1. Introduction

This chapter summarizes experimental observations and related theoretical developments of faulted and intact rock properties related to earthquake nucleation, failure and dynamic slip. We will be concerned primarily with earthquakes occurring in the brittle crust. Intermediate and deep-focus earthquakes have unique mechanical considerations that are discussed in Section 7. We focus on repeatable laboratory observations and their direct implications for fault strength and stability. Important results that may be new, incomplete or controversial have also been included. To distinguish the well-established findings from others, the latter appear in “new and provisional results” sections.

Natural faults in the Earth’s crust are zones of localized shear that are inherently weaker than the surrounding intact rock (Chapter 29 by Sibson). Slip can occur slowly as aseismic creep or rapidly as earthquakes with long interevent times during which little or no slip occurs. Crustal faults have complex and irregular geometry, typically fractal in nature, that includes irregular, interlocked surfaces, offset segments, bends, and junctions. Since compressive stresses increase with depth due to the overburden, the natural tendency for irregular fault surfaces to move apart during sliding is suppressed. Instead, fault slip at depth involves grinding and crushing of grains and must involve continual fracture of asperities or interlocked regions. Breaking of asperities may in fact control the position and timing of earthquake nucleation. For this reason, we will examine the brittle fracture of intact rock as well as frictional properties of preexisting faults.

Experimental and theoretical rock mechanics studies have greatly influenced our understanding of earthquakes and related crustal processes. Many key theoretical advances have either originated or been tested and refined through observations of brittle deformation under controlled laboratory conditions. Although the association of earthquakes with crustal faults was established by the beginning of the 20th century (Lyell (1877) actually suggested a connection before this), important empirical relations for strength of rock and granular material were developed much earlier. The Coulomb failure criterion for initially intact rocks was proposed in the late 18th century and continues to be of great practical use

\[ |\tau| = c + \mu_n \sigma_n \]  

where \( \tau \) and \( \sigma_n \) are shear and normal stress resolved on the eventual fracture plane, \( c \) is cohesion, and \( \mu_n \) is coefficient of internal friction. Here we define fracture as a loss of cohesion across a material plane, and failure as a loss of strength, thus fracture is a subset of failure. For the failure of intact material we conform to the standard definition that the coefficient of internal friction is the slope of the failure surface: \( \mu_n = \partial \tau / \partial \sigma_n \) (Jaeger and Cook, 1971), the failure surface or envelope being the solution to the failure criterion, i.e., for Eq. (1) the failure surface is linear and \( \mu_n \) is a constant. Griffith (1920) demonstrated the importance of flaws, which act as stress concentrators, in controlling the strength of brittle materials, materials whose ability to resist deformation decreases with increasing permanent deformation. Using this approach, fracture mechanics (e.g., Broek, 1982; Atkinson, 1987) has been highly successful in relating defect structures to fracture strength and related properties. Similar ideas have been a fundamental part of frictional contact theory where it is recognized that the real area of contact supporting stresses across a fault is a small fraction of the total fault area (Jaeger and Cook, 1971; Dieterich, 1978b; Dieterich and Kilgore, 1994).

Many of the theoretical advances in rock fracture and friction have their origins in metallurgy. Although the macroscopic phenomenology can be similar, the micromechanical deformation mechanisms are often distinct for metals and framework silicates. Brittle fracture in silicates commonly involves opening of cracks and an accompanying volume increase. Because this dilatational strain must do work against the ambient confining stress, Coulomb materials, unlike metals, are characterized by a pressure sensitivity of yield strength as expressed in Eq. (1).

The basic concept of friction or frictional resistance of two surfaces in contact, the ratio of the shear resistance to the
surface normal stress, is referred to as Amontons’ law and was developed about eighty years before the Coulomb criterion (Rabinowicz, 1965; Jaeger, 1969). In this case, shear resistance \( (f_s) \) is related to the normal force \( (f_n) \) acting on the surface by \( |f_s| = \mu f_n \), where \( \mu \) is coefficient of friction. Dividing by the area of the surface gives

\[
|\tau| = \mu f_n \quad (2a)
\]

For many materials, spanning a wide range of mean stress, \( \mu \) is only weakly dependent on normal stress. Friction will depend both on the material filling the interface (fault gouge) and on the surface geometry (surface roughness, coherence or degree of interlocking of surfaces). For example, joint systems, which are formed by tensile opening of fractures with little net shearing, can result in clean, well-mated surfaces that have higher shear strength than surfaces with uncorrelated roughness. For coherent surfaces, shear strength may not vanish at zero normal stress and a more accurate empirical description of shear strength becomes

\[
|\tau| = S_0 + \mu f_n \quad (2b)
\]

where \( S_0 \) is an ‘inherent’ shear strength similar to cohesion in Eq. (1) (Jaeger and Cook, 1971). The quantity most easily measured and most often used in studies of rock friction is the simple ratio of shear to normal stress. We will refer to this ratio as friction or frictional resistance, denoted by \( \mu \), to distinguish it from the coefficient of friction \( \mu_f \), so that Eq. (2b) gives

\[
\mu = \frac{|\tau|}{f_n} = \frac{S_0}{f_n} + \mu_f \quad (3)
\]

High temperature, high pressure or the addition of reactive pore fluids may allow for activation of ductile deformation mechanisms (mechanisms whereby permanent deformation can be accommodated without loss of strength) such as dislocation glide or pressure solution, and can allow deformation that lacks the pressure sensitivity characteristic of friction and fracture. Thus, both \( \mu \) and \( \mu_f \) are expected to decrease with increasing depth in the crust (Mogi, 1966), and in fact laboratory measurements indicate that \( \mu_c \), \( \mu_f \) and \( \mu \) decrease with increasing mean stress \( (\sigma_1 + \sigma_3)/2 \) or normal stress (Fig. 1b). Normal stress increases with depth and provides a convenient independent variable for identifying trends in shear strength and other fault-related properties. However, when considering failure of intact rock, there is no preexisting fault plane and normal stress becomes a less useful concept. In this case, mean stress provides a good alternative and will be used throughout this chapter.

Although it is tempting, based on observations such as shown in Figure 1a, to treat friction as a material property, this assumption may not be correct. Some effects resulting from surface geometry can be accounted for by adjusting \( S_0 \),

**FIGURE 1** (a) Shear-normal stress plot for Westerly granite showing peak intact strength and frictional strength. (b) Alternative measures of strength for intact and faulted Westerly granite.
However, other geometric properties related to the deformation of granular material between rigid fault blocks also affect shear strength. In an idealized sense, a fault zone can be described mechanically as a thin, weak layer sheared between rigid blocks. The boundary conditions for this geometry require that normal strains in the gouge layer in the plane of the fault must be zero and average flow lines must eventually become parallel to the fault walls, imposing constraints on the stress field that develops within the gouge layer (Spencer, 1982; de Josselin de Jong, 1988; Rice, 1992; Savage et al., 1996; Savage and Lockner, 1997). The strains resulting from average simple shear deformation imposed on a gouge layer by the fault walls are inconsistent with strains resulting from flow of a Coulomb material (Byerlee and Savage, 1992). As a result, complex structures such as Riedel shear bands develop in the fault gouge (Gu and Wong, 1994; Savage et al., 1996) which reflect a heterogeneous stress field that develops within fault zones in response to deformation (Mandl et al., 1977; Morgan and Boettcher, 1999).

Laboratory measurements show that numerous parameters in addition to normal stress or confining pressure affect strength of both fault surfaces and intact rock. These include mineralogy, porosity, cementation, packing geometry of gouge, surface roughness, coherence of surfaces, angularity and size of gouge particles, temperature, pore fluid pressure and composition, deformation rate, deformation and stress history, fabric, foliation, or other anisotropic properties. Sample size is reported to affect intact strength, but appears to have little or no effect on frictional strength. Some of these effects are discussed individually in subsequent sections. Pore fluid pressure, in particular, has a first order effect on rock and fault strength in cases where there is well-connected porosity. This effect is discussed in more detail in Sections 2.3.2 and 5. For now, we note that “effective” normal stress should be substituted into Eqs. (1), (2), and (3) where an effective normal stress component is the normal stress reduced by pore pressure: i.e., $\sigma_{n,\text{eff}} = \sigma_n - p$ (Terzaghi, 1929; Hubbert and Rubey, 1959).

### 2. Brittle Fracture

For theoretical development or numerical modeling, it is common to assume that earthquakes occur in preexisting fault zones. For a number of reasons, however, we will include analysis of the rupture process in both intact and prefRACTured rock (cf. Section 4.2). Although the Earth’s crust is generally permeated with preexisting fractures and joints, it is not known what fraction of earthquakes occur on preexisting faults, what percentage of any single earthquake rupture surface represents new faulting (especially for earthquakes that jump fault segments), or whether rupture nucleation might occur at jogs or locked segments and involve breaking of rock which is at or near the intact rock strength. Also, there are many similarities between the mathematical analysis of intact rock deformation and deformation on faults. The micromechanical mechanisms for fracture and sliding on faults both involve grain breakage, grain rotation, crack propagation, and in many cases plastic deformation mechanisms. Fault zone materials will be altered by passage of pore fluids, but this process also occurs in intact rock. Finally, in the long periods between earthquakes, faults can be expected to heal and restrengthen by an amount which, in the limit, would return to intact rock strength. For these reasons, a proper review of rock friction should also include discussion of rock fracture.

A number of comprehensive studies and reviews of rock failure, faulting and rheology currently exist (Handin, 1966; Jaeger and Cook, 1971; Bock, 1978; Lama and Vutukuri, 1978; Paterson, 1978; Lambe and Whitman, 1979; Kirby and McCormick, 1984; Kirby and Kronenberg, 1987; Mandl, 1988; Scholz, 1990; Evans and Dresen, 1991; Lockner, 1995). Readers are encouraged to refer to these studies for additional information. Many of the effects discussed in this section have direct bearing on how rocks deform along preexisting faults and provide a useful framework for much of the discussion of earthquake processes.

#### 2.1 Basic Relations

The relations between fault orientation angle $\beta$, fault normal angle $\psi = (\pi/2) - \beta$, and stress components, including maximum compressive ($\sigma_1$) and minimum compressive ($\sigma_3$) principal stresses (Fig. 2), follow directly from expressing the state of stress on planes which contain the intermediate principal stress axis as a function of the extreme principal stresses

\[
\tau = \frac{1}{2}(\sigma_1 - \sigma_3) \sin 2\beta \tag{4a}
\]

\[
\sigma_n = \frac{1}{2}[(\sigma_1 + \sigma_3) - (\sigma_1 - \sigma_3) \cos 2\beta] \tag{4b}
\]

Here and throughout we take compressive stresses and contractive strains as positive. As indicated by Eq. (4), the orientation of the failure plane is assumed to contain and be independent of the intermediate principal stress.

Laboratory failure tests conducted at different conditions can be used to construct the failure envelope, defined as the locus of points in shear versus normal stress space which define failure (the Mohr plane) (Fig. 2). The simplest laboratory test geometries are uniaxial compression ($\sigma_1 = \sigma_c$, $\sigma_2 = \sigma_3 = 0$) leading to failure stress state “U” in Fig. 2, and uniaxial tension ($\sigma_1 = \sigma_2 = 0$, $\sigma_3 = \sigma_c$) shown as stress state “T”. Tensile failure occurs on a fracture oriented perpendicular to the maximum tensile stress direction ($\beta = 90^\circ$), whereas compressive failure involves shearing on a fault inclined at smaller $\beta$ (typically in the range $20^\circ < \beta < 45^\circ$). More complicated testing conditions are appropriate for deformation tests relevant to earthquake failure, which require higher normal stress and independent variation of one or more
stress components. The two most common test geometries are shear box or direct shear and triaxial. Direct shear machines allow for independent variation of shear and normal stress on the sliding surface. Triaxial machines are the most common method of testing samples at confining pressures as high as 1 GPa, controlled pore fluid pressure, and temperatures relevant to the crust and upper mantle. A restriction on standard triaxial tests is that a fluid is used to apply confining pressure $P_{\text{conf}}$ so that $\sigma_2 = \sigma_3 = P_{\text{conf}}$. Rotary shear machines, where confining pressure, normal stress, shear stress (torque), and pore pressure can be varied independently, are also particularly useful in friction studies since significantly larger displacements can be achieved in this geometry than in any linear displacement apparatus. However, the direct shear and unconfined rotary shear geometries lead to difficulties in controlling pore pressure and temperature.

The steep slope for the failure envelope in Figure 2a shows the strong pressure sensitivity of failure strength for dilatant materials. A Mohr representation in which shear stress is plotted versus normal stress, is shown in Figure 2b (Jaeger and Cook, 1971). For application to the Earth, it is often assumed and observed that failure occurs when Eq. (1) is first satisfied on the most favorably oriented plane. This assumption leads to additional constraints on the stress state and fault orientation,

$$\phi_i = \tan^{-1} \mu_i = \frac{\pi}{2} - 2\beta$$  \hspace{1cm} (5a)

$$\tan 2\beta = \frac{1}{\mu_i}$$  \hspace{1cm} (5b)

$$\sigma_1 = \sigma_c + \left[ \left( \mu_i^2 + 1 \right)^{1/2} + \mu_i \right] \sigma_3$$  \hspace{1cm} (5c)

Equations (5) are Mohr–Navier–Coulomb relations.

