 19], interface inhomogeneity has been considered in processes which include earthquake arrest due to pre-existing barriers[3, 6, 7, 38, 39, 42] and stick-slip in frictional sliding[43]. It is generally assumed, however, that such inhomogeneities do not arise dynamically.

Let us now look again at the shear loading curve presented in Figure 2. A close look at the curve reveals a sequence of very small drops in F_S (labelled as precursor events in the figure) at applied shear values that are well below the sliding onset. These small stress drops, which could easily be missed, result from minute (~1μm) motion at the system's trailing edge. The fact that they are observed at all in these experiments is due to the relatively high stiffness (10^7 N/m) of the loading system. We shall see, however, that they indicate substantial changes that are occurring along the ostensibly "static" interface.

We can observe these changes by measuring the corresponding dynamics of the contact area[44]. In Figure 5a we present a plot of the spatial and temporal changes occurring in $A(x,t)$, when normalized by the real contact area at the time, $t = 0$, when F_S was first applied. This normalization is important, as it highlights long-time changes in $A(x,t)$. (This is in contrast with the normalization of A used in the Figure 4a, where A was normalized at a time that preceded the onset to sliding by less than 1 msec.) A rather surprising feature of these measurements is the enormous changes in $A(x,t)$ that take place at values of F_S that are far lower than the imposed shear at the sliding onset. Typically, local values of $A(x,t)$ can undergo changes of nearly 50% well before the transition from static to dynamic friction takes place. In addition, we see that the distribution of contact area along the interface evolves from an entirely uniform spatial distribution, when F_S is first applied, to a highly non-uniform spatial profile immediately prior to sliding.

Let us now take a closer look at the dynamics along the interface that mediated these large changes in $A(x,t)$. In Figure 5b we plot the changes, $\partial A/\partial t$, that took place in the data presented in Figure 5a, at each spatial point along the interface. This representation of the data immediately highlights the fact that all of the major changes in $A(x,t)$ occurred as a result of a sequence of discrete, arrested precursor events that initiate at the trailing edge of the sample and arrest within the interface, before reaching the sample's leading edge. Each of these large changes in contact area corresponds to one of the small drops in the values of F_S that were observed prior to onset. Figure 5b also shows that the length, l, of these events progressively increases with each successive event.

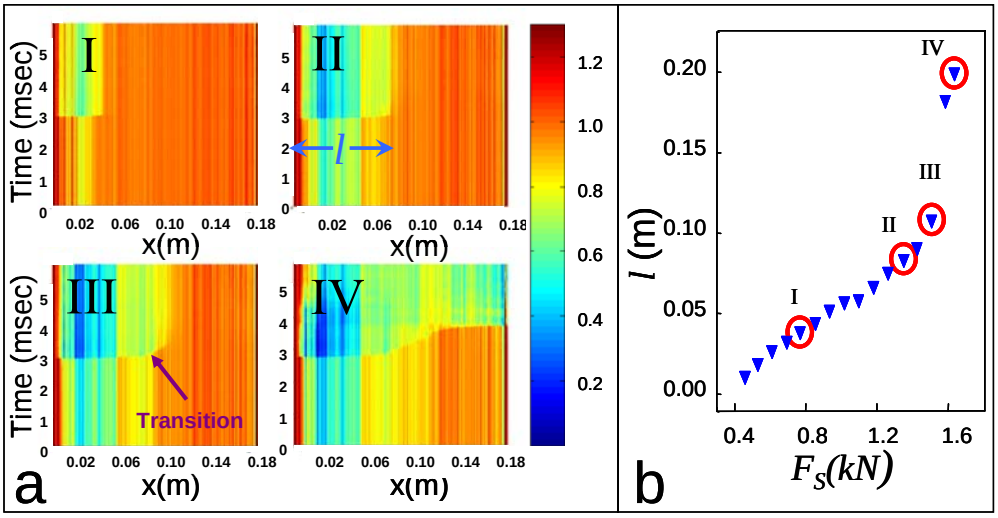

Figure 6. (a). The short-time temporal evolution of $A(x,t)$, [normalized as in Figure. 5(a) by $A(x,0)$ at $F_S = 0$]. Lines are values of $A(x,t)/A(x,0)$ photographed at 14μs intervals for four different precursor events. The plots correspond to the 6 msec bracketing each precursor's propagation. The arrow marks the transition from a sub-Rayleigh front to a slow front in frame III. (b). The precursor length, l, as a function of F_S in a typical experiment. The events denoted by I, II, III and IV correspond to the precursor events shown in (a). Figure adapted from (44).

