

From left to right: Terry Tullis, Roland Bürgmann, Onno Oncken, Allan Rubin, Paul Segall, Sergei Shapiro, Geoff King, Greg Hirth, Chris Wibberley, Massimo Cocco, Jim Rice, Kenshiro Otsuki



7

Group Report: Rheology of Fault Rocks and Their Surroundings

TERRY E. TULLIS, Rapporteur

ROLAND BÜRGMANN, MASSIMO COCCO, GREG HIRTH, GEOFFREY C. P. KING, ONNO ONCKEN, KENSHIRO OTSUKI, JAMES R. RICE, ALLAN RUBIN, PAUL SEGALL, SERGEI A. SHAPIRO, and CHRISTOPHER A. J. WIBBERLEY

OVERVIEW

At depth, all major faults undergo aseismic motion, and some faults slip aseismically at shallower levels. However, most slip on shallow faults occurs seismically, generating earthquakes. Due to the importance of earthquake-related hazards to society and because more remains to be understood about the mechanics of earthquakes than about stable motion, the focus of this report is on understanding the processes involved in seismic motions. In addition, we consider the process by which stress accrues on faults, as well as the interactions between seismic events and the aseismically deforming crust at depth.

We organize the discussion around the seismic cycle, the repeated occurrence of earthquakes, and its stages: prior to an earthquake, slip increases quasistatically during a nucleation phase. The coseismic stage, i.e, when the earthquake occurs and nearly all slip accumulates, coincides with dynamic rupture and is relatively short (seconds to minutes). It is followed by post- and interseismic phases, characterized by gradually decreasing slip and healing of the slip surface, until the next earthquake occurs. Coseismic slip increases stress below the maximum depth of rupture, accelerating aseismic deformation during the postseismic period and reloading seismogenic portions of the fault (Figures 6.1a, b, this volume). During much of the earthquake cycle, strain in the vicinity of active faults accumulates linearly in time, reflecting steady loading by tectonic plate motion. Typically, successive earthquakes differ from one another in their location, nucleation, rupture, and postseismic creep, so that although the concept of an earthquake cycle is useful, it should not be interpreted as a regular series of identical events.

Although fault motion during the vast majority of earthquakes is ultimately caused by plate motion, the details of how faults are loaded differ from one tectonic setting to another. We begin our discussion with considerations of this loading and the distribution of deformation on the largest scale before focusing on the several stages of the earthquake cycle itself. We close with a discussion of the significance of off-fault damage resulting from fault growth and complex geometry, and a section on the implications of our earlier discussions for the magnitude of stresses on faults. The principles and processes discussed below apply to the three main types of faults: strike-slip, thrust, and normal. However, due to the importance of subduction thrusts as the site of the world's largest earthquakes, and because their geometry and kinematics are more complicated than the strike-slip example fault illustrated in Figure 7.1, we include a special section devoted to subduction thrusts. Within each section we focus discussion on topics of importance for future study and offer ideas as to which approaches might be taken during this study.

A schematic representation of a continental plate boundary strike-slip fault, including a more detailed view of the seismogenic portion of the fault, is illustrated in Figure 7.1. This diagram summarizes much of what we know about faults and earthquakes and also serves as a basis for discussion of what still needs to be learned. In addition, time and length scales of the processes underlying faulting and earthquakes are shown in Figure 7.2.

Figure 7.1 Schematic diagrams of a generic strike-slip fault showing details of seismogenic > portion (a) and overall configuration (b). The state of current understanding and outstanding questions are presented. Although the example shown is a strike-slip fault, it highlights issues of fault zone structure, rupture initiation, and propagation that pertain to all fault types. A–B is the product of effective normal stress and $d\mu_{ss}/dt$, where μ_{ss} is the coefficient of steady-state friction; individually, A and B are the products of the effective normal stress and a and b, respectively, in the rate and state friction equation (see text). The lower depth limit of seismicity occurs where A-B becomes positive; peak strength occurs at greater depth. Between these depths, the frictional coefficient increases with sliding velocity, behavior termed velocity-strengthening. At shallow depths, it is still unclear if behavior is velocity-weakening or velocity-strengthening (A-B negative or positive). Earthquakes only nucleate where A-B is negative, but can rupture some distance into A-B positive regions. Nucleation may occur at special sites where gradually increasing static stress locally reaches static friction. During rupture propagation, rock strength τ_v at static friction may be overcome by stress that is dynamically concentrated ahead of the rupture tip in the presence of much lower tectonic stress τ_0 . Stress τ is probably low behind the tip due to low coseismic friction $\tau_{\rm f}$ but may increase slightly after slip ceases due to possible healing.