### 2.2 Conceptual Fracture Model

As mean stress increases with depth in the shallow to mid crust due to the increase in overburden, both frictional and intact rock strength can be expected to increase. If special conditions result in near-lithostatic fluid pressure, that is, pore pressure equivalent to overburden (i.e., magma intrusion or trapped pore fluid), then the minimum effective principal stress can be low or negative and the rock can fracture in tension. In the case of earthquakes, however, focal mechanisms indicate shear rather than tensile failure. Rocks are generally about ten times stronger in compression than in tension (Lockner, 1995), reflecting the difficulty of propagating a large-scale shear
fracture (generally involving dilatancy) through a compressively loaded brittle material. It is well known that the failure process in rock involves microcrack growth. Open pores, contrasts in elastic properties of adjacent minerals, and weak grain boundaries can all act as stress concentrators in compression (Peng and Johnson, 1972; Tapponnier and Brace, 1976) and shearing along surfaces oblique to the maximum compressive principal stress $\sigma_1$ is likely to play an important role in the development of large local stresses (Peng and Johnson, 1972; Horii and Nemat-Nasser, 1986; Sammis and Ashby, 1986; Ashby and Sammis, 1990; Lockner et al., 1992b). The local stresses induced near the crack tips contain a component of tension which leads to local tensile failure before the fracture toughness for failure in shear is achieved (Lawn and Wilshaw, 1975). As a result, tensile “wing” cracks grow aligned subparallel to the $\sigma_1$ direction. Unlike the case of remotely applied tensile stress, two important stabilizing processes take place during the loading of cracks in compression. First, as wing cracks extend, stress intensity decreases, and additional deviatoric stress must be applied to cause further crack growth (Costin, 1983, 1987; Horii and Nemat-Nasser, 1986). Second, diagonal flaws, which are favorably oriented to act as nucleation points for shear failure, propagate out-of-plane (parallel to $\sigma_1$) and cannot maintain the necessary shear geometry (Lawn and Wilshaw, 1975). Eventually, crack densities become sufficiently large for significant crack interaction to occur, leading to the nucleation of a proto-fault: a shear fracture connecting en-echelon arrays of the dilatant microcracks (Peng and Johnson, 1972; Horii and Nemat-Nasser, 1986; Sammis and Ashby, 1986; Costin, 1987; Kemeny and Cook, 1987; Ashby and Sammis, 1990; Du and Aydin, 1991; Lockner et al., 1992b; Reches and Lockner, 1994). Finally, quasi-static fault growth experiments (Lockner et al., 1991, 1992a) have demonstrated that following nucleation, the shear fault propagates in-plane due to its own stress field (Fig. 3). Although the fracture propagates as mixed shear modes II and III, the fracture tip is surrounded by a process zone, much like the earthquake rupture zone described by Rice (1980). In this high-stress region, energy is dissipated by the growth of a halo of microcracks (primarily tensile (Moore and Lockner, 1995)) as the fracture advances through the rock. Similar results have been observed for damage zones surrounding natural faults (Vermilye and Scholz, 1998). The acoustic emission experiments provide an important constraint on nucleation. In past studies it was often argued that the eventual fault plane began forming, as a region of high microcrack damage, well before peak strength. Although this interpretation may hold for samples containing large preexisting flaws, it appears that in homogeneous, isotropic rocks, the location of the eventual fault plane is not correlated in any obvious way to precursory, localized damage. Similar results have been reported in field studies (Linde and Johnston, 1989; Johnston and Linde, 2002) where nucleation zones for moderate-sized earthquakes were inferred to be less than 0.1% of the coseismic rupture surface area.

![Acoustic emission (AE) hypocentral locations during fault formation of initially intact Westerly granite. Time progresses from left to right. Middle: Figures view sample along strike of eventual fault plane which appears as diagonal feature in (e) and (f). Top: Same AE events viewed perpendicular to strike. Bottom: Accompanying stress-displacement curve indicates segments of the experiment from which acoustic emission plots are made. Fault nucleation occurs in (d).](https://example.com/ae.plot)

### 2.3 Factors Affecting Fracture Strength of Rock

#### 2.3.1 Confining Pressure Effect

Confining pressure affects brittle fracture strength by suppressing the growth of dilatant microcracks. Microcracks tend to grow parallel to $\sigma_1$ when a sample is loaded in compression by locally overcoming the ambient compressive stress field near crack tips and developing a localized region of tensile stress. This process requires some specialized mechanism such as slip along grain boundaries or bending of a rigid grain that is adjacent to an open pore or more compliant grain (Peng and Johnson, 1972; Tapponnier and Brace, 1976; Sammis and Ashby, 1986; Reches and Lockner, 1994). Such mechanisms are generally enhanced by deviatoric stress and suppressed by mean stress. In addition, slip along grain boundaries will
not occur until frictional contact strength, which increases with mean stress, is overcome. The overall strengthening effect can be represented in a shear stress–normal stress plot (Mohr–Coulomb failure envelope) such as Figure 1a for Westerly granite. In general, the failure envelope is concave towards the normal stress axis so that a linear failure criterion such as Eq. (1) may not be a satisfactory approximation. This curvature becomes most important when experimental data span a broad range of normal stress. Some minerals, such as calcite, undergo a brittle–ductile transition within this pressure range and exhibit severe curvature of the failure envelope. Furthermore, the difference between intact failure strength and frictional sliding strength vanishes at high pressure (Byerlee, 1967; Ismail and Murrell, 1990). A good empirical fit to the intact rock strength in Figure 1 (Lockner, 1998) is provided by

\[
\sigma_1 = \sigma_3 + (46.660 + 5128.7\sigma_3)^{1/2} - 8.3 \text{ MPa} \quad (6)
\]

where stresses are in MPa.

The majority of strength measurements have been conducted under uniaxial or “triaxial” stress conditions in which \(\sigma_2 = \sigma_3\). A limited number of true triaxial measurements have been performed to explore the effect of intermediate principal stress on failure mode. While the most commonly used failure criteria (e.g., Mohr–Coulomb) assume that failure is independent of intermediate stress, experimental evidence demonstrates that this assumption is not strictly true (Mogi, 1967, 1970; Amadei et al., 1987).

### 2.3.2 Pore Fluid Pressure and Effective Pressure Law

Pore fluids can affect fracture strength through a direct pressure effect as well as through chemical interactions with the rock matrix. Mechanically, pore pressure acts to reduce the normal stress tensor components throughout the rock mass according to the effective pressure law (Hubbert and Rubey, 1959)

\[
\sigma_{n,\text{eff}} = \sigma_n - \alpha p
\]

where \(\alpha\) is a constant related to pore geometry. Although variations in \(\alpha\) have been measured for the effect of pore pressure on transport properties (Walsh, 1981), \(\alpha = 1\) is appropriate for fracture strength of rock with well-connected porosity. If fluid pressures in the crust were low, the increase in mean stress and temperature with depth, due to overburden pressure, should result in large shear stress and the gradual change from brittle to ductile deformation in the mid to lower crust (as suggested by the convergence of frictional and intact strength in Fig. 1). The normal hydraulic head for well-connected porous material will offset the overburden pressure by approximately one-third. However, for isolated porosity at depth, normal consolidation and dewatering processes can compress pore fluids to well above hydrostatic pressures and under proper circumstances may lead to repeated episodes in which the rock is hydraulically fractured. There is currently considerable interest in fluid overpressure phenomena and their relation to oil reservoir partitions (Hunt, 1990; Powley, 1990) as well as to larger-scale crustal processes such as earthquake cycles and the reduction of strength of mature faults (Byerlee, 1990, 1992, 1993; Nur and Walder, 1990; Rice, 1992; Miller et al., 1996; Sibson, 2002) (see Section 3.2).

Since water reacts with silicates to disrupt the Si–O bond, pore fluids can have a profound chemical effect on the physical properties of crustal silicate rocks. CO\(_2\) concentration plays a similar role in the diagenesis and metamorphism of carbonates. Crack healing, cementation, and densification all act to strengthen rock with time (Hickman and Evans, 1992). By contrast, chemically assisted crack growth is the primary mechanism for stress corrosion and static fatigue (subcritical crack growth) at temperatures in the 0–200°C range (Rutter and Mainprice, 1978; Kranz, 1980; Kranz et al., 1982; Lockner, 1993b). Subcritical crack growth is crack extension at crack tip stress below the critical stress intensity \(K_{IC}\) needed to split lattice bonds (\(K_{IC} < K_{IC}\) (Atkinson and Meredith, 1987; Costin, 1987). Subcritical crack growth rate \(v\) is generally related to \(K_{IC}\) through a power law

\[
v = A_1 K_{IC}^n
\]

or exponential form

\[
v = A_2 \exp(b_2 K_{IC})
\]

where \(A, b, \text{ and } n\) are empirical constants. An alternative view relates crack growth rate to energy release rate \(G_{II} (\sim K_{IC}^2)\) (Lockner, 1993b):

\[
v = A_3 \exp(b_3 G_{II})
\]

Experimental data are insufficient to identify which form fits best. However, there are theoretical arguments, based on reaction rate theory, to prefer Eq. (9) or (10) (Costin, 1983; Lockner, 1993b). Nominal values for \(n\) at room temperature and in wet rock (Atkinson, 1987, Table 11.6) are: 15–40 for quartz and quartz rocks; 10–30 for calcite rocks; 30–70 for granitic rocks; and 25–50 for gabbro and basalt. A typical value for \(b_2\) for room temperature granite is 0.04 (MPa m\(^{-1/2}\))\(^{-1}\). The corrosive effect of the pore fluid can be altered significantly by varying the activity of dissolved species as well as altering the electrochemical potential (Dunning et al., 1980; Ishido and Mizutani, 1980; Atkinson and Meredith, 1987; Evans and Dresen, 1991; Hickman and Evans, 1991).

### 2.3.3 Strain Rate

In the brittle deformation field, rocks typically exhibit a pseudoviscous effect prior to failure which is reflected in a strength increase with increasing strain rate. This effect is best known in the mining industry as static fatigue in which a pillar or other load-bearing structure will fail after some period of time under
constant load. Subcritical tensile crack growth has been identified as the principal mechanism responsible for static fatigue in rock (Kranz, 1980; Kranz et al., 1982; Lockner, 1993b) and has been studied in recent years (Atkinson and Meredith, 1987). The strain-rate dependence of rock strength has been measured in the laboratory at rates above $10^{-10}$ sec$^{-1}$ which is generally much faster than geologic rates even in tectonically active regions. However, the lower limit does represent a useful limit for many engineering applications. Experimental results covering a broad range of strain rates are shown in Figure 4. Average strain rates have been obtained from static fatigue tests on Barre granite (Kranz, 1980) by dividing time-to-failure data by total inelastic strain. Various studies (Costin, 1983; Lockner, 1993b) have been successful in relating this macroscopic relation between stress and strain rate to subcritical microcrack growth rate and its sensitivity to stress intensity at crack tips. They determined an apparent activation volume of 0.43 ± 0.04 kJ MPa$^{-1}$ mol$^{-1}$ (4.3 x 10$^{-4}$ m$^3$ mol$^{-1}$) for crack growth in granite creep experiments. This approach provides a link between energetics of microcrack extension and bulk pseudoviscous response of rocks in creep (Lockner, 1998). It may also prove useful as a way to incorporate time-dependent effects in damage mechanics applications (Costin, 1987). This rate sensitivity is often expressed as a power law dependence of the form

$$\sigma_{\text{max}} = a \varepsilon^n$$ \hspace{1cm} (11)

Typical rate sensitivities in this regime are 0.02 $\leq n \leq$ 0.04.

By using shock and other impulsive loading techniques, strain rates in excess of 10$^9$ sec$^{-1}$ have been achieved for samples failing in tension (Grady and Lipkin, 1980). Although some question remains as to the dependence of the measurement on machine effects and boundary conditions, numerous experiments show a transitional strain rate of approximately 10$^2$ sec$^{-1}$ above which significantly larger rate dependence is observed. High-rate tensile experiments, taken from Grady and Lipkin (1980), are summarized in the lower-right quadrant of Figure 4 as trend lines without accompanying data values. An upper limit of $n \leq 1/3$ is indicated by both theory and experiment (Brace and Jones, 1971; Green et al., 1972; Lipkin and Jones, 1979; Grady and Lipkin, 1980; Olsson, 1991) for rate sensitivity under these conditions.

### 2.3.4 Temperature

Mechanisms of crystal plasticity involving dislocation motion, diffusion of point defects, and twinning are thermally activated and will dominate brittle cracking as temperature increases (Kirby and Kronenberg, 1987; Evans and Dresen, 1991). Some minerals, such as halite and calcite, will deform ductily at room temperature if sufficient confining pressure is present to suppress brittle crack growth (Friedrich et al., 1989). However, dry quartz appears to deform brittly at room temperature even at confining pressures in excess of 1 GPa (Tullis and Yund, 1977). As previously mentioned, water has a significant effect on pressure-assisted grain-boundary deformation mechanisms such as pressure solution (Blanpied et al., 1991; Hickman and Evans, 1991). These fluid-assisted mechanisms will often dominate at intermediate temperatures and, over geologic time scales, may play an important role in determining rock strength at room temperature (Rutter and Mainprice, 1978). Thus, even in the brittle regime, increasing temperature tends to reduce fracture strength (Bragg and Andersland, 1981; Sayles and Carbee, 1981; Wong, 1982a). The effect of temperature on fracture strength of Westerly granite is shown in Figure 5.

### 2.3.5 Sample Size and Scaling

Laboratory studies in which pressure, temperature and other environmental conditions are controlled have provided strength data relevant to natural faulting and seismicity. However, the scaling of these data to field applications remains problematic (Pinto da Cunha, 1990). For example, the largest laboratory fault models (approximately 1-m$^2$ surface) are many orders of magnitude less than the fault area involved in a moderate earthquake. A general result of fracture mechanics analysis is that stress intensity at the tip of a flaw scales as the square root of the flaw size. Consequently, larger samples, which can contain larger flaws, are more likely to be weak. This argument assumes, of course, that small samples with visible flaws (and certainly flaws that completely bisect the sample) will be rejected and not included in the strength statistics. However, the degree of weakening has not been well determined and should

---

**FIGURE 4** Effect of strain rate on brittle fracture strength. Trend lines in lower-right quadrant are from dynamic tensile fracture experiments (Goldsmith et al., 1976.)
greatest size effect. Pratt et al. (1972) noted a decrease in size sensitivity for samples greater than 1-m length for both diorite and coal and suggested that this may represent an upper limit to the weakening region. Due to the small amount of existing data, it is not yet clear to what extent these tendencies can be generalized. If flaw size is the correct explanation for the weakening effect, then sample strength will depend as much on crack-size distribution as on mineralogy. Furthermore, the studies cited here are all unconfined. As discussed in Section 3, we may expect large rock masses, especially in tectonically active regions, to support stresses that are comparable to their frictional strength rather than laboratory-measured intact strength (Townend and Zoback, 2000).

3. Friction and Crustal Faults

Numerous theoretical arguments have been proposed to explain frictional properties of rocks; many of which have met with reasonable success in matching experimental observations. Yet, natural faults are complex systems, and many complicated and competing processes probably occur during deformation of fault zones. At present, fault rheologic models should be considered empirical. As more sophisticated and precise experimental observations are made, the corresponding mathematical descriptions have invariably become more complicated. There are many earthquake-related applications that can take advantage of linearized approximations of frictional properties where the non-linear aspects of time-dependent failure or fluid–rock interactions are of less importance.

