A MODEL FOR THE NUCLEATION OF EARTHQUAKE SLIP

James H. Dieterich

U.S. Geological Survey, Menlo Park, California 94025

Abstract. The nucleation of unstable slip on faults with state variable constitutive properties has the following characteristics. First, unstable slip may initiate only on patches exeeding a critical radius, r_c. Below the critical radius slip is always stable. The critical radius is a function of normal stress, loading conditions and constitutive parameters which include D_c, the characteristic slip distance. Second, for a patch larger than the critical radius, slip accelerates to instability if the normalized stress, μ, exceeds the steady-state friction μ_s. Over a wide range of conditions, following application of the initial stress, the logarithm of the time to instability linearly decreases as the difference $\mu - \mu_s$ increases. Using laboratory derived constitutive parameters extrapolated to the time intervals characteristic of earthquake faulting (10^7 seconds or longer) the time delay from the application of μ to the time of instability varies from seconds to times on the order of the interevent time. It is speculated that this delayed failure may be an important process in controlling the timing of a mainshock following a foreshock, or in controlling aftershock sequences. The familiar $1/t$ decay in rate of aftershock occurrence is satisfied by this model assuming a uniform distribution of initial stresses, $\mu - \mu_s$, on the population of nucleation patches for the aftershocks. Displacements during the accelerating slip that leads to instability are proportional to D_c. Unless D_c for earthquake faults is significantly greater than that observed on simulated faults, premonitory displacements prior to earthquakes may be too small to detect using current strain observation methods.

Introduction

Slip instability on faults with rate and rate-history constitutive properties has been the subject of analyses by Dieterich [1979a, 1981]; Ruina [1980, 1983]; Rice and Ruina [1983]; Rice [1983]; Gu et al. [1984]; and Weeks and Tullis [1985]. A result these studies have in common is that slip instability does not occur instantaneously at some threshold stress—an interval of accelerating slip always preceeds instability. This paper discusses the possible sequence of events that must occur to satisfy the conditions necessary for instability on faults with these constitutive properties and then examines the characteristics of the interval of accelerating slip that precedes the instability. The results are believed relevant to understanding the scaling of possible earthquake premonitory phenomena and in explaining some of the temporal characteristics of earthquake occurrence. Of particular interest for the latter are the mechanisms causing foreshocks and aftershocks. Because the intent is to explore the importance of the fault constitutive properties in controlling the characteristics of earthquake occurrence, other aspects of fault interactions have been simplified to the maximum extent considered reasonable.

Constitutive Law for Faulting

Laboratory study of slip on simulated faults has demonstrated the existence of time-, velocity- and history-dependent processes that perturb fault strength [e.g., Dieterich, 1978a, 1981; Scholz and Engelder, 1976; Ruina, 1980, 1983; Weeks and Tullis, 1985]. Other features of the laboratory observations include: displacement weakening, over a distance that scales by the parameter D_c, at the onset of slip if the fault is previously stationary; and time-dependent recovery of fault strength following an episode of rapid fault slip [Dieterich, 1972, 1981]. The displacement scaling parameter D_c, varies from a few microns for semipolished surfaces [Dieterich, 1978a; Ruina, 1980] to about 50 microns for roughened surfaces separated by a thin layer of crushed rock [Dieterich, 1981]. Throughout this paper fault displacements, d, and slip velocity, v, are normalized by the characteristic slip distance D_c, and are indicated by capitalized letters:

$$D = d/D_c, \qquad V = v/D_c \tag{1}$$

The experimental observations may be represented by constitutive laws that incorporate slip rate and rate-history effects [Dieterich, 1979a, 1981; Ruina, 1980, 1983; Rice, 1983; Gu et al. 1984]. In this approach, sliding history effects are represented by a state variable that evolves with displacement toward a steady state value that is governed by the instantaneous slip speed. Several specific forms for the constitutive laws have been employed in the studies cited above. The various constitutive formulations differ in detail, depending largely on approximations employed, but all share the same approach and provide very similar representations of the data. For example, useful simplifying approximations to the relations employed by Dieterich [1979a, 1981] are introduced by Ruina [1980, 1983] by not representing the saturation of the slip rate and history effects at high slip velocity. The results of Ruina [1980] and Weeks and Tullis [1985] suggest that under some conditions the data are better represented by more than one state variable. Some relationships between constitutive laws and comparison of data from dynamic slip events are discussed by Okubo and Dieterich [1986].

In the following discussion, the coefficient of friction, μ, is defined:

$$\mu = \tau/\sigma \tag{2}$$

where τ is the shear stress acting across the fault and σ is the normal stress. Throughout this work, σ is considered to be constant during slip and multiplication of μ by σ is implied when reference is made to fault shear stress. The following single state variable constitutive law for μ has been employed for the study reported here [see Okubo and Dieterich, 1986]:

$$\mu = \mu_0 + B_1 \ln(B_2\theta + 1) - A_1 \ln[(A_2/V) + 1] \tag{3}$$

where μ_0, A_1, A_2, B_1, and B_2 are experimentally determined parameters, V is the normalized slip speed and θ is the state variable that depends on slip history. The state variable is discussed further below. It is noted that (3) is equivalent to that of Dieterich [1979, 1981] with the distinction that the earlier forms employ somewhat awkward quotients that are well represented by expansion to (3). The summing of $B_2\theta$ and A_2/V with 1 is done to represent the saturation of the rate and rate-history effects at high slip speeds. See Okubo and Dieterich [1986] for a discussion of the rate limits. At V and θ well removed from the rate limits (i.e., $B_2\theta \gg 1$, $A_2/V \gg 1$) equation (3) is exactly equivalent to that employed by Ruina [1980] and later by several other investigators. That is:

$$\mu = \mu_0' + \Theta + A_1 \ln V \tag{4}$$

where: $\Theta = B_1 \ln \theta$ and $\mu_0' = \mu_0 + B_1 \ln B_2 - A_1 \ln A_2$.