LARGE-SCALE VIEW OF THE EARTHQUAKE MACHINE

Figure 7.1 illustrates that slip on the seismogenic portion of the generic strikeslip fault is usually driven by plate motion. Although this general picture of faults has held for at least thirty years, several unresolved problems regarding





sixs pol s no (m) elsos lsitsq2

Figure 7.2 Plot of length scale versus timescale with boxes corresponding to intervals of operation for fault processes discussed in this report. The broad range of space-time intervals for these processes is part of the complexity of understanding faulting and earthquakes. Note that because fault zone dimensions parallel to the fault surface are much greater than those that are perpendicular to it, we plot the dimensions believed to be most relevant to the particular process in order to avoid covering the entire spatial scale. the interactions between the seismogenic zone and the aseismic portions of the plate boundary remain (see Chapters 2 and 6):

- How is displacement transferred through the lithosphere to the seismogenic zone?
- To what degree is deformation beneath the seismogenic zone localized?
- What constitutive relationships apply for deformation beneath the seismogenic zone?
- How does the fault zone beneath the seismogenic zone respond to stress perturbations associated with earthquakes?

Many of these issues are currently investigated by modeling inter- and postseismic deformation measured with geodetic techniques. In particular, following a number of large earthquake ruptures, stresses transferred to the lower crust and upper mantle can be thought of as rock mechanics experiments in a natural laboratory of lithospheric proportions. Models of postseismic deformation are sometimes nonunique; that is, more than one relaxation process (e.g., shallow and deep afterslip, poroelastic rebound, and viscous flow in lower crust and upper mantle) may contribute to the observed deformation. However, together with continued improvement in laboratory rock deformation data, and more sophisticated analyses of exhumed fault rocks from appropriate depths, these geodetic studies provide a new opportunity to constrain the rheology of the lower crust and upper mantle. Such information not only has important implications for understanding the earthquake process, it can also provide rheological constraints that are applicable to many other problems in geology and geodynamics.

A review of some recent postseismic deformation studies by Handy et al. (Chapter 6) shows that, depending on fault type, earthquake magnitude, and the composition and environment of the surrounding lithosphere, different processes and rheological parameters fit the deformation data. However, these studies also show that different authors come to different conclusions for the same earthquake. This may, in part, reflect the variety of tectonic settings, which includes as examples the 1992 Landers and 1999 Hector Mine earthquakes in California, two strike-slip earthquakes in Iceland, the 1999 Chi-Chi earthquake in Taiwan, and the 2003 Tokachi-Oki earthquake in Japan. Depending on the method of observation and the earthquake studied, the observed postseismic deformation has been explained by afterslip below the rupture, distributed lower crustal and/or upper mantle flow, and poroelastic readjustments. Rheologies inferred for the lower crust and upper mantle include linear viscous, biviscous rheology, and power law creep. The strain associated with the postseismic deformation is very small, on the order of 0.02 to 0.2%, depending on the width of the deformation zone. Thus, depending on the ratio of the stress perturbation to the long-term stress state, the nonlinear response may largely reflect transient creep processes rather than steady-state deformation.

To clarify the issues raised by the current alternate interpretations, more work is needed to incorporate the the spatial and temporal information contained in all existing and future deformation data. Careful model testing and comparisons should carefully consider all candidate mechanisms. Even with recent improvements in the geodetic community's response following large earthquakes, we are still limited by the lack of sufficiently precise horizontal and vertical deformation data and would benefit from better and more complete spatial and temporal sampling of the relaxation signal. The small strain associated with postseismic deformation illustrates the need for further experimental and theoretical work on transient creep processes.

SEISMIC CYCLE: NUCLEATION LEADING TO UNSTABLE SLIP

Fault nucleation requires further study, not only because it is associated with precursors that could someday provide a means for short-term (weeks to minutes) prediction, but because it is fundamental to the onset of dynamic rupture and coseismic slip (Dieterich and Kilgore 1996; Tullis 1996). Nucleation occurs on a restricted portion of the fault, whereas dynamic rupture typically propagates over a much larger area of the fault (Figure 7.1). The time period of nucleation and the dimensions of the nucleation area are not well known and may be quite variable. Our current knowledge comes from seismic, strainmeter (Johnston and Linde 2002), and theoretical studies. These are consistent with laboratory-derived friction laws discussed below.