3.1 Byerlee’s Law and Incipient Failure

It has already been noted that Eq. (3) provides a lowest order approximation of rock friction, especially in the lower temperature brittle deformation field. Byerlee (1978) compiled laboratory friction data for rocks and minerals and found that many silicates and carbonates had very similar frictional strengths. He suggested that the characteristic concave-downward trend of shear strength versus normal stress could be adequately approximated by a piecewise linear function often referred to as Byerlee’s law:

\[
\tau = 0.85 \sigma_n \quad \text{for } \sigma_n < 200 \text{ MPa}
\]

\[
\tau = 50 \text{ MPa} + 0.6 \sigma_n \quad 200 < \sigma_n < 1700 \text{ MPa}
\]  

As discussed in the next section, Eq. (12) has proved to be a popular approximation for shallow to mid crustal strength where details of mineralogy or stress state may not be known. It should be noted that many hydrated minerals and especially sheet silicates with good basal cleavage or which accommodate interlayer water have frictional strength significantly lower than that described by Eq. (12) (see Section 3.2). Frictional
resistance $\mu$ as determined by Eq. (12) is plotted in Figures 1a and b. This function, or values of frictional shear strength given by Eq. (12) will be used as a convenient reference in this paper.

Let us assume for the moment that the strength of the brittle crust can be adequately represented by granodiorite. Although this is an obvious simplification of the range of lithologies that may be found in the continental crust, it should provide a reasonable approximation of both intact and frictional strength for many low-porosity, unaltered crystalline rocks.

The intact strength curve shown in Figure 1a, which is derived from granitic samples chosen to be free of macroscopic flaws, provides a reasonable approximation to the upper strength limit of the crust for rock volumes free of faults or joints. For fractured rock masses, containing faults or joints of all orientations, the most favorably oriented faults will fail when the stress state intersects the frictional strength curve (Eq. 12). In this case, the frictional strength would provide a practical upper limit to crustal strength in the brittle field. The limit of differential stress $\sigma_d = \sigma_1 - \sigma_3$ in the crust predicted by Eqs. (2b) and (5b) can be written

$$\sigma_d = \frac{2(S_0 + \mu_f \sigma_{n,eff})}{\cos(\tan^{-1} \mu_f)} = \frac{2(S_0 + \mu_f \sigma_{n,eff})}{\cos \phi_f} \tag{13}$$

where the friction angle $\phi_f = \tan^{-1} \mu_f$ and $\tan 2\beta = 1/\mu_f$. Equation (13) specifies the differential failure stress for frictional failure, and by substituting $c$ for $S_0$ and $\mu_i$ for $\mu_f$, Eq. (13) can predict differential failure stress for rock fracture. Although intact rock strength is an appropriate stress limit at the surface of the Earth and in some near-surface applications such as mining, in the case of earthquake faulting, Eq. (13) appears to serve as an upper limit for stress in the upper crust, especially in intraplate regions. This frictional strength limit is confirmed by in situ measurements made at depths up to 9 km (Hickman, 1991; Townend and Zoback, 2000) (also see Chapter 40 by McGarr et al. and Chapter 34 by Zoback and Zoback). These field observations indicate that $\sigma_d$ generally does not exceed laboratory measurements of the differential strength of faulted, dry, anhydrous silicates (i.e., Byerlee’s law).

Using the same assumptions used to derive Eq. (13), the variation of stress with depth in the brittle part of the crust can be estimated from Eqs. (2b) or (12) provided the magnitude of one of the principal stresses is known. The magnitude of one of the principal stresses can be deduced in normal and thrust faulting environments by using Anderson’s assumption that near the Earth’s surface one of the principal stresses is vertical due to the lack of shear stress on Earth’s free surface (Anderson, 1951), and the vertical stress is given with depth by the overburden. This assumption leads to expressions for normal stress on critically oriented faults in reverse fault environments where $\sigma_v = \sigma_3$;

$$\sigma_n = \frac{(S_0 - \mu_f p)(\mu_f - \frac{1}{\cos \phi_f})}{1 + \mu_f^2 - \frac{\mu_f}{\cos \phi_f}} \tag{14a}$$

and for principally oriented normal faults where $\sigma_v = \sigma_1$;

$$\sigma_n = \frac{(S_0 - \mu_f p)(\mu_f + \frac{1}{\cos \phi_f})}{1 + \mu_f^2 + \frac{\mu_f}{\cos \phi_f}} \tag{14b}$$

(Fig. 7) (as follows from Sibson (1991)). (Here we have used the definition $\phi_f = \tan^{-1} \mu_f$.) Equations (14a) and (14b) as depicted in Figure 7 indicate the Anderson–Byerlee predictions for depth varying stress in the shallow crust.

### 3.2 Strength of Crustal Faults

Laboratory measurements of frictional strength indicate that many common silicates should have sliding strengths proportional to effective normal stress as represented by Byerlee’s law ($0.6 \leq \mu \leq 0.85$). In fact, field observations of crustal stress,

![FIGURE 7 Anderson–Byerlee differential stress with depth in reverse (Rev) and normal (Nor) faulting environments, with hydrostatic (Hyd) fluid pressure of 10 MPa km$^{-1}$ and no fluid pressure (Dry).](image-url)
obtained primarily from shallow borehole stress measurements (see, for example, Zoback et al., 1987; Townend and Zoback, 2000, and other chapters in this volume), show many crustal faults in a stress state consistent with a slightly wider range of friction of $\mu \sim 0.5$–1.0. The occurrence of faults with strengths somewhat below Byerlee’s law can be explained by the presence of weak mineral phases (especially at shallower depths). Fault zones are generally wet and contain ground-up and altered gouge with significant hydrated mineral phases. In particular, many clays, micas and other hydrous phases which are common products of chemical reactions between water and anhydrous silicates have lower frictional strength than Byerlee’s law (Raleigh and Paterson, 1965; Byerlee, 1978; Wang et al., 1980; Moore et al., 1983, 1986, 1996, 1997; Kronenberg et al., 1990; Morrow et al., 1992, 2000) (Fig. 8). Conversely, faults that are stronger than Byerlee’s law may be either partially healed by precipitation reactions, or due to complex fault geometry, require breaking of intact rock to accommodate slip.

Although the strength of many faults seems to conform to Eq. (12), a long-standing debate concerns the strength of major plate-bounding faults and the San Andreas fault (SAF) in particular (see, for example, Hickman, 1991; Scholz, 2000). In what has come to be known as the heat flow paradox, field measurements have failed to detect any significant heat flow anomaly associated with the SAF (Brune et al., 1969; Lachenbruch and Sass, 1980). Whether an active fault dissipates energy either through creep or dynamically during earthquakes, only a small portion of the consumed energy is expended in creating new fractures and surface energy (Wong, 1982b). Similarly, at least for strong faults, seismic energy radiated by earthquakes is also a relatively small fraction of the total energy expended (Lockner and Okubo, 1983; McGarr, 1998). The majority of energy associated with crustal fault deformation should be dissipated as shear heating and, for a fault such as the San Andreas that has been active for millions of years, should produce a measurable heat flow anomaly at the Earth’s surface (Brune et al., 1969; Lachenbruch and Sass, 1980). The lack of such an anomaly has been interpreted to imply an average shear strength of the top 14 km of the SAF of less than 20 MPa (or $\mu < 0.2$). Recent stress orientation measurements made at depths of 1–4 km adjacent to the San Andreas (Mount and Suppe, 1987; Zoback et al., 1987; Shamir and Zoback, 1989) indicate that the greatest principal stress is at an angle approaching 90° to the fault which requires the fault to be weak relative to surrounding subsidiary faults that are oriented more favorably relative to the ambient stress field and yet remain locked while the San Andreas accommodates slip. Stress orientations determined along the SAF imply either an average coefficient of friction less than 0.1 or superhydrostatic fluid pressure (Lachenbruch and McGarr, 1990). Recently Scholz (2000) has proposed a model for a strong SAF in which plate-driving forces are applied as horizontal tractions at the base of the crust. This model also predicts rotation of principal horizontal stresses near the SAF and suggests that the inference that large plate-bounding faults such as the San Andreas are anomalously weak, remains open to debate.

If the SAF is in fact weak relative to its surroundings, then a number of possible explanations have been proposed for the apparent discrepancy between field and laboratory strength observations: (1) dynamic rupture effects reduce normal stress across the fault during large earthquakes so that dynamic sliding strength is low; (2) high fluid pressure reduces effective normal stress and therefore shear strength; and (3) intrinsically weak minerals are responsible for low shear strength on mature faults. In terms of field observations of low heat flow, it may be possible that ground-water circulation acts to broaden and suppress a heat flow signature resulting from heating at depth. However, the borehole stress orientations are not subject to this effect and must be explained in some other manner (i.e., Scholz, 2000).

The possibility of low effective normal stress acting across the fault during large earthquakes has significant appeal and continues to be a topic of active research (see Section 6.2). Early models suggested that intense frictional heating would quickly vaporize pore water trapped in the fault zone and lead to high fluid pressure and low effective normal stress (Lachenbruch, 1980). More recent analysis has focused on the

![Figure 8](image-url)  
**FIGURE 8** Inferred shear strength versus depth on SAF for possible fault zone constituent minerals. Heat flow and stress orientation constraints are indicated by linear strength increase with depth.
possibility that large earthquakes propagate in a pulse-like mode in which dynamic normal stresses unclamp the fault and allow slip at low shear stress (Heaton, 1990; Brune et al., 1993; Ben-Zion and Andrews, 1998). Although these explanations are appealing, they only apply to dynamic slip and do not explain low heat flow along the creeping portion of the SAF in central California which appears to accommodate slip without large earthquakes.

Elevated fluid pressure is known to reduce fault strength (Hubbert and Rubey, 1959) and has been demonstrated to induce seismicity through the reduction of effective normal stress during fluid injection (Raleigh et al., 1976) and reservoir impoundment (Talwani, 1976) (also see Chapter 37 of this volume). Furthermore, fluid overpressured zones, trapped fluids, and pore fluid compartmentalization are commonly observed in boreholes drilled for petroleum exploration and development (Hunt, 1990; Powley, 1990). If fluid migration is localized along faults and mineral precipitation and veining occurs, permeability will decrease and could lead to sealing and isolated fluid compartments. With time, faults can become barriers to flow and time-dependent compaction or dehydration reactions can lead to pore pressure buildup. The steady increase of fluid pressure can eventually lead to hydrofracture of the compartment seal and the escape of fluid. This cycle of slow fluid pressure buildup and then sudden pressure release has in fact been incorporated into a variety of episodic fault models (Sibson, 1981, 1992; Sleep and Blanpied, 1994; Fenoglio et al., 1995; Lockner and Byerlee, 1995; Miller et al., 1996).

The crust surrounding the SAF appears to be in compression (relative motion of Pacific and North American plates suggests a small component of convergence along the SAF) and is stressed critically for thrust failure in some areas such as the central Coast Ranges. Under these conditions, in order to avoid hydrofracture, the pore pressure must be less than or equal to a critical value

\[ p_{cr} = \sigma_{H_{\text{max}}} - (\sigma_{H_{\text{max}}} - \sigma_{H_{\text{min}}}) \cos \psi \sin (\psi - \phi_f)/\sin \phi_f \]  

(15)

where \( \phi_f \) is the friction angle and \( \psi = (\pi/2) - \beta \) in Eqs. (4) and (5) is the angle between the fault normal and the maximum horizontal stress \( \sigma_{H_{\text{max}}} \) (Rice, 1992). Ignoring cohesion, for friction coefficients \( \mu = \mu_f = 0.6-0.9 \), Eq. (15) requires that \( p \) will exceed \( p_{H_{\text{min}}} \) and lead to hydrofracture when \( \psi \) is less than 30-40°. Thus, provided the stress state is homogenous across the fault zone, the observed high angle for \( \sigma_{H_{\text{max}}} \) along the fault is an argument against high pore pressure as an explanation of the weakness of the SAF.

This problem of excessive pressure buildup resulting in hydrofracture before fault slip occurs can be avoided for a mature fault zone which is wide enough to sustain a pore pressure gradient. Two mechanisms have been proposed for producing pore pressure which increases from the fault zone boundaries to the fault core axis. Byerlee (1990, 1993) suggested that for fine-grained fault gouge (especially containing hydrophilic clays or other minerals) a threshold pore pressure gradient may exist below which pore fluid will not flow. In this case, time-dependent compaction of the fault gouge material would lead to the establishment of a stable pore pressure profile in which pressure increases toward the fault zone axis by an amount equal to the threshold pressure gradient of the gouge material. For small faults that have narrow gouge zones, this pressure gradient is inconsequential. However, for mature fault zones such as the San Andreas where the fault zone may be hundreds of meters wide, a stable pore pressure profile could exist in which there is no pore fluid flow between, for example, hydrostatic pore pressure on the boundaries of the fault zone and near lithostatic pore pressure along the fault zone axis. In this case, the fault would have essentially no shear strength. In this model, rising fluid pressure within the fault zone leads to loss of strength and yielding. The resulting deformation changes the magnitude and orientation of principal stresses within the fault so that the pore pressure remains less than the minimum principal stress everywhere.

Rice (1992) proposed an alternative model in which a deep-rooted fault such as the San Andreas would have higher permeability than the surrounding crustal rock and would act as a conduit for the upward migration of mantle-derived fluids. Based on laboratory observations that permeability is highly sensitive to effective mean stress (Brace et al., 1968; Pratt et al., 1977; Brace, 1978; Morrow et al., 1994), Rice’s model predicts that pore fluid pressure at the base of the fault zone increases until a steady-state condition is achieved in which fluid flows up the axis of the fault zone and leaks out into the surrounding country rock. In this case, as in Byerlee’s model, a pore pressure gradient is established between the fault zone axis and the country rock, so that pore pressure in the interior of the fault zone can be near-lithostatic and at the same time be less than the minimum principal stress throughout the fault zone. These models must be verified by direct observation over many years since drilling into the fault zone is likely to disturb the local pore pressure and the re-establishment of a steady-state condition may be quite slow.

We now examine the fourth possibility that the San Andreas is weak strictly as the result of the fault gouge mineral strength. Although few minerals have frictional strengths significantly greater than \( \mu = 0.85 \), there are numerous examples of potential fault gouge materials that are weaker than Byerlee’s law. For example, biotite (Kronenberg et al., 1990) and graphite (Morrow et al., 2000) have an easy cleavage and weak interlayer bonding which results in a low shear strength. In this regard, Byerlee’s law can be thought of as a limiting shear strength for deformation of a granular aggregate or Coulomb material between rigid driving blocks. When weak minerals are sheared in similar geometries, frictional strength can be ranked according to mineral bond strength (Morrow et al., 2000). Laboratory determinations of
friction for a range of rock types and minerals that are found in fault zones are summarized in Figure 9. Most mineral strengths in Figure 9 with friction below 0.7 have been tested wet since the presence of pore water can contribute significantly to low strength of some minerals. Good examples are smectites (montmorillonite clay) and chrysotile (serpentine) which have a strong affinity for interlayer or adsorbed water and exhibit remarkably low shear strength when wet (Moore et al., 1983, 1996, 1997; Reinen et al., 1991, 1994; Morrow et al., 1992). In this case, increased temperature or mean stress tends to drive off adsorbed water and leads to increased shear strength (Fig. 9) (Moore et al., 1997).