In Figure 6a we take a closer look at the short-time dynamics of the events highlighted in Figure 5b. Looking now at the dynamics at μsec time scales (each line in the figure is a snapshot of $A(x,t)$ taken at a 14μsec interval), we see that each of the initial arrested events (labelled I and II in the figure) corresponds to an arrested sub-Rayleigh front, whose maximal velocities ranged to about V_R. In Figure 6b we present the lengths, l, of each event within this sequence as a function of the applied value of F_S. The length of the first events in this sequence increased in an approximately linear fashion with F_S. Among these are the two events labelled I and II in Figure 6a. As the transition to sliding motion is approached (in Figure 6 the transition takes place at the event labelled IV), the linear scaling of the precursor lengths breaks down. This break-down from linear to accelerated growth of l is a general feature of all of the event sequences that we have measured. Surprisingly, only when this deviation from linear scaling takes place is a transition to slow fronts from the rapid sub-Rayleigh precursors observed[44]. The transition is shown in the panel labelled III in Figure 6a. In the transition region (the region between III and IV in Figure 6b) slow fronts do not traverse the entire interface, but instead arrest after propagating a short distance. Within the transition region, the sub-Rayleigh fronts that trigger the slow fronts do not increase in length; each sub-Rayleigh front excites a slow front at the same location and the slow front propagates over increasingly longer distances. Sliding occurs when a slow front either continues to the end of the sample or (as in Figure 4) triggers an additional sub-Rayleigh front that reaches the leading edge.

The linear scaling region that accompanies the arrested sub-Rayleigh fronts (e.g. events I and II in Figure 6) is a general feature that is always observed

when edge-loading is applied. This region has been shown[44] to scale with the total length, L, of the interface, F_N and F_S as: $l \propto L \cdot F_S / F_N$. A way to interpret this scaling relation is that the local value shear stress (i.e. the mean shear stress over the length of the precursor), $\sigma_S = F_S / l$, is proportional to the mean normal stress over the entire interface, $\sigma_N = F_N / L$. The scaling is, in a loose sense, equivalent to a "local" Amontons-Coulomb law; $\sigma_S \propto \sigma_N$. In our experiments, the scaling region prevails until l becomes approximately half of the system size. We currently understand this breakdown of scaling to signify that the fronts start to "feel" the finite size of the system at $l \sim L/2$, and the effect of the far boundary (at the leading edge) may somehow be involved in the excitation of the slow fronts and eventual transition to overall sliding.

The $\sigma_S \propto \sigma_N$ scaling does not, however, explain the discrete nature of the precursors. Our measurements have shown[44] that the discrete shear force intervals, ΔF_S, between precursors scale linearly with the height, h, above the interface where F_S is applied. Interestingly, the value of h affects neither the scaling of l, nor the location of the transition region to sliding. Therefore, reducing the value of h will decrease the force intervals between successive precursors and correspondingly *increase* the *number* of precursors before the initiation of sliding.