Rate and State Friction

Laboratory studies have shown that the frictional resistance of rocks involves two main dependences: a nonlinearly viscous dependence on slip velocity, in which an increase in stress causes an increase in slip rate, and a state dependence in which the state of the fault evolves with both time and displacement (Dieterich 1979; Dieterich and Kilgore 1996; Marone 1998; Ruina 1983; Tullis 1996). The slip velocities over which these experiments have been conducted span the range of slip rates that are expected to occur during earthquake nucleation. Theoretical studies of nucleation using empirically derived rate and state friction laws show interesting nucleation behavior and represent the basis for our best expectation of what may occur during nucleation (Dieterich 1986, 1992; Dieterich and Kilgore 1996). The rate and state friction laws are based on fitting experimental data to empirical functions over the range of the laboratory experimental variables and are not based on theoretical descriptions of the operative processes, although some progress has been made in this direction (Chapter 5). Neither of the two evolution laws in common use, the "slowness" or "aging" law and the "slip" law, fit the experimental data over the entire range of slip velocities from which they were derived (Kato and Tullis 2001; Marone 1998). The two laws make quite different predictions for behavior during earthquake nucleation. In particular, the aging law leads to much larger nucleation zones, which are more likely to be seismically detectable, than does the slip law.

Thus, constitutive laws are needed that fit laboratory data better and that are based on identifiable processes in fault rocks. Although rate and state friction is reasonably well understood as a thermally activated process (Chapter 5), we need to determine the processes responsible for this behavior. This would create a firmer basis for extrapolating the laws from experimental conditions to conditions in nature that are relevant to earthquake nucleation and propagation. There may be other processes operating in nature that are not well studied by laboratory experiments and which may produce behavior similar to that of the present rate and state laws, but with different evolution laws and different values of the controlling parameters. Consequently, theoretical studies of mechanisms that produce state evolution on faults should be undertaken to determine the form of their evolution laws. Such processes may include pressure solution, phase equilibrium-induced reactions, and even biologically enhanced mineral precipitation. The effect of pore-fluid-related processes on nucleation needs to be investigated. Although it appears that thermal pressurization (discussed below) alone cannot lead to earthquake nucleation, it may play an important role in accelerating slip at slip speeds exceeding 1 mm s⁻¹ (Segall and Rice 2004).

Direct Studies of Nucleation

Direct measurements of behavior during earthquake nucleation should be undertaken. This can be done in a range of settings: from large laboratory experiments on rock, to small earthquakes induced by mining or by fluid injection in geothermic experiments, to standard earthquakes of a range of sizes. It is critical to place measurement devices, for example, seismometers and strainmeters, as close as possible to the nucleation site of events whose occurrence we can anticipate, ideally at distances comparable to the dimensions of the nucleation area itself. Although this spatial requirement is difficult to fulfill, borehole strainmeter installations are probably best suited for locating sufficiently close to be able to detect signals over the frequency band in which aseismic nucleation signals are expected. Studying the repeating earthquakes where the SAFOD experiment will hopefully penetrate in the summer of 2007 represents one possible opportunity, but better opportunities may exist in the case of mining-induced earthquakes.

The seismic nucleation stage may yield information about the aseismic nucleation process. One strategy for making progress in this area is to develop the capability for imaging earthquake slip using the waves generated during the beginning of seismic slip. Slip that occurs without much rupture propagation is indicative of nucleation over an area, rather than at a point. Therefore, we should try to identify microearthquakes that have stopped soon after nucleation, because much of the character of such earthquakes will be that of the nucleation stage. If such earthquakes can be identified, they may allow us to constrain the dimensions of the nucleation area. Knowledge of the nucleation area will help us evaluate the applicability of predictions from rate and state friction to understanding the nucleation process. Such microearthquakes might be identified by small ratios of radiated energy to seismic moment. One possibility is to seek microearthquakes that delimit the boundaries between areas on faults that creep and areas that slip in large earthquakes, such as those boundaries under Middle Mountain on the Parkfield segment of the San Andreas fault, or those located near the base of the seismogenic zone. The material properties in these areas are presumably marginal for unstable slip, in which case the nucleation zones will tend to be larger and thus more likely to produce detectable signals. A subset of the aftershocks of the Morgan Hill earthquake on the Calaveras fault, California, defines a truncated Omori-Law decay sequence, suggesting that they are just able to slip seismically. This renders them good candidates for study.