Montmorillonite is an expandable clay that, because of its low shear strength and general abundance, has been suggested as a contributor to low strength faults (Wu et al., 1975; Wang, 1984; Morrow et al., 1992). At room temperature wet montmorillonite has a coefficient of friction of about 0.2. As we mentioned, this low shear strength is the result of weakly bonded interlayer water attracted to the charged surfaces of the clay platelets. Under similar conditions, dry montmorillonite has \( \mu \sim 0.4 \) (Moore et al., 1983; Morrow et al., 2000). Consequently, at shallow depths (low pressure and temperature) montmorillonite comes close to satisfying the heat flow and stress orientation constraints for the SAF. However, shear strength tests on montmorillonite at conditions representing increasing burial depth show two effects. First, the higher temperatures and pressures associated with the deeper fault zone tend to drive off the interlayer water and result in increased frictional strength. Secondly, montmorillonite becomes chemically unstable with increasing temperature and first dewatered and then reverts to illite by approximately 400°C (Wu et al., 1975; Wang, 1984). Illite has a significantly higher shear strength than montmorillonite and cannot satisfy the weak SAF constraints. Above approximately 100°C mixed-layer montmorillonite/illite clay may be expected in the fault zone (Freed and Peacor, 1989; Meunier and Velde, 1989). As a result, the deeper part of the fault is expected to rapidly increase in strength (Fig. 8).

Another potential candidate for limiting shear strength of the San Andreas fault is the chrysotile form of serpentine. Due to its strong affinity for adsorbed water, chrysotile is one of the weakest likely fault zone minerals (room temperature friction is \( \mu \sim 0.2 \) and decreases to a minimum strength of \( \mu \sim 0.1 \) at 100°C (Moore et al., 1996, 1997). Furthermore, chrysotile is associated especially with the creeping section of the SAF in central California (Coleman and Keith, 1971; Irwin and Barnes, 1975). However, as in the case of montmorillonite, chrysotile also increases strength with temperature and pressure as shown in Figures 8 and 9. Although it is possible that other less-common minerals exist that may retain a low coefficient of friction to relatively high temperature (i.e., brucite, Morrow et al., 2000), the general trend of increasing strength of common fault gouge constituents with increasing pressure and temperature makes it unlikely that a weak San Andreas fault zone can be explained solely in terms of mineralogy.

In summary, a self-consistent workable model for a weak San Andreas fault can be based on elevated fluid pressure. A combination of elevated fluid pressure and inherently weak fault-zone mineralogy should also satisfy the heat flow constraints and has the advantage that less severe pore fluid overpressures are required. Dynamic weakening can also contribute to reduced frictional heating and needs to be understood more fully. It is possible that all three mechanisms will be found to play significant roles in contributing to the apparent low strength of the San Andreas fault.

3.3 Brittle–Ductile Models

Laboratory observations suggest earthquakes should be limited to conditions where the differential stress does not exceed the ductile flow strength of rocks, because under most circumstances ductile deformation, deformation which allows large strain without loss of strength, precludes dynamic stress drop. Ductile yield strength, equal to the differential stress \( \sigma_y \) during yielding, rapidly decreases with temperature as follows from the creep flow law

\[
\dot{\varepsilon} = A \sigma_y \exp \left( \frac{-E}{RT} \right)
\]
where \( R \) is the gas constant, \( E \) is the activation energy, \( n \) is a dimensionless constant, and \( A \) is a constant (MPa\(^{-n}\) sec\(^{-1}\)). Seismicity occurs at conditions where brittle strength [Eq. (14)] is lower than Eq. (16) (Fig. 10a). Diagrams, originated by Goetze and Evans (1979), such as Figure 10, where quartz strength is assumed representative of the continental crust and olivine of oceanic crust, can explain the general depth dependent features of natural seismicity (Fig. 10b and c). However uncertainties in the form of the flow law (Brace and Kohlstedt, 1980; Kirby, 1980), uncertainty of the Earth’s temperature gradient and the strain rate dependence of Eq. (16) do not allow the strength or other material properties to be well constrained by natural seismicity.

An additional consideration in determining the depth extent of seismicity in the Earth is to allow for the possibility of aseismic fault slip within the brittle field. Studies of frictional sliding indicate that faults can slide stably under some circumstances. Thus it is possible that the maximum depth extent of seismicity is not determined by the transition to predominately ductile deformation but instead by a transition from unstable to stable frictional sliding (Scholz, 1990; Blanpied et al., 1991, 1995; Chester, 1994, 1995).

4. Frictional Instability and Earthquakes

If earthquakes correspond to slip on preexisting fault surfaces, as opposed to failure of intact rock, then the conditions leading to unstable fault slip are relevant to studies of earthquake nucleation. Unstable sliding is only possible if fault strength drops following the onset of slip so that stored elastic energy from the surrounding rock can be used to drive the instability. The simplest description of fault rheology [Eq. (2b)] represents shear resistance strictly in terms of (effective) normal stress and precludes the possibility of weakening due to time-, temperature-, slip-, or slip rate-dependent processes. Under proper conditions, Eq. (2b) can allow for instability driven by increasing pore fluid pressure or possibly a rotation of the stress field loading a fault, but there are clear examples of fault instability in laboratory tests that do not involve these mechanisms. For frictional slip, instability depends only indirectly on the absolute level of strength. Rather, instability depends on inequality between fault strength and the stress available to shear the fault, which is supplied by the loading system. In basic terms, elastic energy is stored in the surrounding rock as tectonic processes load the brittle crust. When the strength of a fault is reached, sliding will commence and stored elastic energy is converted into frictional heat or surface energy as the fault rock is crushed. If an increment of slip on the fault surface reduces the fault strength faster than the driving stress is reduced, an unstable condition develops in which fault slip can accelerate into an earthquake and some of the stored elastic energy is radiated as seismic waves (Byerlee, 1970; Dieterich, 1979). In the laboratory, the test sample, piston, and load frame store elastic energy just as the brittle crust stores energy around a fault. Steady loading followed by dynamic fault slip is referred to as the “stick-slip” cycle and has been proposed as the laboratory counterpart of the earthquake cycle (Fig. 11) (Brace and Byerlee, 1966). Indeed laboratory stick-slip events are similar to earthquakes and if considered in terms of the dimension of rupture, earthquakes and other rock seismic faulting events form a continuum over at least 10 orders of magnitude. Great earthquakes occur on faults with linear
dimensions of hundreds of kilometers, whereas microseismicity and mining rock bursts represent fault instabilities spanning meters to tens of meters. Typical laboratory faults have lengths of 0.01–0.1 m and for specialized test machines have 2 m fault length (Okubo and Dieterich, 1984). At an even smaller scale, microcracking and associated ultrasonic acoustic emissions occur over a range from micrometers to millimeters and have been used as a laboratory analogue for many earthquake phenomena (Lockner, 1993a).

The elastic energy which drives instability can lead to a net drop in stress as shown from direct high speed stress measurements recorded during model earthquakes (Okubo and Dieterich, 1981, 1984, 1986; Lockner et al., 1982; Lockner and Okubo, 1983; Ohnaka and Kuwahara, 1990). The simplest idealization of this stress drop is static and dynamic friction (Jaeger and Cook, 1971; Rice and Tse, 1986; Scholz, 1990), that is, a fault at rest has static friction represented by Eq. (2b) or (12) but once it begins to slip, strength drops immediately to a characteristic dynamic friction level. This concept proves to be unrealistic when applied to rupture nucleation due to the assumption of an abrupt change in fault strength (Ruina, 1983); strength drop for real materials is not instantaneous, and requires an apparent time or displacement weakening. An improvement over the static and dynamic friction model which well represents many aspects of the laboratory observations (Okubo and Dieterich, 1984, 1986; Ohnaka et al., 1986; Ohnaka and Kuwahara, 1990) is the incorporation of strain weakening, also known as displacement or slip weakening. This slip-weakening behavior has been described theoretically (Andrews, 1976; Rice, 1980; Rudnicki, 1980) and has also been confirmed experimentally for brittle failure of intact laboratory samples (Rice, 1980; Wong, 1982b; Lockner et al., 1991) (see Section 4.2). In this case, stress decreases from static to dynamic friction over a characteristic displacement $d_c$ or a characteristic strain $\varepsilon_c = d_c/v$, where $w$ is the thickness of the deforming layer. Simple slip weakening functions include linear and exponential decay, and because of their mathematical simplicity, are useful in dynamic rupture propagation models.

If the elastic stiffness of the loading system is expressed as $k = d\tau/db$, where $b$ is fault slip, then a requirement for instability at constant normal stress can be expressed in terms of a critical stiffness (Dieterich, 1979)

$$ k_c = \frac{d\mu}{db} \sigma_n \approx \frac{\Delta\mu\sigma_n}{d_c} \tag{17} $$

where $\Delta\mu$ is the drop in coefficient of frictional resistance. A slip instability can then occur for $k < k_c$. This relation has been verified experimentally for samples containing well-characterized displacement weakening (Lockner and Byerlee, 1990). In laboratory experiments, stiffness is the combined stiffness of the intact portion of the sample and the testing apparatus, whereas in the Earth the stiffness is related to the elastic modulus of the rock and the geometry of the portion of the fault that is slipping (e.g., Byerlee, 1970; Walsh, 1971; Dieterich, 1978a, 1979).

### 4.1 Characteristics of Earthquake Nucleation

In the context of earthquake nucleation, failure criteria such as Amontons’ law establish a yield stress but are unable to explain premonitory phenomena such as creep or rupture nucleation. Although some moment inversions of radiated seismic energy suggest that earthquakes have distinct nucleation and propagation phases (e.g., Ellsworth and Beroza, 1995, 1998), only the final, rapid slip portion of the nucleation phase can be studied with seismic wave analysis. Attempts to measure natural fault slip during the early stages of fault nucleation have been unsuccessful (Johnston et al., 1987; Linde and Johnston, 1989; Chapter 36 by Johnston and Linde). Given the high sensitivity of the strain meters in these studies, analysis of limited data suggests premonitory creep for moderate-sized earthquakes is restricted to nucleation patches that are tens of meters or less in linear dimension. Although field measurements of deformation have been unsuccessful in detecting the early strain associated with rupture nucleation, laboratory studies consistently demonstrate a characteristic accelerating slip phase that develops into the dynamic instability. Figure 11 shows the primary characteristics of

FIGURE 11 Stick–slip. Frictional strength vs. time during sliding on initially bare surfaces of Westerly granite at room temperature, 50 MPa confining pressure and 0.1 μm sec$^{-1}$ loading velocity. The onset of detectable slip ($t_{onset}$) and the peak stress ($t_{peak}$) differ from the failure time ($t_{fail}$). Unstable slip shown by vertical line segments. The inset shows the spring and slider block model appropriate for laboratory experiments.
slowly nucleating laboratory failure. (1) Fault slip $\delta$ accretes as the fault is loaded at a constant rate, as evidenced by the deviation from linear increase in stress with time [shear stress on the fault $\tau = k(V_f t - \delta)$]. This behavior, if appropriate for natural earthquake failure allows for the possibility of detectable precursors such as strain or strain-related phenomena. For example, precursory strain should be associated with dilatancy (Nur, 1972; Scholz et al., 1973) and a number of precursory phenomena measured in the laboratory [i.e., changes in acoustic velocity and attenuation (Lockner et al., 1977, 1992a; Yanagidani et al., 1985; Masuda et al., 1990; Satoh et al., 1990), Gutenberg–Richter b-value (Weeks et al., 1978; Meredith et al., 1990), electrical resistivity (Lockner and Byerlee, 1986), and permeability (Zoback and Byerlee, 1975; Morrow et al., 1981)]. (2) Peak stress occurs before failure and post-peak accelerating slip is accompanied by slip weakening (e.g., Okubo and Dieterich, 1984; Ohnaka et al., 1986; Ohnaka and Kuwahara, 1990; Lockner and Beeler, 1999) indicating no unique threshold failure stress. (3) Failure is delayed: there is lag time between the onset of detectable slip ($t_{\text{onset}}$) or the time of peak stress ($t_{\text{peak}}$) and the time of failure ($t_{\text{fail}}$). The reader is reminded that all of these features have direct counterparts in intact rock failure (Fig. 3 and, for example, Lockner et al., 1992a; Lockner, 1998).

4.2 Slip-weakening Model and Unstable Rupture

Any mechanism that results in a loss in shear resistance that is more rapid than the unloading of the surrounding region has the potential for producing slip instability. One obvious example is the fracture of intact rock or, in the case of an irregular or segmented fault, the rupture of an asperity or competent step-over. Laboratory measurements (Fig. 1) indicate that strength reduction resulting from brittle fracture can be 20–50% of peak strength whereas strength loss associated with stick-slip of nominally flat, ground rock surfaces (see Section 4.3) is approximately 5–10% (Okubo and Dieterich, 1986; Ohnaka and Yamashita, 1989). Laboratory measurements have shown that dilatancy-induced pore pressure loss in a hydraulically isolated fault can lead to a 20% increase in effective normal stress (Lockner and Byerlee, 1994). Collapse of this open pore structure with continued deformation (Blandpied et al., 1992) could repressurize pore fluid and provide a mechanism for triggering instability. The idealized flat, parallel fault surfaces used in laboratory tests to determine fault constitutive properties as described in the following section by Eqs. (18) are designed specifically to avoid complicating factors related to geometric fault complexity (and fluid pressure effects). In reality, natural faults are commonly irregular (fractal) over a broad range of scale (Brown and Scholz, 1985b; Scholz and Aviles, 1986; Power et al., 1988; Moore and Byerlee, 1989, 1991). Some aspects of the geometric complexity of faults have been explored (Andrews, 1989) but no framework exists for systematically analyzing this issue.