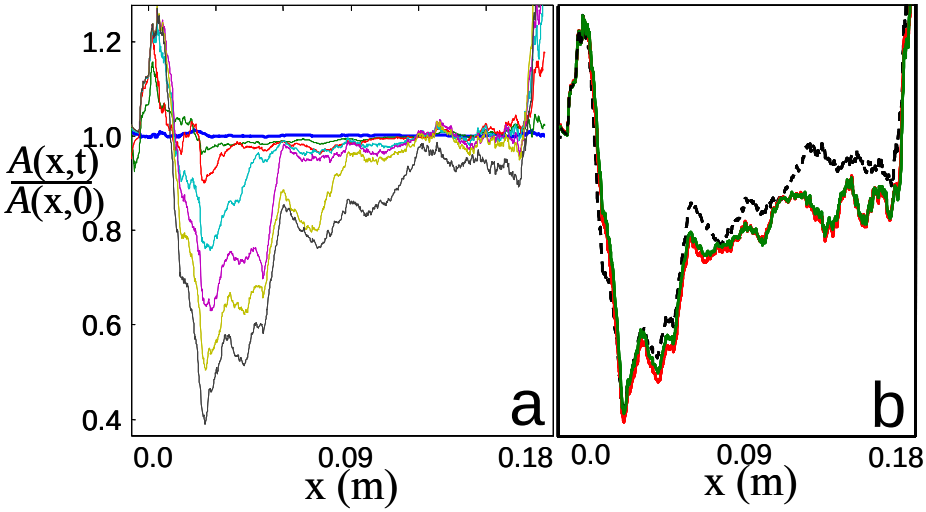

Figure 7. The discrete precursor sequences significantly alter the contact area profile, creating a strong "barrier" adjacent to the edge and a weakened region beyond it. (a) *A(x,t)* profiles subsequent to every third precursor presented in Figure. 6b. (b) Three *A(x,t)* profiles, one taken immediately prior to sliding (dotted line) and two immediately following sliding in consecutive stick-slip events of approximately 50μm in length, show that *A(x)* profiles are not modified during successive slip-stick events. Figure adapted from (44).

In the preceding subsection we have described how a discrete sequence of rapid, crack-like precursors, launched from the edge at which shear is applied, leads up to the transition to motion. We have also seen that the details of how the shear load is applied (i.e. the size of h) have a large effect on the size and periodicity of these precursors. How can we understand the role of both the

loading and details of the loading in governing the dynamics of this process?Insight to the origins of these dynamics is obtained by taking a closer look at the contact area profile, *A(x,t)*. In Figure 7a we present a number of individual *A(x,t)* profiles, each taken at 3 precursor intervals, in the experiment shown in Figure 5a. The experiments start from a uniform contact profile and, with each successive precursor the contact profiles become increasingly more non-uniform. The profiles have a number of characteristic features. Near the trailing edge (at x=0) a localised increase in *A(x,t)* is observed. The height of this strong patch ("barrier") increases with each successive precursor, but its width remains constant. We find that the spatial extent of this barrier is proportional to *h*. This is, perhaps, not surprising since *h* is the system's only relevant length scale (besides the overall interface length, *L*). Beyond this barrier, each precursor reduces *A(x,t)* until arresting at length *l*. The spatial extent of reduced *A(x,t)* region becomes successively greater with each precursor. In addition, with the passage of each successive precursor, the overall depth of the reduced contact area increases. In other words, each precursor serves to redistribute the contact area in the regions traversed. *A* is increased in a region of extent *h* adjacent to the trailing edge but significantly reduced beyond this region. During this entire process, the overall contact area, however, remains nearly constant[44], remaining within 90-100% of its initial integrated value.

It is interesting to see how the contact area profile, *A(x,t)*, is affected once the transition of overall motion has occurred. Intuitively, one would expect that once large-scale sliding of the interface commences, the contact surface would be renewed and *A(x,t)* would, once again, become uniform. Profiles of *A(x,t)* are presented in Figure 7b for successive stick-slip sliding events in which the relative motion in each event was a many mean contact points (~50μm) in length. Surprisingly, there is virtually *no* change in the shape of *A(x,t)*. Once the contact area profile is established by the precursor sequence, it is retained by the system in subsequent motion. The above results suggest that non-uniformity of *A(x,t)* is, in some sense, the preferred state of the system during sliding.

A model for precursor dynamics

The behavior described above suggest the following intuitive picture for the precursor dynamics. Since F_S is applied at the trailing edge and, initially, all the interface is pinned, the shear stress can only increase within the medium in a finite range surrounding the loading point. This range is determined by the scale *h*. Thus, a region of spatial extent *h* from the trailing edge finds itself under locally high stresses ($\sigma_S \sim F_S/h$), while the remainder of the interface is relatively unaffected. As F_S is increased, the stresses in the 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 – since the stresses beyond the area adjacent to the trailing edge are not under sufficient stress to precipitate fracture. Since at the tip of even an arrested crack a stress singularity exists, the area surrounding the tip of the arrested 'crack' is under high shear. This shear cannot

diffuse away, because the frictional contacts along the "broken" region of the interface rapidly re-establish and effectively entrap these residual stresses.

As a result of this process the trailing edge undergoes a slight frictional motion and is displaced from its original position. On the other hand, the interface remains pinned beyond the point where the crack arrested. 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 causes the increased contact area (the "barrier" in Figure 7) of scale h adjacent to the trailing edge together with the region of reduced contact area in its wake, as shown in Figure 7.