The variable θ represents sliding history effects and consequently incorporates displacement dependence. θ has been interpreted [Dieterich, 1979a; Dieterich and Conrad, 1984] as the average age of the load supporting contacts between the sliding surfaces. A number of equations for the evolution of θ have been employed in studies reported in the literature, with similar results. The following evolution law discussed by Ruina [1980] and employed by Dieterich [1981] is used here:

$$d\theta/dt = 1 - \theta V \tag{5}$$

Hence, at steady state slip, where slip speed is constant for large displacements:

$$d\theta/dt = 0, \qquad \theta = 1/V \tag{6}$$

For a nominally stationary fault in which the contacts are not disturbed by displacement, $d\theta/dt = 1$ and θ increases with the time of stationary contact. Consequently, for conditions of periodic earthquake slip separated by long intervals of nominally stationary fault contact, θ is approximately the time since the last slip event. In practice this is only an upper limit for θ because some amount of creep always occurs in simulations of the interevent period that tends to reduce θ. Below, initial values of θ in the range 10^5–10^9 s are employed for calculations of the processes leading to slip instability. The larger values of θ are considered representative of the conditions on a fault prior to earthquake recurrence. Under conditions of constant velocity (5) can be solved for θ using the chain rule to replace $d\theta/dt$ with $d\theta/V\,dD$ yielding:

$$\theta = 1/V + (\theta_0 - 1/V)e^{-D} \tag{7}$$

where θ_0 is the value of θ at $D = 0$. Below, use is made of (7) in numerical computations by treating slip as a series of small constant velocity steps.

Conditions for Slip Instability

It is assumed for the remainder of the paper that conditions are well removed from the effects of the saturation of θ that occurs at small contact times (i.e., it is assumed that $B_2\theta \gg 1$). This assumption is reasonable because the experiments indicate that $B_2 > 1$. For the conditions of interest recall that θ is greater than the time since the last fault slip. Therefore, $\theta \gg 1$. Hence, in the following the term $(B_2\theta + 1)$ of (3) is replaced by θ, with a suitable change in the constant μ_0 as given above.

Some characteristics of the constitutive relations are useful for discussing the nucleation of unstable fault slip. Note first that the maximum limit of the frictional strength, $\mu_{\max}$, occurs when $V \gg A_2$ with the result that (3) becomes:

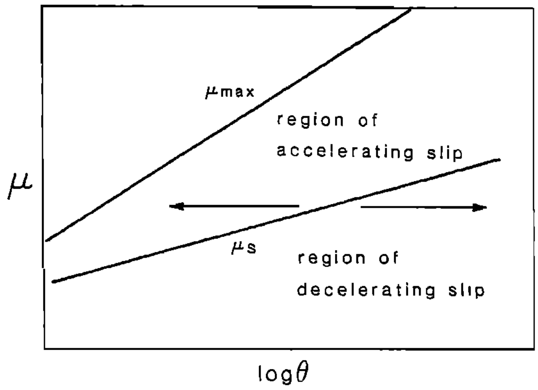

Fig. 1. Plot of maximum friction, $\mu_{\max}$, and steady state friction, μ_s, against state variable θ. This plot assumes that $B_2\theta \gg 1$. Under conditions of constant stress, μ, slip velocity and θ will remain constant only at $\mu = \mu_s$. Under conditions of $\mu > \mu_s$ and constant stress, velocity increases and θ decreases with time until $\mu = \mu_{\max}$, at which time instability occurs. At $\mu < \mu_s$, under conditions of constant stress, velocity decreases and θ increases. The arrows indicate the direction in which θ evolves away from μ_s.

$$\mu_{\max} = \mu_0 + B_1 \ln \theta \tag{8}$$

As μ approaches $\mu_{\max}$, V approaches infinity. Also note that in general, frictional strength, μ is uniquely defined only if both θ and V are known. However, at steady state slip, $\theta = 1/V$, which yields the steady-state friction

$$\mu_s = \mu_0 - B_1 \ln V - A_1 \ln(A_2/V + 1) \tag{9a}$$

or equivalently:

$$\mu_s = \mu_0 + B_1 \ln \theta - A_1 \ln(A_2\theta + 1) \tag{9b}$$

A dynamic slip instability occurs when the frictional strength decreases at a rate that exceeds the capability of the applied stress to follow (i.e., the applied load exceeds the frictional resistance). Below $\mu_{\max}$, if the friction is less than the applied load, then the system needs only to slide at a higher rate to balance the friction to the applied load and stabilize the slip. This is because of the direct velocity-dependent term in the constitutive relation. Therefore, under quasi-static conditions, instability can only occur at the instant the applied stress reaches the limiting strength $\mu_{\max}$. Unless the stress state steps instantaneously to $\mu_{\max}$ there is an interval of accelerating slip rate immediately prior to the onset of instability as the friction approaches $\mu_{\max}$. Additionally, for an instability to occur, the stiffness of the system loading the fault at the point of instability must be less than a critical value. At stiffnesses greater than the critical stiffness the displacements that occur during accelerating slip decrease the applied stress at a rate that is sufficient to move the applied stress away from $\mu_{\max}$ which terminates the acceleration.

Under conditions of constant μ (corresponding to zero stiffness) in the range $\mu_s < \mu < \mu_{\max}$, the state variable θ will continuously decrease with displacement resulting in accelerating slip. The decrease of θ with displacement may be seen by noting that: when $\mu > \mu_s$, then $V > 1/\theta$ (equations (3) and (9)) and consequently from (5) $d\theta/dt < 1$. Because μ_s and $\mu_{\max}$ depend upon θ (equations (8) and (9)), the accelerating slip continues until either μ equals μ_s or $\mu_{\max}$ at which point the slip will become either steady state slip or unstable, respectively (Figure 1). Note from Figure 1 that at $\mu > \mu_s$ slip will always accelerate to instability when $B_1 > A_1$. This result was noted by Dieterich [1979a] and proved by Ruina [1980]. From this it is seen that μ_s and $\mu_{\max}$ define the range of frictional stresses in which slip may accelerate to instability. Below μ_s slip rates decelerate under conditions of constant or decreasing stress and sliding is intrinsically stable.