Ohnaka and co-workers developed an empirical model for unstable rupture that is described solely in terms of fault slip. The model is intended to describe the nucleation and growth of unstable rupture as measured on a biaxial press with granite sample dimension of $280 \times 280 \times 50$ mm (Ohnaka et al., 1986). This apparatus, similar to but smaller than the 2-meter biaxial press (Okubo and Dieterich, 1981), has a fault surface that is large relative to the nucleation patch size for unstable rupture. The idealized slip instability is based on a slip-weakening model as discussed by Rice (1980) and Rudnicki (1980). The model includes a breakdown zone associated with an advancing crack tip in which shear stress $\tau$ near the crack tip is a function of shear displacement $\delta$ (Fig. 12a). High-speed records of the passage of a rupture front in a 2-meter fault are shown for comparison in Figure 12b (after Okubo and Dieterich, 1986). Note the characteristic stress rise as the rupture front approaches, followed by a rapid weakening to the dynamic sliding strength over displacement $\delta_c$. When shear resistance at a point on the fault is plotted as a function of fault

![Figure 12](image-url)
slip (Rice, 1980) (Fig. 13), energy release rate $G_{II}$ associated with the rupture propagation is represented by the area under the stress curve in excess of the frictional sliding resistance $E_f$. Similar results have been reported on the 0.4-meter biaxial press (Ohnaka et al., 1986; Ohnaka and Yamashita, 1989; Ohnaka and Kuwahara, 1990; Yamashita and Ohnaka, 1991) (Fig. 13b). Scaling relationships have been proposed (Ohnaka and Shen, 1999) for critical slip distance, nucleation patch size and energy release rate assuming a dependence of nucleation on a characteristic surface irregularity size (Ohnaka and Yamashita, 1989; Ohnaka and Kuwahara, 1990; Yamashita and Ohnaka, 1991). The issue of fault heterogeneity arises in many contexts in earthquake processes including nucleation and stopping phases of earthquakes, abundance of high-frequency radiated energy, and earthquake statistics (Rice and Ben-Zion, 1996).

Fracture data for an initially intact sandstone sample are shown in Figure 13c where differential stress is plotted versus inelastic (or permanent) axial strain. This plot might represent, for example, the rupture of an asperity or interlocked region on a fault. Fracture nucleation coincides with the onset of rapid slip weakening. This portion of the breakdown curve is similar to the breakdown curve for the preexisting fault (Fig. 13b). The stress drop can be significantly greater for fracture of crystalline rock such as granite that has significant cohesion. The most notable difference between Figures 13b and c is the large amount of work done in loading the intact rock to peak stress. This work is expended in growing microcracks and in dense crystalline rock is accompanied by a volume increase and changes in a host of related material properties such as wave speed, electrical resistivity and permeability. Since these changes occur quasi-statically in response to tectonic loading, this is the effect most commonly appealed to in searching for premonitory phenomena.

This section ends with a comparison of time–position plots of fracture growth on a preexisting fault and in an intact rock. Figure 14a shows the growth of an instability in a precut granite sample with a 0.4-meter fault (Yamashita and Ohnaka, 1991). Solid circles denote the time the shear stress reaches the peak value, indicating the onset of instability. Open circles denote the time when the shear stress has dropped to the dynamic frictional stress level, indicating the passage of the breakdown zone. Note the finite width of the breakdown zone once the fracture is growing at near shear wave speed. Figure 14b is a space–time plot of acoustic emission rate recorded during shear fracture growth in an initially intact 0.2-meter long granite sample (Lockner and Byerlee, 1992). In this example, the fracture propagated quasi-statically as indicated by the slow advance rate of acoustic emission activity across the sample. The nucleation phase had a width of 2–3 cm in both the intact and faulted samples. Following nucleation, the narrow band of intense acoustic emission activity indicated the position of the advancing process zone.

**FIGURE 13** (a) Idealized stress–displacement relation for rupture propagation shown in Figure 12a. Energy release rate per unit area of fault surface is represented by $G$. (b) High-speed data recording showing stress-displacement relation during passage of an unstable rupture. (c) Normalized stress–inelastic axial strain plot for failure and fracture propagation in initially intact Berea sandstone.
4.3 Rate and State Friction Equations

The displacement-dependent rupture model described, for example, in Ohnaka and Shen (1999) describes fault conditions following the onset of rapid slip. Although it can account for failure characteristics such as post-peak loss of strength (Fig. 11), it does not address the characteristic prefailure deformation of slow, time-dependent creep (referred to as phase I in Ohnaka and Kuwahara, 1990). The dependent variables in frictional sliding, especially at slip rates of $10^{-9} - 10^{-4}$ m sec$^{-1}$, are more easily revealed in rate-stepping tests in a stiff testing machine ($k > k_c$) than in stick-slip or rupture nucleation experiments. Experimental measurements indicate the characteristic response of rock friction to changes in sliding velocity is somewhat complicated, consisting of two apparently additive log velocity-dependent effects: a positive, instantaneous dependence (direct effect) and a negative dependence which evolves approximately exponentially with displacement (evolution effect) (Dieterich, 1978b, 1979) (Fig. 15). These effects are generally small and of similar magnitude (typically a few percent of the total friction for a decade change in velocity (Blanpied et al., 1998a). Since the effects are opposite in sign, they tend to cancel and the net dependence on slip velocity is even smaller and can be either positive or negative (Dieterich, 1978b, 1979; Tullis and Weeks, 1986).

The detailed response of fault frictional strength, at room temperature and modest normal stress (typically 5–200 MPa), to changes in sliding velocity, whether during rate-stepping tests or during stick-slip sliding, can be well described by what are usually referred to as rate and state variable constitutive equations, developed by Ruina (1983) as a generalization of equations of Dieterich (1978a, 1979),

$$\mu = \mu_0 + a \ln \frac{V}{V_0} + b \ln \frac{\theta V_0}{d_c}$$  \hspace{1cm} (18a)

$$\frac{d\theta}{dt} = 1 - \frac{V\theta}{d_c}$$  \hspace{1cm} (18b)

(Rice and Ruina, 1983; Ruina, 1983). Here $\mu_0$ is the frictional resistance to slip at an arbitrary reference sliding velocity $V_0$. $\mu_0$ can be thought of as the first order frictional strength, such as described by Eq. (2) or (12). The other two terms in Eq. (18a) are the rate and state dependencies which are second order. The $a \ln(V/V_0)$ term describes the instantaneous rate dependence of strength (direct effect in Fig. 15) and the last term in Eq. (18a) represents the state dependence (evolution effect in Fig. 15). As written here, $\theta$ is the “state variable,” having the dimension of time. The form of $\theta$ prescribed in Eq. (18b) is one of a number of empirical equations that have been used in this context. [See Linker and Dieterich (1992) and Ruina (1983) for detailed discussions of the different forms of state and their properties.] The steady-state value of state, as deduced from Eq. (18b) is $\theta_{\infty} = d_c/V$. When substituted in Eq. (18a) this defines the

![Figure 14](image_url)  \hspace{1cm} (a) Nucleation and initial propagation phase of unstable rupture event on a preexisting fault in granite. (b) Plot of acoustic emission rate in region across center of initially intact granite sample.

![Figure 15](image_url)  \hspace{1cm} (b) Typical slip-rate dependence of fault strength for granite at 25 MPa normal stress and room temperature in response to an imposed change in sliding velocity. The loading velocity is decreased by an order of magnitude at displacement corresponding to the vertical dashed line. The net rate dependence $d\mu/d\ln V$ is the sum of two opposing effects.

Rocks Failures and Earthquakes
steady-state rate dependence of fault strength \(d\mu_s/d\ln V = b - a \) which determines the sliding stability.

In the case where the dependence of steady-state frictional strength on the logarithm of sliding velocity is negative \((d\mu_s/d\ln V < 0)\), any perturbation that causes the fault to accelerate will lead to a loss of strength and will encourage further slip acceleration. In this case, slip instability is possible. Conversely, for neutral or positive rate dependence \((d\mu_s/d\ln V \geq 0)\) sliding will tend to be stable (Rice and Ruina, 1983). Linearized perturbation analysis of Eq. (18) (Rice and Ruina, 1983; Ruina, 1983) results in an expression for critical stiffness analogous to Eq. (17)

\[
k_c = \frac{d\mu_s}{d\ln V} \sigma_n
\]

(19)

where \(\mu_s\) is the steady-state frictional resistance and \(d_c\) is the characteristic displacement of the evolution effect. Accordingly, instability is favored by more negative values of \(d\mu_s/d\ln V\) and by low values of \(k\) (more compliant loading systems) (Rice and Ruina, 1983; Gu et al., 1984).

Laboratory measurements of rate dependence have been conducted using two-fault configurations: initially bare rock surfaces which consist of ground, flat, parallel rock which has been roughened with a coarse (typically 60–80 grit) abrasive, and simulated gouge layers which consist of powdered rock flour (typically <90 \(\mu\)m diameter) sheared between rigid blocks. The rate dependence of quartzo-feldspathic rocks, on which the constitutive relations Eq. (18) were based, can depend on whether or not the fault zone contains a thick gouge layer. Thick quartzo-feldspathic gouge layers have positive rate dependence unless shear displacement becomes localized within the layer (Dieterich, 1981; Marone and Kilgore, 1993) (also see review by Marone, 1998). Conversely initially bare rock surfaces of quartzo-feldspathic rock (Dieterich, 1978b; Tullis and Weeks, 1986) are rate weakening. In our review of rock friction measurements relevant to earthquakes we focus primarily on measurements that relate directly to frictional instability. We have therefore omitted studies on the frictional properties of gouge layers which show variability in the sign of the rate dependence (see summary by Beeler et al., 1996), we focus predominately on the results from rate-weakening experiments on initially bare rock surfaces. Slip between such surfaces remains localized even at large displacement (Beeler et al., 1996) and such faults show little tendency for fault zone width growth after an initial period of rapid wear (Biegel et al., 1989) because they are flat at long wavelengths [see Power et al. (1988) and as discussed below].

In contrast to experimental faults, a continually expanding fault zone thickness is expected for natural faults (Robertson, 1983). For example, the San Andreas fault zone, with total offset of hundreds of kilometers, has a thickness of as much as a kilometer in places (Chapter 29 by Sibson). Increased thickness of natural faults may result from geometrical and dynamic effects during shear. Natural rock surfaces such as fracture and joint surfaces have a fractal roughness (Brown and Scholz, 1985a). As a result, increasing displacement of an initially matched shear fracture (proto-fault) leads to an increasing volume of mismatched surface irregularity and therefore increasing wear material (gouge) with displacement (Power et al., 1988). During earthquake slip, large dynamic stresses that develop in front of an advancing shear rupture during propagation may be higher on off-fault planes than on the fault plane itself (Rice, 1980). These high stresses may induce branching or off-fault deformation which may lead to nonplanar ruptures, surface roughening and gradual growth of the damage zone with successive earthquakes. The experimental observations thus require that slip during earthquake nucleation is localized so that \(d\mu_s/d\ln V < 0\), even within very wide fault zones. At least in experiments to date on gouge-filled faults, fault slip at large displacement does not localize to the degree that the rate dependence is negative (Beeler et al., 1996). It is not known how such localization arises in mature natural fault zones.

### 4.4 Origin of Delayed Failure in Friction and Fracture

The two opposing rate-dependent effects seen in rate stepping tests (Fig. 15) can be qualitatively related to the onset of failure in stick-slip experiments (Fig. 11) in a straightforward manner. Obviously, frictional failure requires that the rate weakening term exceeds the rate strengthening term for the stress to drop as slip accelerates. However, negative rate dependence alone leads to rapid failure following a threshold failure stress and therefore does not allow for delayed failure or precursory slip. It is the long duration of the acceleration to failure (Fig. 16) which most clearly distinguishes observed failure from failure predicted by simple static and dynamic strength. Prior to the peak strength, which occurs when the sliding velocity equals the loading velocity \(V = V_L\), the sliding velocity increases (Fig. 16)

![FIGURE 16 Acceleration to failure during slip of granite at room temperature and 50 MPa confining pressure.](image)
and yet the stress on the fault also increases (Fig. 11) indicating that fault strength depends positively on sliding velocity. Thus this precursory creep during delayed failure is directly attributed to the direct effect (Fig. 15). A similar positive dependence of strength on deformation rate is seen during intact failure (e.g., Lockner, 1998). Delayed failure in response to stress perturbations, which is characteristic of both frictional failure and fracture, is consistent with natural observations of Omori aftershock sequences (Scholz, 1968b; Lockner, 1993a; Dieterich, 1994) (see Section 4.5) and other triggered seismic phenomenon (Gomberg et al., 1998) in which failure is delayed following a stress increase.

Brittle fracture at high normal stress is involved in both abrasion during frictional sliding of bare fault surfaces and grain fracture during shearing of granular gouge layers. Therefore, all kinds of experimental shear deformation in the brittle field should share characteristics associated with fracture. Fracture theory as well as experimental evidence suggest log rate dependent resistance during nucleation of a single fracture. The stress–inelastic strain relation for subcritical fracture (i.e., slow, stable fracture growth at stress intensity less than the critical value as discussed in Section 2.3.2), results from equating strain rate (equivalently velocity $V$) resulting from crack propagation to the rate of chemical reaction at the crack tip (Lockner, 1993b, 1998) where reaction rate is

$$ v = v_0 \exp \left( \frac{-E + \Omega m \sigma - \Omega m \gamma / \chi}{RT} \right) $$

(e.g., Charles and Hillig, 1962). $E$ is the activation energy, $\Omega$ is the activation volume, $\Omega m$ molar volume, $\gamma$ is the surface energy, $\chi$ is the radius of curvature of the tip, $v_0$ is a constant, $R$ is the gas constant, $T$ is temperature and $\sigma$ is the stress at the crack tip. The stress term is akin to the pressure dependence normally associated with an activation volume, such that the reaction rate is faster if the stress is elevated. As has been suggested by numerous authors (see review by Costin (1987)), this stress sensitivity of crack propagation rate likely underlies the established creep-like behavior of rock at low temperature (Scholz, 1968a) and the log rate dependence of the failure stress of intact rock (e.g., Lockner, 1998). Based on consideration of Eq. (20), and due to the involvement of fracture in frictional sliding, the subcritical component of fracture is a likely cause of the instantaneous positive rate dependence of rock friction.