Another strategy would be to study foreshock sequences. There is evidence that foreshocks may involve dynamic rather than static triggering of the subsequent mainshock (Dodge et al. 1996). Alternatively, both the foreshocks and the mainshock are manifestations of an aseismic nucleation process.

SEISMIC CYCLE: DYNAMIC RUPTURE

Dynamic Weakening

Once the earthquake rupture propagates at speeds approaching the shear wave velocity or faster, slip velocities jump to rates of a few m s⁻¹ behind the rupture tip, and continued slip occurs at about 1 m s^{-1} . It is likely that slip resistance drops to values much lower than the stress needed to initiate slip (Figure 7.1), but much of what occurs on the rupture surface is presently unknown. The frictional parameters measured in conventional friction experiments at slip rates less than 1 mm s⁻¹ are probably no longer applicable because many other processes can also operate at high slip rates. These processes induce reductions in shear resistance to values much lower than those obtained from a typical frictional coefficient of ~ 0.6 . A detailed constitutive description for many of these processes is not yet known, nor do we understand the relative importance of the potential mechanisms during earthquakes, despite some recent progress (see review by Rice and Cocco, Chapter 5). We focus our suggestions for future work on the dynamic weakening mechanisms of flash heating, wholesale frictional melting, gel formation, and thermal pressurization of pore fluids (Figure 7.2) because we feel that these processes (see Chapter 5, section: DYNAMIC WEAK-ENING PROCESSES (THERMAL, FLUID) DURING SEISMIC SLIP) are potentially important for better understanding the coseismic stage of the earthquake cycle.

Weakening due to flash heating appears to occur after a few mm of slip at slip speeds above about 100 mm s⁻¹ in all rocks tested (quartzite, granite, gabbro) for which melting occurs at room pressure (Goldsby and Tullis 2003). It does not occur for calcite, which breaks down to CaO and CO₂. Due to the importance of clay minerals in fault zones, it is important to determine if weakening due to flash heating occurs for clay. The breakdown of clay minerals via dehydration may prevent weakening, although the very short times involved in flash heating (Figure 7.2) may not allow dehydration to occur.

With continued slip, the transition from local melting via flash heating to wholesale melting along the fault surface is a complex process involving a temporary increase in shear resistance (Hirose and Shimamoto 2004; Tsutsumi and Shimamoto 1997). Complicated negative feedback can occur with melting, as high heat generation leading to melting occurs when the frictional resistance is high; however, production of enough melt can reduce the resistance to the point that it freezes due to lower heat generation. Consequently, interesting cyclic behavior is possible where sufficient slip occurs. Unless other weakening processes reduce the shear stress to low values, melting is expected to occur on faults with any significant displacement below depths of about 5 km. A better understanding of the shear melting process may be obtained from experimental, theoretical, and field studies; pseudotachylite in natural fault rocks shows the importance of this process. It is commonly believed that melting may not be widespread on faults because pseudotachylite is quite rare in nature. Yet, pseudotachylite may be more common than its apparently rare occurrence would lead one to believe (Figure 7.1), especially given that later faulting may overprint it, or that slow cooling near the base of the seismogenic zone favor devitrification and/or static recrystallization.

Frictional sliding on smooth surfaces of a rock with silica contents of 70% or greater facilitates the generation of a layer of silica gel that results in large reductions in shear resistance (Di Toro et al. 2004; Goldsby and Tullis 2002). This occurs at lower speeds, 1 to 100 mm s⁻¹, than required for weakening caused by flash heating, but it requires displacements of about 1 m to attain steady state. The shear resistance can be extrapolated to zero at seismic slip speeds of 1 m s⁻¹. However, it is not known whether this weakening would occur at speeds above about 100 mm s⁻¹ because the weakening caused by flash heating and thermal pressurization may inhibit or preclude the gel weakening. Large displacement experiments at slip speeds above 100 mm s⁻¹ should be conducted to ascertain whether gel weakening operates in this velocity range. Related experiments at elevated ambient temperature would allow us to determine whether gel production is inhibited when the water required for the gel is driven off.