A general result of investigations of the stability of simple spring and slider systems [e.g., Dieterich 1979a, 1981; Ruina, 1983; Rice and Ruina, 1983; Gu et al., 1984] is that slip will accelerate to instability if the stiffness of the spring, k, is less than a critical value, k_c, given by:

$$k_c = \xi\sigma/D_c \tag{10}$$

Where ξ is parameter that depends on the constitutive parameters and the conditions of the experiment. For steady state slip at $\mu = \mu_s$, Ruina [1980, 1983] shows that perturbations will grow to instability at stiffness less than a critical value. For the constitutive relations (3) and (5) the Ruina result is:

$$\xi = (B_1 - A_1) \tag{11}$$

for steady state sliding removed from the saturation of the θ and V dependence (i.e., $B_2\theta \gg 1$, $A_2/V \gg 1$). Note from (11) that for systems with finite positive stiffness, instabilities growing from perturbations about μ_s can occur only when $B_1 > A_1$. For sliding above the steady state friction ($\mu > \mu_s$) equation (11) is the lower bound for the critical system stiffness. Gu et al. [1984] examine the conditions for instability with $\mu > \mu_s$ under various loading conditions.

The results for a spring slider system may be applied to approximately determine the conditions for stability of a fault patch embedded in an elastic media. We consider the simple case of slip on a circular fault patch of radius,

θ, subject to a uniform change of shear stress, $\Delta\tau$. The displacement at the center of the patch due to $\Delta\tau$ is greater than any other point on the patch and is given by:

$$d = \frac{24\Delta\tau r}{7\pi G} \tag{12}$$

where G is the shear modulus, [Chinnery, 1969]. The effective stiffness of every point on the patch is $\Delta\tau$ divided by the local displacement. Because the displacement is maximum at the center point, the effective stiffness of the center is less than any other point. Below it is demonstrated that a low stiffness system tends to reach instability before a higher stiffness system under equivalent initial conditions. Consequently, the center point is assumed to control the onset of instability for the entire patch. From (12) the effective stiffness of the center of the patch:

$$k = \frac{7\pi G}{24r} \tag{13}$$

Combining (10) and (13) yields an approximate relationship for the minimum critical crack radius for unstable fault slip:

$$r_c = \frac{7\pi G D_c}{24\sigma\xi} \tag{14}$$

Similar approaches for displacement weakening cracks are given by Andrews [1976] and Day [1982] and for displacement weakening arising from rate and state dependence by Dieterich [1979a].

The result of equation (14) indicates an interval of premonitory stable fault slip will occur on faults with heterogeneity that is independent of the premonitory slip seen in the single degree of freedom spring-slider systems. For example, consider a fault with uniform frictional properties and heterogeneous shear stress. The constitutive relation (3) indicates that every point on the fault will slip at some velocity depending upon θ and the applied stress. However, because the coefficient A_1 is small, very large variation in slip rates arise from small variations in stress. Therefore in terms of the stress state and relative displacement rates we treat some sections of a fault as being essentially locked while others are slipping. Patches with stress at or above μ_s will slide stably as long as $r < r_c$. A patch will enlarge as the stress at the edge of the patch increases due to increasing tectonic stress and stress redistribution from slip on the patch. This slip is intrinsically stable until $r > r_c$. At that point the slip may accelerate to instability if $\mu > \mu_s$ without further increase of the stress on the fault.

These considerations suggest that two distinct phases of stable slip can be identified leading to confined unstable slip on a laterally varying fault. The first is associated with the growth of the slipping patch to a radius exceeding some critical radius. Until the radius is reached, slip is intrinsically stable and is largely externally driven. The second phase consists of the interval of accelerating slip that may begin on the patch only if $r > r_c$ and $\mu > \mu_s$. These patches with $r > r_c$ and $\mu > \mu_s$ are the sites at which unstable slip can eventually nucleate. The large stress drop that occurs at the time of the instability on the nucleating fault patch may subsequently drive the instability to dimensions in excess of the dimensions of r_c.

Stable slip on a fault patch has been observed experimentally [Dieterich, 1978b; Dieterich et al., 1978] and has been analyzed, in part, with a 2-D numerical model [Dieterich, 1979b]. Those studies show the existence of two readily distinguished phases of premonitory slip. The first phase is associated with the growth of a stably sliding fault patch. The growth of the patch is controlled by the magnitude of the stress/strength variations on the surface and by the rate of loading on the fault. In situations where the dimensions of the minumum nucleation zone exceed the dimensions of the experimental fault, the second phase always begins after slip has propagated along the entire fault [Dieterich, 1979b]. In this case the experimental apparatus and the fault behave approximately as a spring slider system with the stiffness, K, determined by the machine stiffness. In large scale tests on a 2-m fault, the expected minimum size for the nucleation zone is less than the dimensions of the fault. In this case rapidly accelerating slip may begin before the growing patch reaches the ends of the sample and often the slip instability does not rupture to the ends of the fault [Dieterich et al., 1978]. Where the embedded creeping patch nucleates unstable slip, the stiffness of the system is given by the parameters of the slipping patch (as in (13)) and not by the apparatus.

The interval of accelerating slip on a patch with $r > r_c$ is now considered in more detail. Unlike previous studies cited above which emphasize stability criteria, the following treatment emphasizes: 1) the scaling of the magnitude and form of the premonitory displacements that occur during the interval of accelerating fault slip, and 2) the factors that control the duration of the period of accelerating slip. The former is of interest in attempting to assess the potential utilization of premonitory fault slip for earthquake prediction. The latter may be of interest for illuminating some factors that control the timing of earthquakes, particularly following a perturbation of stress on the nucleating patch. Each of these is discussed below.