### 4.5 Implications of Delayed Failure

Laboratory observations of delayed failure predict that stress changes affect the timing of earthquakes, and many natural observations of delayed failure are consistent with these predictions. For example, assuming that aftershocks result from a stress change superimposed on a constant tectonic loading rate, Omori aftershock sequences indicate that the sensitivity of an aftershock source fault to stress depends on its temporal proximity to failure. That is, because aftershock rates immediately following a main shock static stress change are higher than they are at later times, faults closer to failure are advanced toward failure by a different amount than those that are far from failure. Such a time dependence of aftershock rates follows directly from Eqs. (18) (Dieterich, 1994). Assuming, based on the results of studies of static stress change (Stein and Lisowski, 1983; Harris and Simpson, 1992; Reasenberg and Simpson, 1992), that aftershock failures are triggered by a stress increase and that ongoing tectonic loading provides a constant rate of stress increase, then seismicity rate $r$ as a function of time $t$ following an increase in shear stress $\Delta \sigma$ is

$$ r(t) = \frac{r_0}{1 - [1 - \exp(-\Delta \sigma / a)] \exp(-\gamma t / a)} $$

where $r_0$ is the seismicity rate prior to the stress change, $\gamma$ is the stressing rate, and $a$ is the constitutive parameter which scales the size of the direct effect in Eqs. (18) (Dieterich, 1994). Equation (21) predicts an Omori decay consistent with field observations (Fig. 17); a similar prediction was made by Scholz (1968b) who considered aftershocks as small fracture events, the natural analogue to acoustic emissions, instead of unstable slip of pre-existing faults.

Delay between the application of a stress change and failure, such as seen in laboratory experiments and predicted by constitutive Eqs. (18), are also consistent with the lack of tidal triggering of nonvolcanic earthquakes. Because daily variations in stress due to tidal forces (0.001–0.004 MPa) greatly exceed the daily accumulation of tectonic stress, earthquake occurrence would be strongly correlated with the tides if failure coincided with a stress threshold [e.g., Eq. (2)].

![Figure 17](image-url) Time-dependent effects. Seismicity rate following a stress step from Eq. (21). The constitutive equation in this calculation is a simplification of Eq. (18). Seismicity rate follows an Omori decay and returns to the background rate over a characteristic time $t_o = a \sigma^2 / \gamma$ (Dieterich, 1994).
However, analysis of natural occurrence suggests weak (Vidale et al., 1998) or no (Heaton, 1982; Hartzell and Heaton, 1989) correlation between earthquakes and Earth tides. Numerical simulations of combined tectonic and tidal loading of fault populations obeying Eqs. (18) suggest that failure is insensitive to stress change with the amplitude and frequency of the tidal stress (Dieterich, 1987). Similar conclusions are reached from stick-slip experiments where constant loading was modulated by a small amplitude periodic signal. Correlation between the periodic loading and the time of failure occurred when the periodic amplitude was 0.05–0.1 MPa (Lockner and Beeler, 1999). Simple extrapolation of this result to the Earth suggests that approximately 1% of earthquakes on the San Andreas fault should be correlated with the tides; a result that is consistent with absent or weak correlations in field data.

4.6 Limits of Frictional Instability Predicted by Laboratory Observations

The rate and state variable constitutive Eqs. (18), which well describe many laboratory observations, can qualitatively explain a wide range of averaged seismic observations, e.g., realistic earthquake cycles (Tse and Rice, 1986), Omori aftershock sequences (Dieterich, 1994), the lack of tidal triggering of earthquakes (Dieterich, 1987; Lockner and Beeler, 1999), variations in stress drop with recurrence and loading velocity (Gu and Wong, 1991), foreshocks and earthquake clustering (Das and Scholz, 1981; Dieterich and Kilgore, 1996), earthquake aftershock (Marone et al., 1991), precursory slip (Dieterich, 1994), and remotely triggered earthquakes (Gomberg et al., 1997, 1998). Most important for frictional instability to be a viable explanation for natural seismicity is that transitions from stable (aseismic) and unstable (seismic) frictional sliding seen in experiments correspond to the depth range of observed seismicity. For example, seismicity on the San Andreas fault system is most active between about 4 and 15–20 km (Fig. 10). Two recent complementary experimental studies of quartz gouge and granite gouge attempt to directly predict the depth extent of seismicity by determining the rate dependence of frictional strength at temperatures spanning the entire seismogenic zone (Blanpied et al., 1991, 1995; Chester and Higgs, 1992; Chester, 1994, 1995).

At the top of the seismic zone, corresponding to low temperature, the rate dependence of granite frictional strength is positive, in agreement with room temperature experiments (Byerlee and Summers, 1976; Solberg and Byerlee, 1984; Morrow and Byerlee, 1989; Marone et al., 1990). The rate dependence of granite gouge undergoes a transition to rate weakening between 25 and 50°C (Blanpied et al., 1991, 1995) (Fig. 18). However when converted to depth this prediction corresponds closely to the field observations of a seismic/aseismic transition at around 4 km only if the loading rates are orders of magnitude faster than the plate motion rate on the San Andreas. A similar conclusion is reached from application of the quartz data. Quartz does not show rate strengthening at the lowest strain rate and the lowest temperatures (Chester and Higgs, 1992), although thicker gouge layers are rate strengthening at room temperature and slightly elevated temperature (Marone et al., 1990; Chester, 1994). Another explanation of the aseismic to seismic transition, that the depth extent of a thick gouge-filled fault zone controls the lack of shallow seismicity (Marone and Scholz, 1988; Marone et al., 1991), is consistent with field evidence. However, the Marone study is based on data at sliding velocities much higher than plate motion rates and does not consider the temperature dependence observed in the Chester and Blanpied studies.

At higher temperatures, a lower seismic/aseismic transition is also predicted from the laboratory data. However, the depth of the transition is not fully consistent with observed seismicity. Blanpied et al. (1991, 1995), Chester and Higgs (1992), and Chester (1994, 1995) studied friction in the presence of water at temperatures up to 600°C. In the Blanpied study, steady-state rate dependence of granite was determined at 100 MPa water pressure and 400 MPa confining pressure. The Chester study involved stress relaxation tests at 100 MPa water pressure and 250 MPa confining pressure. Both studies show a broad range of temperatures where the rate dependence is measured or inferred to be negative, permitting frictional
instability from 25–350°C for sliding velocities of 0.1–0.01 μm sec⁻¹ (e.g., Blanpied et al., 1991) (Fig. 18). At higher temperatures, ductile processes such as solution transport creep (pressure solution) begin to affect the strength of load-bearing contacts in the experimental fault zones (Chester and Higgs, 1992). This high temperature regime shows strong velocity strengthening and stable sliding. Although these experiments imply a seismic zone that corresponds roughly to natural observations, the sliding rates, corresponding to loading rates on natural faults, used in the steady-state sliding experiments are 10–100 times faster than typical plate motion on the San Andreas fault. If the results of these studies are extrapolated in strain rate to an appropriate natural condition, the base of the seismogenic zone would occur at about 180°C for quartz (Fig. 18b) and about 240°C for granite (Fig. 18a). This is in agreement with the temperature at the base of the seismogenic zone for the Landers rupture, but much less than the 350–400°C inferred for the San Andreas based on heat flow measurements (Williams, 1996). Thus, while existing experimental studies qualitatively predict many aspects of natural seismicity, the field observations are not matched closely. Better knowledge of fault zone mineralogy and the rate dependence of more typical fault zone gouges (e.g., clays and other alteration products) at these deeper hypocentral depths would be useful in providing more accurate representation of fault properties, particularly in instances such as the San Andreas fault where fault strength may deviate significantly from that of quartzo-feldspathic (Byerlee’s law) materials. The effects of fluids and the role of fluid pressure at elevated temperatures, not taken into account in the friction studies we have discussed so far, may also be important in frictional failure in the Earth.

5. Pore Fluid Effects

Pore fluid can affect fault failure strength in two fundamental ways: (1) pore fluid pressure can reduce the effective normal stress via the effective pressure law Eq. (7) and therefore reduce the failure shear stress in the brittle portion of the crust [Eq. (2) or (12)] as discussed above in the context of fracture, and (2) pore fluid at elevated temperatures can lead to dissolution and precipitation or other chemical reactions which may alter fault strength (see summary by Hickman et al., 1995). Interest in pore fluids has arisen because dry friction cannot explain all of the field observations of faulting. For a classic example, the lateral dimensions of thrust sheets are large enough relative to sheet thickness that the force necessary to overcome the frictional resistance is sufficient to break the hanging wall block unless high pore fluid pressure reduces the effective stress on the sliding surface (Hubbert and Rubey, 1959). Recently, models involving elevated fluid pressures have been proposed to explain the apparent weakness of the San Andreas (Byerlee, 1990; Byrne and Fisher, 1990; Rice, 1992; Sleep and Blanpied, 1992; Lockner and Byerlee, 1995) (see Section 3.2). Presently there is also interest in the role of pore fluid pressure in Coulomb failure models [based on Eq. (2) or (12)] of triggered seismicity [see reviews by Harris (1998), Stein (1999) and Chapter 73 by Harris]. Although few laboratory studies have directly addressed the role of pore fluid in failure, some useful inferences can be drawn from studies of elastic coupling between stress and pore pressure (summarized in Section 5.1) and from experimental studies carried out on faults under conditions where no excess internal fluid pressure occurs (summarized in Section 5.2). Most evidence for the influence of chemical effects on faulting comes from field observations of healed fractures, pressure solution and veining, and alteration products (Power and Tallis, 1989; Parry and Bruhn, 1990; Parry et al., 1991; Chester et al., 1993; Bruhn et al., 1994). The limited relevant laboratory observations summarized in Section 5.3 do not yet make a strong case for chemical effects in earthquake mechanics.

5.1 Role of Poroelastic Effects in Models of Coulomb Failure (New and Provisional Results)

Studies of triggered seismicity have used the stress changes produced by slip of a large earthquake to explain the spatial distribution of subsequent earthquakes in the surrounding region. Such studies commonly assume that earthquakes in the surroundings obey the Coulomb failure criterion [Eq. (2)] and that changes in proximity to failure of a particular fault plane can be determined if the stress changes are known. The change in proximity to failure is the change in Eq. (2) due to earthquake-induced stress

\[ \Delta \sigma_c = \Delta \tau - \mu_f (\Delta \sigma_n - \Delta \rho) \]  

(22a)

Here, \( S_0 \) is assumed constant, and the effective stress law has been invoked. Positive values of \( \Delta \sigma_c \) indicate that the fault plane in question is closer to failure, whereas for negative values the fault is farther from failure. The stress changes used to evaluate Eq. (22) are commonly calculated from dislocation models of the source earthquake slip, assuming the surroundings are linearly elastic. Although laboratory studies provide limits on \( \mu_f \) (e.g., Byerlee, 1978), \( \Delta \rho \) must be estimated using some model of pore fluid response to stress change. In most studies of stress-triggered seismicity, pore pressure change is not considered explicitly and \( \mu_f \) in Eq. (22a) is replaced by a free parameter,

\[ \Delta \sigma_c = \Delta \tau - \mu' \Delta \sigma_n \]  

(22b)

The spatial distribution of earthquakes is consistent with positive values of Eq. (22b) if \( \mu' = 0.4 \) (Stein, 1999).

Current experimental and field observations cannot provide constraints on the choice of appropriate pore pressure response models to use in studies of triggered Coulomb failure.
Experiments suggest that in many cases change in pore pressure in rock and soils is determined by the change in mean stress (Skempton, 1954; Roeloffs, 1996). For a homogeneous isotropic poroelastic medium (Rice and Cleary, 1976; Roeloffs, 1996), the relation between stress change and pore pressure change is

$$\Delta p = B\Delta\sigma_{kk}/3 = B\Delta\sigma_m$$  \hspace{1cm} (23)

where $\sigma_{kk}$ indicates summation over the diagonal elements of the stress tensor and $\sigma_m = \sigma_{kk}/3$ is the mean stress. The parameter $B$ is Skempton’s coefficient, and is limited to the range $0 \leq B < 1$. Relatively sparse experimental determinations of $B$ for rocks indicate a range from 0.5 to 0.9 for granites, sandstones, and marbles (Rice and Cleary, 1976; Roeloffs and Rudnicki, 1985; Roeloffs, 1996).

However, Eq. (23) is based on studies of intact rock samples, not faulted rock, and there are no analogous experimental measurements of the poroelastic response of fault zones. Because they may have accommodated large shear strains and are often associated with hydrothermal alteration, it is unlikely that the poroelastic properties of mature fault zones are identical to the surrounding country rock as assumed in Eq. (23). Cocco and Rice (2002) suggest an inhomogeneous isotropic poroelastic response model,

$$\Delta p = B' \frac{K' K_{nn}}{K_n} \left[ G M_n \Delta\sigma_m + G - G' \frac{G}{G} \Delta\sigma_n \right]$$  \hspace{1cm} (24)

Here the primed and unprimed constants are associated with the fault-zone and country rock, respectively; $M = \lambda + 2G$, $K = \lambda + 2G/3$, and $G$ and $\lambda$ are the Lame elastic constants. According to Eq. (24), the poroelastic response of fault zones depends on changes in both mean stress and normal stress. In the case where $G = G'$, Eq. (24) reduces to Eq. (23); however if $G \gg G'$, as might be expected for mature fault zones, $\Delta p \propto \Delta\sigma_n$ (Cocco and Rice, 2002) [cf. Eq. (22b)].

Similarly, in the case where the fault zone is anisotropic, stress components other than mean stress may be important in determining the pore pressure response to stress change. For example, simulated laboratory faults, mismatched tensile fractures and natural joint surfaces have large fault-normal compliance (Brown and Scholz, 1985b) because normal stress is supported by a few highly stressed asperities. Although the orientations of fluid-filled fractures along active fault zones at depth are generally unknown, if natural faults and associated fault-parallel fractures have an asperity structure similar to mismatched joints and are hydraulically isolated from the adjacent country rock, then

$$\Delta p \propto \Delta\sigma_n$$  \hspace{1cm} (25)

would be the expected pore fluid response (Cocco and Rice, 2002). In this case Eq. (22b) would be an appropriate Coulomb failure model.

While Eqs. (23), (24), and (25) are plausible models of poroelastic response, other measurements indicate that elastic stress–pore–fluid–pressure interactions are more complicated. For example, volumetric changes can be produced by changes in deviatoric stress alone, both in unfaulted (Wang, 1997; Lockner and Stanchits, 2002) and faulted rocks (Woodcock and Roeloffs, 1996). Furthermore, inelastic effects, discussed in the next section, may alter the pore fluid pressure response of seismic fault zones near failure (Segall and Rice, 1995, and references therein).