Once the first precursor has propagated, the contact area adjacent to the edge is once again re-formed and the previously high shear stresses were relieved by the precursor's passage. As F_S is further increased, the process will repeat itself at 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 further elastic deformation of the slider, thereby both strengthening the barrier adjacent to the trailing edge and further weakening the region beyond. As F_S is still further increased, this process continues as successive cracks extend into the interface. Since 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 (i.e. $F_S \sim \mu_S F_N$).

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 we use. In our experiments we find that the strains imposed during the precursory sequence can only be fully relieved if the normal forces are entirely relaxed.

Ramifications for earthquake dynamics

The work described here was performed in the laboratory at scales that are 10-20cm in size. It is never the less tempting to use these measurements to obtain insight about processes related to earthquake dynamics that occur along rough interfaces that can reach hundreds of kilometres in length. In the case of earthquakes, sliding (overall motion along the fault) very rarely occurs. Slip along such an extended interface is limited to sections of the faults that may extend tens to hundreds of kilometers. Is this huge extrapolation of scales justified? It has long been thought that earthquakes result from the frictional stick-slip motion along faults that form the frictional boundary between slowly creeping tectonic plates[6, 45]. Earthquakes take place over segments where elastic energy, stored within scales ranging over 1-100km, is released due to the slip of relatively smooth faults (where it is not clear what the relevant dissipative scales are). This

large separation of scales in earthquakes is somewhat mirrored in the experiments, as the laboratory system's size (~20cm) is 5 orders of magnitude larger than the characteristic features within the interface (~1μm).

The results described above lend support to these ideas. We have shown that the transients that lead to sliding are entirely governed by the motion of fronts that propagate across the frictional interface. The sub-Rayleigh fronts observed in our experiments are analogous to the vast majority of earthquakes, which generally propagate at velocities of about $0.8V_R$. We have observed fronts that also propagate at intersonic velocities in our experiments. Do these have analogs along a fault? For many years there did not seem to be any evidence for such rapid earthquakes. Shortly after the first experimental observations of intersonic cracks[27] in 1999, evidence for intersonic earthquakes started to appear[46]. Although still somewhat controversial, these observations are now taken seriously in the geophysical community[47-49].

We have also demonstrated that, nearly every time large-scale sliding occurs, slow fronts are observed. Are there analogs to these fronts in earthquake dynamics? The answer to this question is still unclear. There are a variety of slow phenomena that have recently been observed. As our main source of information in the study of fault dynamics is by means of seismic sensing, these "slow" phenomena had, until recently escaped detection since they have an extremely small seismic signature. Recent use of GPS sensors, strainmeters and applications of satellite imagery to geophysically active regions have now given us the capability to accurately measure displacements that occur along faults. These measurements have uncovered a host of slow processes along active faults in which significant "aseismic" slip has been detected. These events are currently the subject of active study in the geophysical community, but there is still no consensus of either what they signify or even how to best categorize them. Slow slip events in the literature[50-57] include phenomena coined "slow" earthquakes, "silent" earthquakes, and "aseismic tremors".

This intriguing situation brings us back to our experiments. What can we learn from them about earthquake dynamics? The answer is that, potentially, quite a lot. The key difficulty in unravelling earthquake dynamics is that earthquakes occur at seismogenic depths that range from tens of kilometers beneath the earth's surface. Our main source of information about source dynamics of earthquakes comes from the inversion of seismic signals. This technique is currently the best that we can do, but it is extremely limited. Seismic inversion essentially involves reconstruction of earthquake source dynamics from a number of discrete acoustic measurements. The measurements themselves are highly bandwidth limited (as all high frequencies are highly attenuated within the earth, before they can be detected by surface measurements) where the medium through which they propagate is far from ideal. Under these conditions, the best spatial resolution that one can currently achieve in large earthquakes is localization of a source to within 3 km, with significant uncertainties due to the large degree of heterogeneity within the host media through which seismic signals propagate.

Experiments such as the ones described here are not subject to these limitations. Slip events can be directly imaged and quantitative measurements of other key quantities (e.g. slip, high-bandwidth acoustics, strains and stresses) can be easily obtained. On the other hand, these are extremely simplified systems (e.g. using highly controlled loading of single, perfectly uniform faults) and it is entirely possible that important features that are critical to earthquake dynamics are omitted. The challenge is to perform these real-time measurements, make sense of them, and then to relate them to known phenomena within the earth. We will then be in a realistic position to determine what features are important to what types of processes.