Numerical Model

The model is represented as a single slider attached to spring with a constant stiffness, k. The fault obeys the constitutive law (3) and the evolution law of (5). Follow-

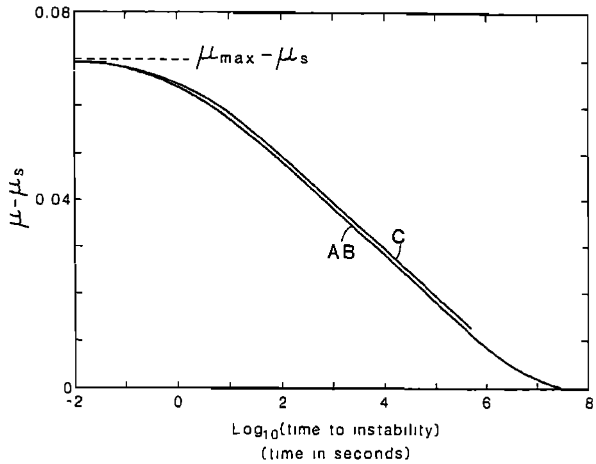

Fig. 2. Time to instability following a step in applied stress to an initial level given by $\mu - \mu_s$, for a fixed loading point ($V_l = 0$). This plot is generated from a series of numerical solutions, in which the initial stress for the computation was incremented through the range of initial stress from μ_s to $\mu_{\max}$. The constitutive parameters for these simulations were chosen to be representative of the experimental results: $B_1 = .015/2.3$, $A_1 = .010/2.3$, $A_2 = 1.0$. At the beginning of each simulation $\theta = 10^7$ seconds. The normalized stiffnesses are $K = 0$, $K_c/10$, $K_c/1.1$ for curves A, B and C, respectively. K_c is the normalized critical stiffness for instability arising from perturbation of steady state slip.

ing the discussion of (12)–(14) it is argued that this model provides a simple treatment for slip stability on a confined fault patch, assuming the zone of slip does not expand during the period covered by the computations and that the stress is uniform over the fault patch. The computations employ normalized displacements and velocity for the slider, D and V, respectively. Displacements and velocity of the load point attached to the spring, D_l and V_l, respectively are also normalized by D_c. Stress acting perpendicular to the sliding surface, σ, is constant during slip permitting the shear stress acting on the surface and fault frictional strength to be normalized by σ. The focus of interest for these calculations is the factors that control the duration and characteristics of the long interval of slow fault creep that precedes slip instability. Therefore, we assume quasi-static motions and the applied stress equals the friction at all times. The change in normalized shear stress, $\Delta\mu$, caused by displacement of the load point and slider is:

$$\Delta\mu = \frac{kD_c(D_l - D)}{\sigma} \qquad (15)$$

or:

$$\Delta\mu = K(D_l - D) \qquad (16)$$

where K is the normalized spring stiffness:

$$K = kD_c/\sigma \qquad (17)$$

The stiffness employed for the computations described below are expressed as a factor of the critical stiffness for steady state sliding, K_c, which is obtained from (10).

The computations employ a time marching procedure to follow the slip on the patch. Slip is treated as a series of constant velocity time steps utilizing the evolution law of (7). A numerical integration is performed to assure that the average frictional resistance is equal to the average stress applied by the spring during each time step. An instability occurs when the applied elastic stress exceeds $\mu_{\max}$.

Only stiffnesses less than the critical stiffness have been examined. In some cases it is expected that the radius of the patch may equal r_c when the stress is still well below μ_s, the minimum stress for self-driven acceleration to instability. In such cases growth of the patch will in all likelihood continue to increase and be greater than r_c by the time $\mu = \mu_s$. Also during the interval of accelerating slip the displacements on the patch will presumably drive additional growth of the patch. The interest here is in illuminating controlling processes, therefore crack radius and consequently stiffness are held constant during a computation. Below, several different stiffnesses have been examined to evaluate the possible effect of variable crack radius on the process.

For the purpose of this study the loading of the patch on the fault is treated as the sum of a constant rate of displacement of the loading point and discrete jumps in applied load. The former may be viewed as the model equivalent of the large-scale background tectonic loading rate of a fault. The latter corresponds to strain events that alter the strain field on the nucleating patch. An obvious example of a deformation event would be the occurrence of an earthquake at some other location on the fault, i.e., a foreshock or a mainshock. Another common source of a load step might be the rapid intrusion of magma in a volcanic region. Of particular interest below are the form and magnitude of displacements during accelerating slip under various loading conditions and alteration of the time to instability caused by a load step.

Results

Figure 2 gives the times to instability from a series of simulations in which the loading point was stepped to apply an initial stress, μ, in the interval $\mu_s < \mu < \mu_{\max}$. The stress step is plotted as $\mu - \mu_s$. Following application of the stress step the loading point is held fixed. The state variable θ in these simulations is set initially at

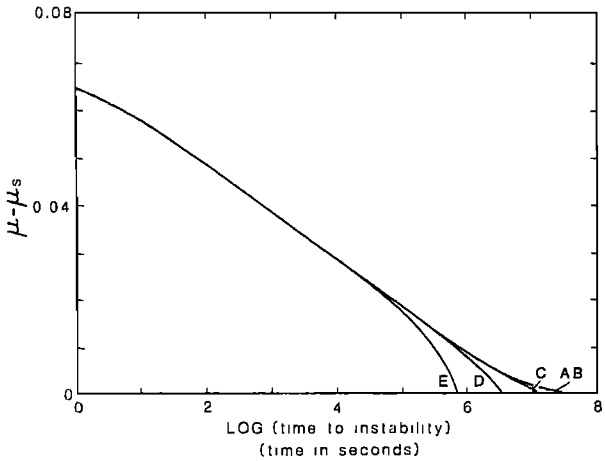

Fig. 3. Time to instability following a step in applied stress to an initial level given by $\mu - \mu_s$, for different loading point velocities. The constitutive parameters and initial value of θ are the same as that of Figure 2. The stiffness is $K = K_c/10$. The loading point velocities are $V_l = 0$, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, for the curves labeled A, B, C, D and E, respectively. For the initial value of $\theta = 10^7$ the steady state slip velocity is 10^{-7}.