### 5.2 The Role of Dilatancy and Inelastic Pore Fluid Effects in Frictional Failure (New and Provisional Results)

Drained measurements of pore volume or fault thickness suggest that fault zone pore pressure will vary with velocity, time, and displacement during frictional sliding and frictional failure. This follows from measurements originally made of gouge volume or fault normal displacement during rate step tests on gouge layers by Morrow and Byerlee (1989), Marone et al. (1990), and Lockner and Byerlee (1994) and subsequently also for initially bare surfaces (Wang and Scholz, 1994; Beeler and Tullis, 1997) (Fig. 19a). Following a velocity change there can be both immediate and steady-state changes in dilation or compaction rate within the fault gouge layer. The steady-state changes in dilation or compaction due to displacement rate are generally small compared to changes in friction (Beeler et al., 1996), but immediately after the velocity change can be large (Marone et al., 1990; Sammis and Steacy, 1994). Much like strength evolution, porosity of the fault gouge evolves to steady state over a characteristic displacement. Beeler and Tullis (1997) have argued that rate-dependent changes in porosity are directly related to the state variable in Eqs. (18), at least for initially bare surfaces, based on similar time and slip dependencies between fault thickness and state. Compaction occurs even when the fault sliding rate is negligible, suggesting that fault zone thickness is time dependent (Fig. 19b). The time-dependent compaction under nominally static shear strain follows the same log linear form as fault strength under the same conditions (Dieterich, 1972) (Fig. 19b inset). After a shear zone compacts statically, subsequent shear induces dilatancy (e.g., Fig. 19c). The fault normal displacement $\delta_n$ resulting from opposing effects of linear shear-induced dilatancy and log time-dependent compaction can be represented by Eq. (18b) and

$$\delta_n = \delta_{n0} + \varepsilon \ln \frac{V_0\theta}{d_e}$$  \hspace{1cm} (26)

where compaction is reckoned positive, $\delta_{n0}$ is the steady-state normal displacement when $V = V_0$, and $\varepsilon$ is the negative of the steady-state velocity-dependence of fault thickness $\varepsilon = -d(\delta_n)/d\ln V_{st}$ (Marone and Kilgore, 1993). In Eq. (18b), state changes exponentially with slip and linearly with time and when
used in Eq. (26) yields the linear slip and log time-dependent thickness changes required by the data (Fig. 19b and c). Equations (26) and (18b) lead to a rate-dependent response of fault thickness to changes in sliding velocity consistent with the observations (Fig. 19a). In experiments on initially bare surfaces, the characteristic displacement of thickness evolution following changes in sliding velocity is approximately equivalent to the characteristic displacement of strength evolution. This association between fault thickness or porosity with state is much less pronounced for thick, gouge layers (Beeler and Tullis, 1997), where the characteristic displacement to recover strength differs from that to recover the porosity.

It should be noted that fault gouge porosity is sensitive to normal or mean stress and that laboratory tests have been carried out at either constant normal stress or constant confining pressure. Loading histories of natural faults may be considerably different from either constant normal stress or confining pressure (although the latter case of increasing normal stress with increasing $\sigma_s$ may be closer to the tectonic loading history experienced by thrust or strike-slip faults). Normal faulting may be preceded by a decrease in minimum compressive stress and normal stress. The importance of these naturally occurring stress paths and the effect that they may have on pore pressure and shear strength have not yet been properly explored.

Important implications of experimental observations of rate-dependent porosity change for earthquake nucleation have been considered by Segall and Rice (1995). They extended the original stability analysis of rate and state equations (Rice and Ruina, 1983; Ruina, 1983) to consider the role of pore pressure and conducted numerical simulations of earthquake cycles to determine the effects of pore pressure on strength and fault stability. In this analysis they ignore the small changes in steady-state displacement rate of dilation and assume a steady-state drained porosity

$$\eta_{st} = \eta_0 + \varepsilon \ln \frac{V}{V_0}$$

(27)

where $\varepsilon$ is a constant, and a dependence of porosity on state such that porosity evolves to steady-state over the state characteristic displacement $d_c$, consistent with Eq. (26) and with the laboratory observations. For an undrained porosity, the stiffness for unstable slip must be less than the critical value

$$k_c = \frac{\sigma_{st,eff}}{d_c} \frac{d\mu_{st}}{d\ln V} - \frac{\varepsilon \mu_{st}}{B d_c}$$

(28)

where $B$ is a poroelastic constant (Segall and Rice, 1995). Thus, dilatant faults are found to be more stable in comparison to nondilatant faults; the critical stiffness predicted by Eq. (28) is less in comparison to Eq. (19) due to the additional dilatant term.

Numerical simulations of earthquake cycles using laboratory values of fault strength and porosity conducted by Segall and Rice (1995) do not lead simultaneously to high pore pressure and instability. Thus, to the degree that the Segall and Rice model reflects the dominant poromechanical interactions of fault zones, it implies that high pore pressure is an unlikely cause of the weakness of seismogenic faults such as the San Andreas. However, this conclusion is based on analysis of dilatancy observed during stable sliding of drained faults at room temperature and high porosity (e.g., 11–26%, Marone et al., 1990), conditions which may not be representative of earthquake failure at seismogenic depths. In particular, Eq. (27)

FIGURE 19 Dilatancy during frictional sliding. (a) Changes in fault thickness (proportional to changes in porosity) with sliding velocity. (b) Time dependence of fault thickness. (c) Fault thickness as a function of slip following a hold test for bare surface granite and quartzite (unpublished data courtesy of T. Tullis).
requires that fault thickness (porosity) increases monotonically during failure and that the maximum fault zone porosity is determined by the peak velocity during unstable slip. These predictions are contradicted both by some soil mechanics observations, and by measurements of fault thickness during stress drop. When overcompacted soils are sheared at constant loading rate, porosity may not increase monotonically. Instead, the steady-state thickness can be reached by compaction (e.g., Mandl et al., 1977). In experiments on thick gouge layers where shear stress is artificially reduced at the start of a static hold, stress drop is accompanied by compaction (Karner and Marone, 1998; Nakatani, 1998), presumably consistent with compaction during earthquake stress drop. Compaction accompanying stress drop has also been reported during stick-slip sliding (Beeler and Tullis, unpublished data).

The observations of compaction accompanying shear stress reduction have untested implications for the evolution of fault fluid pressure throughout the seismic cycle. Stress drop will lead to a reduction in effective normal stress under undrained conditions possibly providing a simple “dynamic weakening” of fault strength. This effect would occur at constant normal stress such as in strike-slip environments and may be significant in normal faulting environments where shear stress drop is accompanied by normal stress increase. On the other hand, following compaction, much larger dilatancy occurs during loading to failure (Nakatani, 1998) which should tend to stabilize sliding through dilatancy hardening. Constitutive equations describing the experimental observations have been proposed by Nakatani (1998). However, the implications that this theoretical model has for fault stability have not yet been addressed.

5.3 Solution Mass Transport Phenomenon
(New and Provisional Results)

Although hydrothermal studies of frictional slip at temperatures greater than 180–240°C and displacement rates corresponding to typical plate motion rates, indicate stable, pressure-solution-like deformation within a fault zone (Blanpied et al., 1995; Chester, 1995), other solution mass transport phenomena such as overgrowths on load bearing contacts and fault zone sealing and compaction could operate at these conditions, and lead to rapid and large changes in fault zone material properties. Laboratory measurements of fault strength (Blanpied et al., 1992) and fault strengthening (Fredrich and Evans, 1992; Karner et al., 1997), and inferences of fault strength (S. Hickman, unpublished data) at hydrothermal conditions (>300°C) where solution transport phenomenon can be rapid, lead to two models where fault failure stress differs from the predictions of conventional rate and state friction. In the following sections we describe these two proposed mechanisms: (1) a solution mass transport compaction and fault sealing model and (2) a solution mass transport lithification model, and the experimental evidence that supports them.

5.3.1 Pressure Solution and Fault Sealing

If minerals are sufficiently soluble in water, then dissolution, transport, and precipitation can occur in response to gradients in chemical potential due to spatial variation of temperature, stress, or surface energy. Numerous studies have reported rapid reduction of permeability of rocks and joints (Moore et al., 1994), and granular aggregates (Tenthory et al., 1998) at such conditions. Additional studies have demonstrated rapid sealing due to flow along a temperature gradient (Summers et al., 1978; Moore et al., 1984; Vaughan et al., 1985). In faulting experiments at 600°C, Blanpied et al. (1992) inferred permeability reduction so rapid that in the presence of shear induced compaction, effective pressure increased, leading to a 50% reduction in shear strength. The observations of reduced strength (Fig. 20) suggest a potential earthquake mechanism for weak faults where, within a fault zone hydraulically sealed with respect to its surroundings, shearing leads to compaction and increased fluid pressure. Instability is initiated when the fault failure strength is reached by the combined action of tectonic loading and effective stress decrease (Blanpied et al., 1992), i.e., this failure mechanism differs from standard rate and state models in that porosity and \( \sigma_{\text{n,eff}} \) are functions of time and displacement.

A qualitative instability model based by the observations by Blanpied and co-workers (Blanpied et al., 1992; Sleep and Blanpied, 1992, 1994) is not unlike the Segall and Rice (1995) model, and accordingly, its potential for instability will be reduced in comparison to dry friction-based instability models through dilatancy and reduction in the critical stiffness. If dilatancy accompanies increases in sliding velocity (Morrow and Byerlee, 1989; Marone et al., 1990; Lockner and Byerlee, 1994) during failure, the accompanying effective stress increase stabilizes the fault (dilatancy hardening) (Rice, 1975; Rudnicki and Chen, 1988; Lockner and Byerlee, 1994). Decreases in effective stress brought on by compaction during

![Figure 20](image-url) Measured shear stress (upper graph solid, left axis scale), \( \tau/(\sigma_n - P) \) (upper graph dashed, right axis scale) and pore pressure (lower graph) versus axial displacement for Granite gouge at 600°C and 400 MPa confining pressure.
loading to failure reduce the critical stiffness, stabilizing the fault as in Eq. (17) or Eq. (19). Compaction-induced instability models are untested, as they require creep compaction behavior that has not yet been observed experimentally at the appropriate temperatures (200–400°C). Theoretical studies of creep compaction instability are ambiguous. Sleep and Blanpied (1994) report such instabilities using faults with uniform rate and state-like properties, as do Lockner and Byerlee (1995) using a Burridge–Knopoff-type model with inhomogeneous properties. Dilatancy hardening results in a delay but not an inhibition of dynamic failure for the Lockner and Byerlee model. On the other hand, strict interpretation of the existing experimental compaction data from rate-stepping tests, albeit at much lower temperatures (discussed in Section 5.2) does not lead to both weakness and instability (Segall and Rice, 1995).

5.3.2 Lithification

It has been suggested that at hydrothermal conditions, rates of fault healing may be greatly enhanced due to lithification processes that are not present in room temperature friction experiments (Kanamori and Allen, 1986; Scholz et al., 1986). Increases in static stress drop with recurrence for large and small repeating earthquakes may exceed the rate predicted by low-temperature friction (Kanamori and Allen, 1986; Scholz et al., 1986; Nadeau and McEvilly, 1999; Beeler et al., 2000), consistent with lithification. Accordingly, a fault-strengthening model based on this hypothesis would predict fault-strengthening rates that exceed those predicted by rate and state implementations of room temperature friction. Such rapid fault strengthening has been suggested to occur in experiments at elevated temperatures and hydrostatic pressure on quartz gouge at rates which exceed those observed at low temperature (Fredrich and Evans, 1992; Karner et al., 1997). However, to achieve the apparent enhanced rate of strengthening, fault gouge must be loaded to failure at a lower temperature (230–235°C) than the healing temperature (500–800°C). Experiments where the healing and loading temperature are the same [636°C (Karner et al., 1997)], as would be the case for faults in the Earth, show no time-dependent strengthening. The lack of time dependence indicates that these studies of fault strengthening at very high temperature may have limited relevance to natural earthquake failure. These studies are designed to measure healing due to local redistribution of material, i.e., dissolution from pore walls or fine or angular grains followed by deposition at grain contacts. More rapid or more significant cementation may occur due to mass transport of silica or calcite over large distances within a fracture network. Certainly quartz vein filling is a well-known example of this strength recovery process (Chapter 29 by Sibson).

Lithification processes other than pressure solution which may lead to fault strengthening can be induced at hydrothermal conditions where a reduction in surface energy, not a reduction in contact stress, drives solution mass transport processes (Hickman and Evans, 1992). Recent experiments on single contacts in quartz at $T = 350–530°C$, $P_r = P_{HLO} = 150$ MPa and non-zero contact normal stress indicate extremely rapid contact overgrowth (S. Hickman, unpublished data). Because of the controlled geometry and in situ monitoring, the mass flux as a function of time is known and the results can be extrapolated to other contact geometries. When extrapolated to a contact size typical for laboratory friction experiments, the contact area increase in these experiments greatly exceeds those of frictional strengthening (Fig. 21). The area increase is so great as to imply that fault zones subject to low effective stress may become entirely lithified during the interseismic period. In this case, the fault failure strength would rapidly approach the strength of an intact rock. However, the low contact stress in the experiments prevents easy extrapolation of these results to seismogenic conditions.

6. Dynamic Fault Strength

Although laboratory observations of frictional strength as represented by Eq. (19) have been applied to model the entire seismic cycle (e.g., Rice and Tsai, 1986; Boatwright and Cocco, 1996), this assumes shear heating is not significant during dynamic slip. Fault slip speeds during earthquake rupture are on the order of meters per second and for large earthquakes will last many seconds at hundreds of MPa normal stress (appropriate for seismogenic depths). Under these conditions significant shear heating should occur, resulting in dynamic weakening due to shear melting or pore fluid pressurization due to thermal expansion (Chapter 29 by Sibson). Consequently, conventional laboratory measurements of strength are unlikely to apply. Effects unrelated to shear heating but not considered in Eqs. (18) are also possible. Dynamic reduction of fault strength by normal stress reduction has been predicted on theoretical grounds (Weertman, 1980) and observed in numerical models (Ben-Zion and Andrews, 1998).

Many important unresolved issues in seismology are related to or determined by the strength during dynamic slip. If
dynamic weakening occurs, dynamic stress drops can be much larger than static drops (Brune, 1970, 1976), and rupture is more likely to propagate as a self-healing slip pulse (Heaton, 1990; Cochard and Madariaga, 1994; Perin et al., 1995; Beeler and Tullis, 1996) rather than as a conventional crack-like rupture (Zheng and Rice, 1998; Chapter 12 by Madariaga and Olsen). If the dynamic stress drop exceeds the static stress drop, larger ratios between the radiated energy and the seismic moment are expected for larger earthquakes than for small events (Abercrombie, 1995).