Weakening due to thermal pressurization of a pore fluid requires that slip be sufficiently localized for temperatures to rise rapidly, and that permeability be low enough that fluid pressure cannot decay rapidly. Thus, it is important to determine whether the narrow zones of slip found for fault zones exhumed from about 3 km depth (see Figure 7.1a and Chapter 5, Figure 5.1, and references

therein) are characteristic of the fault structure at that depth and whether faults at deeper levels have the same degree of localization (Figure 7.1; see also Chapter 2). Finding suitable exposures in exhumed faults where overprinting from shallower brittle deformation did not occur is a formidable task and systematic searches for suitable large-displacement faults with structures preserved from the 5 to 10 km depth interval need to be undertaken (see Chapter 6, Figures 6.5 and 6.6 and Chapter 4, Figure 4.2,). It is also important to collect data on the permeability structure of the damaged rock immediately surrounding the narrow slip surfaces in such exhumed faults (Wibberley and Shimamoto 2003). The thermal pressurization mechanism also requires that fault zone dilatancy does not increase with progressive slip by as much as the increase in volume of the pore fluid due to its thermal expansion (Junger and Tullis 2003). Thus, it is important to evaluate the role of dilatancy in counteracting the thermal pressurization effect. Experiments designed to test the thermal pressurization mechanism should be undertaken, but they should be conducted in the absence of weakening by other mechanisms (Goldsby and Tullis 1998; Roig Silva et al. 2004) to ensure that sufficient heat is generated. Furthermore, the poroelastic and inelastic response to transient increases and decreases of stress during dynamic rupture needs to be included when considering pore-pressure changes.

Geological and seismological field studies are necessary to evaluate these various weakening mechanisms during earthquakes. Searches for thermal signatures immediately adjacent to localized slip surfaces in exhumed faults should be undertaken. Thermobarometric and geochronological studies of samples from exhumed fossil fault zones that were active at depths corresponding to large earthquakes may help to constrain the maximum heat released during earthquakes (d'Alessio et al. 2003). Models of heat generated during dynamic slip suggest that sampling would have to focus on a zone of only a few mm in width, rather than the 2–5 cm sampled by d'Alessio et al. (2003). The core of exhumed fault zones should be also examined for microstructural and physiochemical signatures that may be diagnostic of dynamic weakening mechanisms. Dating of such rocks would require applying geochronometers sensitive to short heat pulses (e.g., electron spin resonance, fission track dating, and U/Th-He methods).

Better seismological ways to determine fracture energy and slip weakening distances for earthquakes are needed to compare with model predictions from several of the high-speed weakening mechanisms discussed above. The current methods are subject to more uncertainty than is desirable for comparison with predictions.

Slip Distributions and Propagation during Dynamic Rupture

Kinematic inversions of slip during dynamic rupture yield very heterogeneous patterns, but their details may be incorrect. The character and significance of these heterogeneities should be further explored by conducting kinematic inversions that are dynamically consistent, in contrast to standard kinematic inversions. Such models could better define what combination of stress, constitutive properties, and fault geometry gives rise to an observed heterogeneity. In addition, we must understand the relative roles of geometry and stress heterogeneity in causing heterogeneity of slip, as well as in causing where ruptures initiate and terminate. It is also important to understand how any slip deficits are accommodated, whether by microseismicity, fault creep, or complementary behavior in successive earthquakes. Continued seismic and geodetic monitoring of strike-slip faults is an important way to do this, in particular at Parkfield, California, an area with many useful existing data. Evaluating the effect of lithology on the behavior of different parts of faults may be approached by analyzing the head waves traveling along faults (McGuire and Ben-Zion 2005).

One approach to understanding earthquake mechanics is to image the earthquake source using seismic waves. Currently, there is a large deficit in observational studies of large earthquakes. Excluding sites in Japan, we do not have instruments that can record large earthquakes on scale in relatively low-noise (and simple in terms of wave propagation effects) borehole environments. The recent Parkfield earthquake (September 28, 2004) was recorded by dozens of strong motion instruments; however, the data are very difficult to interpret for several reasons:

- 1. The instruments are old, and hence the analog signals must be digitized, which results in a loss of signal-to-noise ratio.
- 2. The instruments are at Earth's surface and not in solid rock, and hence nonlinear attenuation renders the records difficult to interpret. This also means that high-frequency ground motion, which might allow one to constrain the slip-weakening displacement, will be lost to seismic attenuation.
- 3. The instruments cannot record smaller earthquakes. Digital recordings of smaller earthquakes would allow us to calibrate wave propagation effects to estimate how much nonlinear attenuation might occur, and hence to isolate source effects in the data.