10^7 s at the beginning of each simulation. This choice of θ is arbitrary. The value of θ is large to simulate the large interevent times for earthquake faults. Below, results are given showing the effect of θ on the time to instability. Constitutive parameters for the simulations are representative of the laboratory results reported by Dieterich [1981] with $\mu_0 = 0.6$, $2.3B_1 = 0.015$, $2.3A_1 = 0.010$, $A_2 = 1.0$. (Note this value of A_2 is arbitrary. The effect of A_2 on these results is to shift the time at which the curves bend to become asymptotic to $\mu_{\max}$.) Several features of the simulations of Figure 2 are of note. First, although the necessary conditions for the instability are satisfied when μ exceeds μ_s and $K < K_c$, the instability process is time-delayed. The time to instability is strongly dependent on the amplitude of the initial stress step. The time to instability goes to zero asymptotically as μ goes to $\mu_{\max}$ and the time to instability becomes very large, presumably approaching infinity, as μ approaches μ_s. Over a wide range of initial stresses the logarithm of the time to instability is proportional to the amplitude of the stress step $\mu - \mu_s$. Second, in the interval $\mu_s < \mu < \mu_{\max}$ the system may accelerate monotonically to instability without the necessity of further displacement of the load point—the acceleration to instability is self-driven. For small stress steps as K approaches K_c an instability does not occur. Third, similar results are obtained for the time to instability over the entire range of stiffnesses from $K = K_c$ to $K = 0$, with the exception that the time to instability increases slightly as K approaches K_c. This insensitivity of the time to instability indicates that patch radius does not exert a strong influence on the details of the acceleration to instability.

Figure 3 gives the time to instability for simulations in which the loading point attached to the slider was subjected to a uniform velocity, V_l, following the initial load step. The stiffness in this series of simulations was set at $K_c/10$. For comparison the zero-load rate result of Figure 2 is included. The effect of moving the loading point at a uniform velocity is to reduce the time to instability for the smaller initial loading steps. At loading rates $V_l < 10/\theta$ the time to instability is little affected by the motion of the loading point. Results from simulations with lower spring stiffnesses are similar, but the effect of loading point velocity is reduced. With stiffness approaching K_c and small initial stress steps, slip does not accelerate monotonically to instability. Instead, motion of the loading point causes slip to progress through a series of stable oscillations of increasing amplitude before reaching instability. The narrow range of conditions yielding oscillations has not been considered in the current study.

In the range of initial stress and loading conditions which yield an approximately linear relation between initial stress and logarithm of the time to instability, the results are found to scale by:

$$\mu - \mu_s = C + A_1 \ln \theta_0 / t \tag{18}$$

where t is the time to instability following the stress step $(\mu - \mu_s)$, C is a constant that depends on K and A_2 and θ_0 is the value of θ at the time of the stress step. This result is illustrated by the series of simulations of Figure 4 and Figure 5. The simulations of Figure 4 employ different values for A_1 while using otherwise identical constitutive parameters and model conditions. For the computations of Figure 5 only θ has been varied. From earlier discussion recall that θ is approximately the time since the last rapid slip event. The result of (18) also agrees with a recently obtained analytic solution for this problem. It is noted that as A_1 goes to zero the time to instability also goes to zero. In this case μ_s equals $\mu_{\max}$ and becomes the threshold stress for instability. This result is evident from (8) and (9) for $\mu_{\max}$ and μ_s.

Figure 6 plots displacement against time showing the interval that includes the final 50 hr before the onset of instability. The constitutive parameters for these simulations are those of the simulations of Figure 2 and are considered representative of the experimental measurements. The velocity of the loading point, V_l; stiffness; initial stress; and θ are at the indicated values. Note that the form and magnitude of the displacements in the 50 hr before the instability are essentially the same in all cases. The curves are offset because of differing durations of slip before the 50-hour cutoff of the plots. However, the slip rates are initially very low, with the

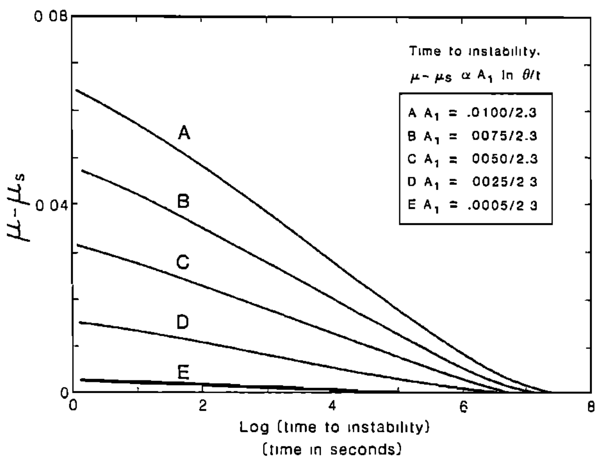

Fig. 4. Time to instability following a step in applied stress to an initial level given by $\mu - \mu_s$, for different values of A_1. In each series of computations the loading point is held fixed, the initial value of θ is 10^7, and $K = 0$, $B_1 = .015/2.3$, $A_2 = 1.0$. The parameter A_1 is .010/2.3, .0075/2.3, .005/2.3, .0025/2.3, and .0005/2.3 for the curves A, B, C, D, and E, respectively. Plotted against $\log_{10} t$ the slopes of the linear portions of the curves are $-.010$, -0.0075, -0.0050, -0.0025, -0.0005 for A, B, C, D, and E, respectively. When plotted against the natural log of t the slopes are equal to $-A_1$.

result that most of the slip occurs in the final stages of the accelerating slip and the total amount of slip is also less sensitive to the duration of slip.

Of possible practical interest is the possibility of recording premonitory displacements for the purpose of earthquake prediction. From that perspective these results indicate that expected displacement rates in the early stages of the accelerating slip rates are very low and would be difficult to detect. Somewhat arbitrarily the interval 10 days to 10 min has been taken as the interval of practical interest for earthquake prediction. For simulations with these constitutive parameters, having an interval of slip of 10 days or more, the amount of stable slip in the interval from 10 days to 10 min is always about $d = 5D_c$. This result is discussed below.