Summary

We have reviewed a number of recent measurements of short-time phenomena that give rise to frictional motion. We have seen that the transition from static to dynamic friction is mediated by the interplay of a number of qualitatively different "detachment" fronts that live within the contact plane that separates two bulk materials under shear. We have seen that the main "players" among these fronts (i.e. intersonic, sub-Rayleigh, and slow fronts) are *always* observed in our experiments. Their existence is not dependent on the loading conditions that are imposed on the system.

We then considered how frictional motion occurs for a *specific* but still quite general loading configuration; when shear is imposed at an edge of the system. For this type of loading, we showed that a specific sequence of events takes place that "prepares" the system for eventual macroscopic motion by transporting the imposed shear across the interface in a well-defined way. The vehicle for the transport of shear is a discrete sequence of sub-Rayleigh precursors, each excited at the trailing edge of the system and each extending further into the interface before arresting. These precursory events are excited with imposed shear forces that are significantly below the shear levels needed to drive frictional sliding. With their passage, these events precipitate a large and systematic redistribution of the contact area along the interface so that, when sliding finally occurs, the initially uniform contact area profile is highly non-uniform. In this configuration, the transition to sliding only takes place when slow fronts are excited in the system. We surmise that this transition may be related to the influence of the system's finite size. We have also shown that, surprisingly, upon the onset of stick-slip motion, the contact area profile is not renewed by the sliding but, rather, the profile that is built-up by the initial precursor sequence is retained.

Acknowledgments

The authors acknowledge the support of the Israel Science Foundation (FIRST Grant No. 1116/05) as well as the US-Israel Binational Science foundation (Grant No. 2006288).

References

1. R. W. Carpick, D. F. Ogletree, M. Salmeron, *Journal of Colloid and Interface Science* **211**, 395-400 (1999).
2. J. H. Dieterich, *Tectonophysics* **211**, 115-134 (1992).
3. N. Lapusta, J. R. Rice, *Journal of Geophysical Research-Solid Earth* **108**, 2205 (2003).
4. Y. Ben-Zion, *Journal of the Mechanics and Physics of Solids* **49**, 2209-2244 (2001).
5. M. Ohnaka, *Earth Planets and Space* **56**, 773-793 (2004).
6. C. H. Scholz, *Nature* **391**, 37-42 (1998).
7. B. D. Thompson, R. P. Young, D. A. Lockner, *Geophysical Research Letters* **32**, L10304 (2005).
8. B. Q. Luan, M. O. Robbins, *Nature* **435**, 929-932 (2005).
9. E. Gerde, M. Marder, *Nature* **413**, 285-288 (2001).
10. M. Urbakh, J. Klafter, D. Gourdon, J. Israelachvili, *Nature* **430**, 525-528 (2004).
11. B. N. J. Persson, *Sliding Friction Physical Principles and Applications* (Springer-Verlag, New York, ed. 2, 2000).
12. J. R. Rice, A. L. Ruina, *Journal of Applied Mechanics* **50**, 343-349 (1983).
13. A. Ruina, *Journal of Geophysical Research* **88**, 359-370 (1983).
14. J. Dieterich, *J.Geophys. Res.* **84**, 2161-2168 (1979).
15. C. Marone, *Annual Review of Earth and Planetary Sciences* **26**, 643-696 (1998).
16. B. D. Kilgore, M. L. Blanpied, J. H. Dieterich, *Geophysical Research Letters* **20**, 903-906 (1993).
17. T. Baumberger, P. Berthoud, C. Caroli, *Physical Review B* **60**, 3928-3939 (1999).
18. C. Caroli, T. Baumberger, L. Bureau, *Journal De Physique Iv* **12**, 269-273 (2002).
19. F. P. Bowden, D. Tabor, *The Friction and Lubrication of Solids* (Oxford Univ. Press, New York, ed. 2, 2001).
20. J. H. Dieterich, B. D. Kilgore, *Pure and Applied Geophysics* **143**, 283-302 (1994).
21. A. Ovcharenko, G. Halperin, I. Etsion, *Wear* **264**, 1043-1050 (2008).
22. L. Bureau, T. Baumberger, C. Caroli, *European Physical Journal E* **19**, 163-169 (2006).
23. S. Rubinstein, G. Cohen, J. Fineberg, *Physical Review Letters* **96**, 256103 (2006).
24. L. B. Freund, *Dynamic Fracture Mechanics* (Cambridge, New York, 1990).
25. J. Fineberg, M. Marder, *Physics Reports-Review Section of Physics Letters* **313**, 2-108 (1999).
26. A. J. Rosakis, O. Samudrala, R. P. Singh, A. Shukla, *Journal of the Mechanics and Physics of Solids* **46**, 1789-1813 (1998).
27. A. J. Rosakis, O. Samudrala, D. Coker, *Science* **284**, 1337-1340 (1999).