Stress heterogeneity in the fault zone also depends on dynamic fault strength. Geometrical effects (roughness of the fault surface) and nonuniform material properties lead to stress heterogeneity, but heterogeneity may also result during dynamic rupture propagation and arrest (e.g., Cochard and Madariaga, 1994). Such stress roughening during dynamic propagation is common in numerical models, particularly discrete representations of rupture such as Burridge–Knopoff models, and is associated with simulated earthquake populations which obey the power law Gutenberg–Richer relation between occurrence and event size. Aspects of the association of rough spatial stress distribution and Gutenberg–Richer event populations are summarized by Rice (1993) where it is argued that discrete models are not satisfactory representations of a continuum, and continuum models of dynamic rupture propagation using laboratory values for fault strength generally do not result in a rough residual stress and Gutenberg–Richer populations. The latter is true in particular if Eq. (18) are used (Rice, 1993; Rice and Ben-Zion, 1996). However, more extreme rate dependence of fault strength, such as would result from dynamic weakening (e.g., Zheng and Rice, 1998), does lead to stress roughening (N. Lapusta and J.R. Rice, unpublished data). Thus many open questions and unresolved debates might be resolved by appropriate laboratory determinations of dynamic fault strength.

### 6.1 Frictional Strength at High Sliding Rates

Measures of fault strength during dynamic rupture events (Okubo and Dieterich, 1986) which have average sliding velocities on the order 0.1 m sec\(^{-1}\) are consistent with frictional strength, i.e., the predictions of Eqs. (18) during stick-slip. However, the implied temperature change in those experiments as estimated from

\[
\Delta T = \frac{\tau \Delta \delta}{w \rho c}
\]

where \(w\) is the fault zone width, \(\rho\) is density and \(c\) is the specific heat (Lachenbruch, 1980) is not sufficient to induce weakening due to melting (there is no pore fluid) (Lockner and Okubo, 1983). This is primarily due to the small total slip \(\Delta \delta\) (80–100 \(\mu m\)) and low normal stress in the experiments. Measurements of fault strength at constant but lower sliding velocity (up to 3.2 mm sec\(^{-1}\)), dry, and at much higher cumulative displacement (500 mm), lead to strengths that are also generally consistent with Eq. (18) (Blanpied et al., 1998b) (see also Yoshioka, 1985; Mair and Marone, 1999). However, in the experiments of Blanpied et al. (1998b), sliding at the highest velocity and the highest normal stress (25 MPa induced anomalous weakening with displacement. This weakening did not occur at the lower normal stress (10 MPa).

### 6.2 Dynamic Weakening (New and Provisional Results)

Recently, dynamic weakening has been measured experimentally. The results, summarized in this section, fall into the two classes mentioned at the outset of this section: dynamic weakening due to reduction of normal stress and weakening due to shear heating. However, to date, none of the experimental measurements of weakening are definitive and some of the results summarized in this section are unconfirmed. In some cases the results or their relevance to natural faults are uncertain.

Dynamic weakening due to normal stress reduction was first observed in this context during stick-slip sliding between surfaces of foam rubber (Brune et al., 1993; Anoshehpoor and Brune, 1994). The observations consist of particle trajectories as the rupture passed and indicate opening displacement during the passage of a dislocation-like slip pulse (Fig. 22a). Because the frictional properties of foam rubber are quite unlike rocks [e.g., \(1.0 < \mu < 3.7\) (Brune and Anoshehpoor, 1997)], the experiments which are conducted under very low normal stress (0.9–7 \(\times\) 10\(^{-6}\) MPa) are difficult to scale to the Earth. Some indication of the extreme differences between stick slip of foam rubber and rock is that for foam rubber, stick slip only occurs at the lowest normal stress; increasing normal stress leads to stable sliding (Anoshehpoor and Brune, 1994), in contrast to the observation for friction of rocks and other brittle materials where increased normal stress promotes instability [e.g., Eq. (17) or (19) (Walsh, 1971; Goodman and Sundaram, 1978; Dieterich, 1979)]. These criticisms do not negate the applicability of the foam rubber observations to seismic failure, but they do emphasize fundamental differences between the behavior of foam rubber and rocks at laboratory conditions. The association of true opening displacements and stick slip for foam rubber suggest that this mechanism of instability is encouraged by low normal stress.

Dynamically reduced normal stress has been measured or inferred for materials other than foam rubber (Tosti, 1967; Bodin et al., 1998; Brown, 1998), suggesting that the foam rubber may be a valid laboratory rock analogue. Significantly, observations relevant to dynamic reduction of normal stress have been made for rock by Brown (1998). Heat generation during stick-slip between Westyler granite surfaces occurs at a lower rate than during stable sliding (Fig. 22b). Brown also found the mean stress during stick slip to be lower than during stable sliding. Assuming that the mean frictional strength is the same for both stick slip and stable sliding, the observations...
can be interpreted to be due to dynamic reduction in normal stress during stress drop.

Dynamic weakening has been reported due to shear heating in unconfined experiments at modest normal stress on gabbro (Tsutsumi and Shimamoto, 1997). Fault strength expressed as frictional resistance drops to the range $\mu = 0.4–0.5$ (Fig. 23) [$\sigma_n = 1.5$ MPa (Tsutsumi and Shimamoto, 1997)] at sliding velocities of $0.43–1.3$ m sec$^{-1}$ during melting and at sliding velocities of $100–550$ mm sec$^{-1}$ when melting is not observed.

The effect in the shear melting experiments is not as dramatic as one might expect, perhaps because of high melt viscosity (Tsutsumi and Shimamoto, 1997), but more likely because the melt is unconfined and is extruded by centrifugal force at these high sliding velocities (experiments are done in a rotary shear geometry). Unconfined experiments conducted by Spray (1987, 1988) on metabasite and amphibolite at $\sigma_n = 50$ MPa and mean sliding velocity of $0.24$ m sec$^{-1}$ show no evidence of weakening though accuracy of the stress determination is poor.

Preliminary results on confined and unconfined surfaces of quartzite and granite at slip rates of only $3.2$ mm sec$^{-1}$, normal stresses of $28–112$ MPa, and slip displacements of $0.2–10$ m, show that friction can dramatically decrease with slip. In some cases the friction drops below 0.3, or roughly 30% of the value typically found in quartz friction experiments (D. Goldsby and T. Tullis, unpublished data). Calculations of both average surface temperature and local temperatures at contacting asperities (i.e., so-called “flash temperatures”) predict maximum temperatures of $100–600$°C. In most cases, the dramatic frictional weakening occurs at estimated temperatures below $300$°C. Hence, the weakening is difficult to reconcile with melting, either of the entire fault surface or at asperities. It appears, however, that both temperature and displacement are important variables affecting the shear resistance.

High temperature triaxial quartzite experiments (restricted to a few millimeter displacements), at temperatures that span the range of frictionally generated temperatures in the Goldsby and Tullis experiments, yield typical values of the friction coefficient of $0.6–0.8$ (Stesky et al., 1974). Large displacement experiments (sheared to several meters) at velocities greater than $100$ μm sec$^{-1}$, which generate much less heat than the $3.2$ mm sec$^{-1}$ experiments over similar displacements, also result in typical frictional strength. A possible explanation is that the low friction coefficient results from shear heating of profoundly altered, likely amorphous, material along the fault.

**FIGURE 22** Dynamic weakening: normal stress effects. (a) Shear (S) and normal (N) displacements in foam rubber during a stick–slip event. (b) Heat generation during stick–slip versus stable sliding on unconfined, initially bare Westerly granite.

**FIGURE 23** Dynamic weakening due to shearing: strength plotted as a function of sliding velocity for unconfined gabbro at $1.5$ MPa normal.
7. Deep Earthquakes

Seismicity can extend well below the conventional brittle ductile transition (Fig. 10). This is particularly true in subduction zones where earthquakes occur to depths around 600 km (see Fig. 10c inset). Following Froelich (1989) we refer to intermediate focus events as those occurring in the depth range 70–300 km and deep focus events as those occurring at depths greater than 300 km. If conventional fracture and frictional instability were the only mechanisms of earthquake failure then seismicity would not exist below depths where the resolved shear stress on a fault is equivalent to 2000 MPa (e.g., Raleigh, 1967), the maximum flow stress above which all rocks and minerals deform by stable creep. For example, in a thrust faulting environment, assuming that $\sigma_1$ is fixed at overburden, pore pressure is hydrostatic and $\mu = 0.7$ (ignoring cohesion); failure of optimally oriented faults at this shear stress would occur at about 75 km depth. Obviously, the depth distribution of natural seismicity which indicates significant seismicity at 600 km cannot be interpreted in this simple manner. A number of mechanisms have been proposed to explain deep earthquakes. In this summary we focus on those that are supported by laboratory experiments. Not considered, for example, are models such as those based on a thermal instability (Griggs and Baker, 1969; Hobbs and Ord, 1988) which currently lack experimental support.

Laboratory observations indicate that earthquakes could accompany phase changes in the otherwise ductile portion of the crust. For example, brittle fracture occurs following reduction of the failure stress due to increased pore fluid pressure from a dehydration reaction. This dehydration embrittlement mechanism can explain intermediate events if hydrated minerals are present. Failure can also be triggered by heat from a phase transformation as well as differential stress around inclusions resulting from a phase transformation. Transformation-induced faulting provides a likely mechanism for deep focus events. Radiated stress and inelastic shear displacements result from other phase transformations such as melting (Griggs and Handin, 1960) and amorphization (Meade and Jeanloz, 1991).

7.1 Intermediate Focus Events

The explanation of intermediate focus earthquakes follows directly from the effective pressure law $\sigma_{\text{eff}} = \sigma_n - \rho$, Eq. (7), which was previously discussed in the context of Eq. (1a). Failure strength in excess of cohesion can be reduced to arbitrarily low levels if the pore pressure is high. At intermediate focal depths, where the high overburden stress and high temperature lead to time-dependent creep and recrystallization, porosity and pore connectivity are significantly reduced (Lockner and Evans, 1995). Under these conditions, water, provided by dehydration reactions, can exist at near lithostatic pressures and shear fracturing can occur at very low shear stress (Hubbert and Rubey, 1959). Experimental confirmation of this theory was made by Raleigh and Paterson (1965) who observed brittle failure of serpentinite at its dehydration temperature, but ductile behavior at lower temperatures (Fig. 24). Acoustic emissions were detected during dehydration at pressures up to 9 GPa (Meade and Jeanloz, 1991) confirming that this mechanism is capable of the rapid unstable deformation associated with earthquakes.

There are four requirements for dehydration embrittlement to be a viable model for intermediate events:

1. The brittle failure strength must be smaller than the ductile shear strength because the ductile shear strength would otherwise limit the stress available to drive failure.
2. The sliding frictional strength must be smaller than the failure strength so that a coseismic stress drop is possible. These first two requirements are met for serpentinite experiments (Raleigh and Paterson, 1965).
3. Faulting must occur under undrained conditions, since a loss of pore pressure would increase the shear resistance and stabilize the incipient rupture. At intermediate depths where porosity and permeability are extremely low, undrained conditions are expected to be the norm. In this case the stress drop during dehydration faulting is well described by a loss of cohesion.
4. Dilatancy must be suppressed or else the pore pressure will be reduced due to dilatancy hardening (Rice, 1975; Rudnicki and Chen, 1988).

FIGURE 24 Dehydration weakening. Strength of serpentinite deformed at a strain rate of $7 \times 10^{-4}$ sec$^{-1}$, and confining pressures of 100, 350, and 500 MPa as a function of temperature.
Recent experimental observations of failure of serpentine even at low confining pressure indicate that deformation is accommodated on shear microcracks that are aligned preferentially with the cleavage direction, as opposed to macroscopic shear fractures initiated from dilatant microcracks (Escartin et al., 1997). Thus little or no dilatancy can be expected during deformation of serpentine at high pressure. Application of the Hubbert and Rubey mechanism to explain intermediate focus events is limited to regions of the crust where dehydration is on-going and also by the constraint that the pore pressure not exceed the least principal stress. However, the conditions necessary for instability have not been rigorously determined for this mechanism.

### 7.2 Deep Focus Earthquakes

Failure of the crust in deep focus earthquakes occurs well below the depth where the frictional failure strength exceeds the flow strength of rock, and below temperature and pressure conditions of dehydration reactions. Deep focus events therefore require a mechanism that is different from shallow seismic failure in fundamental ways. Among the exotic mechanisms called upon to explain deep focus events is transformation induced faulting (Vaisnys and Pilbeam, 1976), where failure is associated with a phase transformation such as the pressure-driven olivine–spinel transformation. This mechanism is well documented in experiments, notably in water ice I to ice II where faulting is accompanied by release of radiated energy (Kirby et al., 1991), a key requirement for any explanation of deep earthquakes. Transformational faulting also occurs in Mg$_2$GeO$_4$ which has the same structure as Si olivine (Fig. 25) (Burnley et al., 1991) and, at temperatures, pressures, and mineralogy consistent with subduction zones, in (Mg,Fe)$_2$SiO$_4$ (Green et al., 1990). The key observations which distinguish this mechanism from conventional brittle failure are the presence of the transformed phase in the fault zone and in transformed lenses in fault parallel lineations, and the lack of precursory acoustic emissions which are associated with dilatancy during nucleation in conventional frictional failure. That the lenses of transformed material tend to be oriented with their long axis perpendicular to the stress field indicates that the transformation itself is controlled by the deviatoric stress. However, the exact mechanism and mechanics of transformational faulting are not well understood. Suggested mechanisms include the coalescence of transformed material into a shear zone (Kirby, 1987) or transformed material initiating a throughgoing shear fracture (Ringwood, 1972; McGarr, 1977). Similarly it is not entirely clear whether the weakens of the shear zone which leads to a stress drop in experiments is due to the fault being made of fine-grained transformed material with reduced shear (flow) strength or because of a loss of cohesion during the formation or propagation of the shear zone.

Advances have been made in theory to accompany and explain the laboratory observations. Lui (1997) has determined conditions necessary for formation of a transformed shear zone and the associated stress drop based on thermodynamic considerations and the kinetics of the phase transformation. However, the conditions necessary for rapid (seismic) failure have not been determined. Other phase transformations may lead to shear instabilities. Meade and Jeanloz (1991) report acoustic emissions and shear deformation associated with a crystal to glass transformation in serpentine between 6 and 25 GPa. Other hydrated minerals may also amorphize in this manner.

### Acknowledgments

We greatly appreciate the thoughtful reviews provided by J. Byerlee, C. Sammis, C. Scholz, R. Sibson, and W. Stuart. This work was supported in part by the National Earthquake Hazards Reduction Program.

### References


**Editor’s Note**
Due to space limitations, three useful items are placed on the Handbook CD-ROM under directory of \32Lockner: (1) symbols used in this chapter are given in LocknerTable1.pdf; (2) references with full citation are given in LocknerFullReferences.pdf; and (3) detailed figure captions are given in LocknerFullFigCaptions.pdf.