Discussion and Conclusions

We have outlined some aspects of the processes leading to unstable fault slip that may be inferred from laboratory observations of fault constitutive properties. The applicability of the laboratory results to faulting in the earth, of necessity, involves extrapolation to conditions not yet examined in the laboratory. In the absence of observations to the contrary, it is assumed that the structure of rate- and state-dependent constitutive relations as employed for the laboratory results is valid for the interevent times, slip rates and temperature conditions for shallow earthquake-active faults. The wide range of slip phenomena observed in the laboratory and the obvious parallels between phenomena observed on simulated faults and active faulting in nature give some support for this assumption.

On faults with rate- and state-dependent constitutive properties unstable fault slip does not begin spontaneously at some threshold stress on a previously locked fault except if μ jumps instantaneously to μ_{max}. In general, stress will increase continuously or in small increments, and unstable slip will be preceded by an interval of stable accelerating slip. This result leads to the notion of a nucleation phase for unstable slip. The principal condition for unstable slip is that accelerating slip must take place on a patch that exceeds a minimum radius, r_c, given approximately by (14). Slipping portions of a fault with $r > r_c$ which may be slipping very slowly until a short time prior to instability are therefore the locations of interest for the nucleation of unstable fault slip. Once unstable slip begins, the large stress drops associated with the instability may permit the instability to propagate well beyond the dimensions of the nucleation patch.

The creation of nucleation patches on faults in nature is presumably controlled by the processes stressing the fault and by heterogeneity of fault stresses and fault constitutive properties. Stresses at the edge of a patch will increase because of remote tectonic stressing and because of slip on the patch. If slip begins on a patch that is of subcritical size the dimensions of the patch will tend to increase to the critical size with time as the stresses at the edges of the patch increase. This process of stable patch growth to

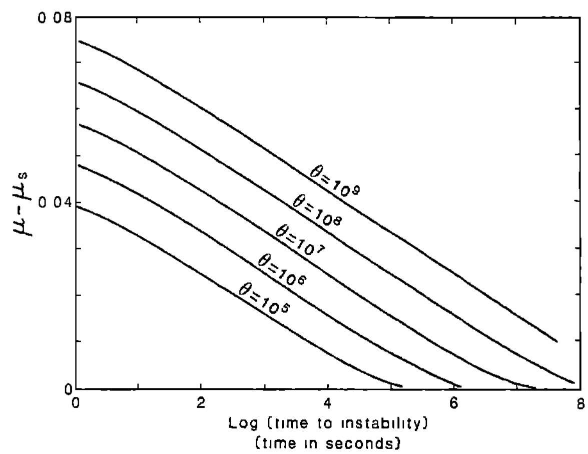

Fig. 5. Time to instability following a step in applied stress to an initial level given by $\mu - \mu_s$, showing the effect of different initial values of r. For these computations the loading point was held fixed, $K = 0$ and the constitutive parameters are those of Figure 2. The initial values of θ are indicated on the plot. Note that the time to instability following a stress step is proportional to θ.

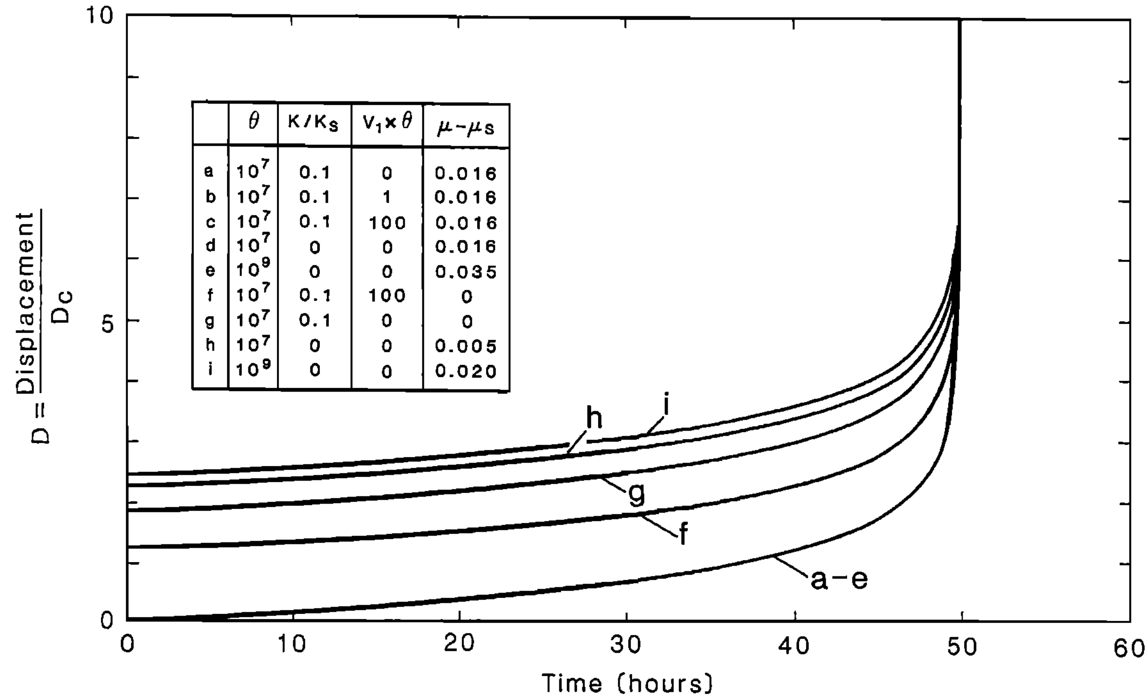

Fig. 6. Displacements plotted against time beginning 50 hr before instability for a variety of conditions and model parameters. The constitutive parameters are those of Figure 2.

create the nucleation patches is therefore associated with intrinsically stable fault slip.

On patches exceeding a critical size, slip will begin to accelerate to instability when the applied stress exceeds the steady-state friction, μ_s, for the patch. The results of Figures 2 through 6 outline some characteristics of the interval of accelerating slip on the nucleation patch. An interesting feature of the results is that the time to instability and the form of the premonitory displacements are relatively insensitive to stiffness (i.e., patch radius) and loading rates over a wide range of conditions. This suggests that the model assumption of constant patch size during accelerating slip and the approximation of (14) which reduces a 3-D problem to one with a single dimension parameter may not severely affect the quantitative results for time to instability and premonitory displacements. At $\mu > \mu_s$, the time to instability is very sensitive to the magnitude of the stress step. Over a wide range of conditions the logarithm of time to instability following a stress step decreases linearly with the amplitude of the stress step.