28. H. J. Gao, Y. G. Huang, F. F. Abraham, *Journal of the Mechanics and Physics of Solids* **49**, 2113-2132 (2001).
29. A. Needleman, *Journal of Applied Mechanics-Transactions of the Asme* **66**, 847-857 (1999).
30. S. M. Rubinstein, G. Cohen, J. Fineberg, *Nature* **430**, 1005-1009 (2004).
31. S. M. Rubinstein, M. Shay, G. Cohen, J. Fineberg, *Int. J. of Fracture* **140**, 201-212 (2006).
32. L. B. Freund, *J. Geophys. Res.* **84**, 2199-2209 (1979).
33. A. J. Rosakis, O. Samudrala, D. Coker, *Materials Research Innovations* **3**, 236-243 (2000).
34. K. W. Xia, A. J. Rosakis, H. Kanamori, *Science* **303**, 1859-1861 (2004).
35. Z. Reches, D. A. Lockner, *Journal of Geophysical Research-Solid Earth* **99**, 18159-18173 (1994).
36. M. L. Falk, J. S. Langer, *Physical Review E* **57**, 7192-7205 (1998).
37. J. Rottler, M. O. Robbins, *Physical Review Letters* **95** (2005).
38. N. Lapusta, J. R. Rice, Y. Ben-Zion, G. T. Zheng, *Journal of Geophysical Research-Solid Earth* **105**, 23765-23789 (2000).
39. M. Ohnaka, *Pure and Applied Geophysics* **161**, 1915-1929 (2004).
40. H. Kanamori, G. S. Stewart, *Journal of Geophysical Research* **83**, 3427-3434 (1978).
41. S. Das, *Pure and Applied Geophysics* **160**, 579-602 (2003).
42. M. Ohnaka, L. F. Shen, *Journal of Geophysical Research-Solid Earth* **104**, 817-844 (1999).
43. S. L. Ma, C. R. He, *Tectonophysics* **337**, 135-145 (2001).
44. S. M. Rubinstein, G. Cohen, J. Fineberg, *Physical Review Letters* **98**, 226103 (2007).
45. W. F. Brace, J. D. Byerlee, *Science* **153**, 990-& (1966).
46. M. Bouchon *et al.*, *Geophysical Research Letters* **28**, 2723-2726 (2001).
47. B. T. Aagaard, T. H. Heaton, *Bulletin of the Seismological Society of America* **94**, 2064-2078 (2004).
48. E. M. Dunham, *Journal of Geophysical Research-Solid Earth* **112** (2007).
49. Y. Liu, N. Lapusta, *Journal of the Mechanics and Physics of Solids* **56**, 25-50 (2008).
50. G. C. Beroza, T. H. Jordan, *Journal of Geophysical Research-Solid Earth and Planets* **95**, 2485-2510 (1990).
51. R. E. Abercrombie, G. Ekstrom, *Journal of Geophysical Research-Solid Earth* **108** (2003).
52. M. R. Brudzinski, R. M. Allen, *Geology* **35**, 907-910 (2007).
53. W. L. Ellsworth, G. C. Beroza, *Geophysical Research Letters* **25**, 401-404 (1998).
54. P. F. Ihmle, T. H. Jordan, *Science* **266**, 1547-1551 (1994).
55. H. Kanamori, *Proceedings of the Japan Academy Series B-Physical and Biological Sciences* **80**, 297-316 (2004).
56. T. I. Melbourne, F. H. Webb, *Science* **300**, 1886-1887 (2003).
57. M. M. Miller, T. Melbourne, D. J. Johnson, W. Q. Sumner, *Science* **295**, 2423-2423 (2002).