It appears plausible that the frequency versus time relationships of aftershock sequences and of mainshocks following foreshocks arise from the strong dependence of the time to instability of the magnitude of a loading step. Specifically, we propose that mainshocks or foreshocks cause a stress step on nearby nucleation patches that advance the onset time of subsequent earthquakes originating on those patches. For example, in the case of aftershock sequences the mainshock provides the loading step to the population of nucleation patches of the aftershocks. It follows then, that the time of occurrence of earthquakes in a population is determined by the population distribution of stresses on the nucleation patches following the stress step. We assume the initial stresses, $\mu - \mu_s$ on the population of nucleation patches, affected by the earthquake does not vary significantly from a uniform distribution:

$$N = N_0 - R(\mu - \mu_s), \qquad \mu < \mu_{\max} \tag{19}$$

where N is the cumulative number of patches with initial stress in the interval from μ to μ_s following an earthquake-related step in stress. The magnitude of the parameter R gives a measure of the size of the stress step and the area affected by the stress step. Combining (18) and (19) and differentiating yields the familiar result for aftershock rates:

$$\frac{dN}{dt} = \frac{RA_1}{t} \tag{20}$$

Similar frequency versus time relationships are found for foreshock-mainshock pairs [Jones and Molnar, 1979; Jones, 1984] suggesting similar mechanisms for foreshocks and aftershocks.

It is also noted that the explanation presented here

Table 1. Premonitory Displacement and Minimum Nucleation Patch Radius

D_c	Premonitory Slip 10 Days to 10 Min Before Stability	Minimum Radius of Nucleation Patch ($\sigma = 100$ MPa)	Moment of Premonitory Slip ($\sigma = 100$ MPa, $r = r_c$)
50 mm	250 mm	5.3 km	5.5×10^{17} N-m
5 mm	25 mm	0.53 km	5.5×10^{14} N-m
0.5 mm	2.5 mm	53 m	5.5×10^{11} N-m
50 μm*	0.25 mm	5.3 m	5.5×10^{8} N-m
5 μm	25 μm	0.53 m	5.5×10^{5} N-m

* Upper limit of D_c observed in experiments.

for aftershocks at least qualitatively explains observations of the spatial characteristics of aftershocks. Commonly, aftershocks cluster at the edges of the mainshock rupture and spread to progressively greater distances from the rupture with increasing time. The spatial clustering may be explained as arising from the stress concentration at the ends of the rupture. The high stress steps at the rupture edges would give rise to the greatest number of nucleation patches with stresses in excess of μ_s and the greatest number of patches at the upper end of the stress range. The earliest and the greatest numbers of aftershocks would therefore tend to arise at the rupture edges. At progressively greater distances from the rupture, the stress step will be smaller and consequently the times to aftershock instability will be longer.

The results of Figure 6 show that premonitory displacements are insensitive to patch size and loading conditions. This permits some generalizations independent of assumptions for specific models. At long times before instability the slip rates are low, but slip rapidly accelerates as the time of instability approaches. Consequently, the largest premonitory displacements occur in the final stages of the stable slip interval. Using constitutive parameters representative of the laboratory results, the displacements in the interval from 10 days to 10 min before instability are found to be approximately $5D_c$.

Table 1 gives premonitory displacement, minimum critical patch radius and minimum moment, M_0, of premonitory strain for a range of possible values for D_c. In Table 1, equation (14) has been used with the criteria for instability at steady-state slip $\xi = B_1 - A_1 = 0.0022$, $G = 2.5 \times 10^4$ MPa and a normal stress of 100 MPa. Direct measurements of D_c for prepared laboratory faults fall in the range 1–50 μm, yielding 5–250 μm of premonitory slip if extrapolated to the large θ for earthquake interevent times—a result that is not particularly encouraging for use of premonitory slip as a practical means for earthquake prediction. However, the magnitude of D_c for natural faults is an open question at this time.

The laboratory observations of D_c indicate that this parameter is sensitive to surface roughness and gouge particle dimensions. Because natural faults are much rougher than laboratory faults and also have much larger and more heterogeneous gouge fragments it is plausable to expect that D_c for natural faults may be substantially greater than any of the laboratory measurements. Upper limits for D_c can be approximately established from earthquake source parameters by assuming the earthquake instability does not propagate beyond the nucleation patch. Equation (14) can then be employed to find a limiting maximum value for D_c by taking the critical patch radius to be equal to the earthquake rupture radius.

These considerations suggest two very different limiting hypotheses for the characteristics of D_c of faults in nature. The first hypothesis is for D_c to be the same everywhere. In this case, the smallest earthquakes occurring on a fault are those with a source radius equal to the critical radius of the nucleation patches. Larger earthquakes occur only when the earthquake rupture is able to propagate beyond the nucleation zone. Under these conditions, the magnitude of premonitory displacements of earthquakes of all sizes would be the same. The smallest recorded earthquakes in many regions appear to have a source radius on the order of 100 m or possibly much less. Using $r_c = 100$ m, $A_1 = 0.010/2.3$, $B_1 = 0.015/2.3$, $G = 2.5/10^4$ MPa, and $\sigma = 100$ MPa in (14) yields a maximum D_c of 1 mm. At $\sigma = 10$ MPa, D_c is 0.1 mm. If this hypothesis is correct, it is evident that accelerating premonitory displacements would be difficult to detect in most situations.

Recently Johnston et al. [1986] reported on a number of high quality strain records obtained for the few hours immediately preceding several moderate earthquakes. At the resolution of these records, there is no evidence for premonitory strains for any of the events studied. For the different earthquakes Johnston et al. [1986] have calculated the maximum possible moments for premonitory strain at the source that could have escaped detection by falling below the limit of strain resolution on the records. The maximum possible moments fall in the range 1.5×10^{12} N-m to 3.3×10^{17} N-m. When applied to the results of the present study those maximum moments provide an upper limit to D_c in the nucleation zone. Assuming,

σ = 100 MPa, $G = 2.5 \times 10^4$ MPa, and $\xi = 0.0022$ in (14) the maximum values of D_c are 0.9 mm and 200 mm, respectively for the end-member moments of 1.5×10^{12} N-m and 3.3×10^{17} N-m, respectively.

The second limiting hypothesis calls for scaling of the dimensions of earthquake ruptures by the dimensions of the nucleation zone:

$$r_c = r/F \tag{21}$$

where r is the rupture radius and F is a constant equal to or greater than 1. This hypothesis implies a heterogeneity of nucleation patch dimensions. Consequently, from (14) and (21) D_c would have variability comparable to the size-frequency statistics of earthquakes. This hypothesis might obtain if heterogeneity of fault-strength to fault stress is great enough to inhibit rupture propagated much beyond the nucleation zone. Premonitory displacements would scale by rupture dimensions and large earthquakes could have very sizable premonitory displacements.

Available evidence is very limited, but possibly favors the second hypothesis or some intermediate case. The intermediate case would consist of heterogeneous D_c with the radius of some ruptures scaling by r_c, but with many others resulting from runaway propagation from small nucleation patches. Faults in nature appear extremely inhomogeneous compared to the prepared sawcuts generally used to simulate faults in the laboratory, perhaps suggesting large and spatially variable values for D_c. Estimates of fracture energy for earthquakes are often large and require breakdown displacements (i.e., D_c) on the order of centimeters or tens of centimeters. Because small earthquakes cannot originate on faults with such large D_c values the occurrence of small earthquakes elsewhere on a fault implies great heterogeneity of D_c. Conversely, the small earthquakes might be evidence of problems with fracture energy determinations. The occurrence of large premonitory displacements, as we have seen, also indicate large values for D_c in the nucleation region. The failure to observe evidence for premonitory creep for many moderate earthquakes in well monitored locations may be interpreted as rupture propagation well beyond a small nuleation patch.

Acknowledgments. During the course of this study several discussions with P. Segall were especially useful. P. Segall, P. Okubo, and B. Hobbs are thanked for reviews and comments.

References

Andrews, D. J., Rupture velocity of plane strain shear cracks, *J. Geophys. Res.*, 81, 5679–5687, 1976.

Chinnery, M. A., Theoretical fault models, Symp. Processes in the Focal Region, Publ. Dominion Observatory, Ottawa, 37, 211–223, 1967.

Day, S. M., Three-dimensional simulation of spontaneous rupture: The effect of non-uniform prestress, *Bull. Seismol. Soc. Am.*, 72, 1881–1902, 1982.

Dieterich, J. H., Time-dependent friction in rocks, *J. Geophys. Res.*, 77, 3690–3697, 1972.

Dieterich, J. H., Time-dependent friction and the mechanics of stick-slip, *Pure Appl. Geophys.*, 116, 790–806, 1978a.

Dieterich, J. H., Preseismic fault slip and earthquake prediction, *J. Geophys. Res.*, 83, 3940–3948, 1978b.

Dieterich, J. H., Modeling of rock friction: 1. Experimental results and constitutive equations, *J. Geophys. Res.*, 84, 2161–2168, 1979a.

Dieterich, J. H., Modeling of rock friction: 2. Simulation of preseismic slip, *J. Geophys. Res.*, 84, 2169–2175, 1979b.

Dieterich, J. H., Constitutive properties of faults with simulated gouge, in *Mechanical Behavior of Crustal Rocks*, Geophysical Monograph 24, 103–120, AGU, Washington, D.C., 1981.

Dieterich, J. H., and G. Conrad, Effect of humidity on time- and velocity-dependent friction, *J. Geophys. Res.*, 89, 4196–4202, 1984.

Dieterich, J. H., D. W. Barber, G. Conrad, and Q. W. Gorton, Preseismic slip in a large scale friction experiment, Proc. U.S. Rock Mech. Symp., 19th, 110–117, 1978.

Gu, J.-C., J. R. Rice, A. L. Ruina, and S. T. Tse, Slip motion and stability of a single degree of freedom elastic system with rate and state dependent friction, *J. Mech. Phys. Solids*, 32(3), 167–196, 1984.

Johnston, M. J. S., A. T. Linde, and M. T. Gladwin, Fault failure with moderate earthquakes, *Tectonophysics*, in press, 1986.

Jones, 1984.

Jones, L. M., Foreshocks and time-dependent earthquake hazard assessment in southern California, *Bull. Seismol. Soc. Amer.*, in press, 1985.

Jones, L. M., and P. Molnar, Some characteristics of foreshocks and their possible relation to earthquake prediction and premonitory slip, *J. Geophys. Res.*, 84, 5709–5723, 1979.

Okubo, P. G., and J. H. Dieterich, State variable fault constitutive relations for dynamic slip, this volume, 1986.

Rice, J. R., Constitutive relations for fault slip and earthquake instabilities, *Pure Appl. Geophys.*, 121, 443–475, 1983.

Rice, J. R., and A. L. Ruina, Stability of steady frictional slipping, *J. Appl. Mech.*, 50, 343–349, 1983.

Ruina, A. L., Friction laws and instabilities: a quasi-static analysis of some dry friction behaviour, Ph.D. dissertation, Brown University, Providence, R.I., November 1980.

Ruina, A. L., Slip instability and state variable friction laws, *J. Geophys. Res.*, 88, 10359–10370, 1983.

Scholz, C. H., and J. T. Engelder, The role of asperity indentation and ploughing in rock friction, I. Asperity creep and stick-slip, (abstract), *Int. J. Rock Mech. Min. Sci. Geomech.*, 149–154, 1976.

Weeks, J. D., and T. E. Tullis, Frictional sliding in dolomite: a variation in constitutive behavior, *J. Geophys. Res.*, 90, 7821–7826, 1985.