Neither EarthScope nor ANSS (the Advanced National Seismic System in the U.S.) will address this observational gap adequately: EarthScope because there is no strong motion component to that project, ANSS because it is focused on instrumenting urban environments rather than faults.

One possibility is to set up an array of monitoring instruments in boreholes along large faults. These might monitor ground motion, strain, and pore pressure, yet be designed to stay on scale during earthquake ruptures. This would amount to monitoring a large biaxial laboratory experiment, but conducted on the scale of Earth's crust in the field. The Plate Boundary Observatory component of EarthScope is doing this, but not with downhole seismometers that would stay on scale during strong shaking.

FAULT ZONE RESTRENGTHENING DURING CO-, POST-, AND INTER-SEISMIC PERIODS

Immediately following an earthquake (postseismic stage) and until the time of the next one (interseismic stage), the rock along and adjacent to the fault heals. The rate of healing likely occurs in a manner that is proportional to the log of the time since the earthquake. Studies, however, are needed to determine the kinds of healing processes that operate and the rate at which strength recovery actually occurs (Chapter 12). This involves learning about the same processes of strength evolution discussed earlier. Because interseismic times are typically long compared to times available for laboratory experiments, it is important to know which processes operate and their corresponding healing laws in order to extrapolate beyond the times of lab data (Figure 7.2). Geological evidence for healing in exhumed faults is needed to identify and understand the processes involved. Geochemical studies of pore-fluid compositions sampled in situ and over a period of time in borehole experiments might help constrain the processes and their rates (see Chapter 14). Seismological studies of changes in wave velocities as a function of time after earthquakes, together with magnetotelluric studies to estimate changes of fluid content, are desirable.

Concepts

Healing rates may be greatest immediately after the dynamic weakening, as slip velocities decrease during the coseismic rupture propagation. Rapid healing is required in a homogeneous fault to explain the finite duration of slip (see Figure 5.6, this volume). In the recent literature, several modifications to the constitutive relations governing fault friction have been proposed to model healing and to simulate dynamic traction evolution; they show fast and immediate restrengthening at the end of the dynamic weakening phase (Figure 7.1). Stress (and/or strength) heterogeneities might explain the finite slip (and slip velocity) durations without requiring such rapid restrengthening (Figure 7.1). Whether finite slip durations are due to heterogeneities or to high-velocity constitutive behavior is an unresolved issue requiring further investigation.

More generally, co-, post-, and interseismic restrengthening is a particular example of the time dependence of strength. In the well-known expression for strength, each term may be thought of as being time dependent:

$$\tau(t) = \mu(t) [\sigma_{\rm n}(t) - P_{\rm f}(t)] .$$
(7.1)

Rate- and state-dependent friction have been proposed to explain the time dependency of the frictional coefficient $\mu(t)$ under constant normal stress and at low slip rates:

$$\mu(t) = \mu_0 + a \ln \left(V/V_0 \right) + b \ln \left(\theta V_0 / D_c \right) , \qquad (7.2)$$

where *V* is slip velocity, μ_0 is the frictional coefficient at steady state at slip velocity V_0 , D_c is the e-folding displacement for evolution of friction with slip following a step in slip velocity, and θ is a state variable with units of time. Clearly it is not correct to describe friction as being solely time dependent, as might appear to be the case from writing $\mu(t)$. In fact, time dependency arises both from the dependence on slip velocity *V*, and the evolution of the state variable θ with time and slip according to several possible evolution laws (Marone 1998; Ruina 1983; Tullis 1996).

In this equation, the state variable θ represents the fault surface memory of previous slip episodes. For the short times and high slip velocities relevant to coseismic and postseismic restrengthening, the frictional coefficient may depend even more strongly on time and slip velocity than is the case for typical rate and state friction. In addition, both pore-pressure evolution and normal stress variations can contribute to this time dependency of fault strength over a wide range of time scales.

Observations

Evidence of fault restrengthening comes from several observations. Laboratory experiments illustrate the importance of restrengthening both during coseismic and interseismic periods (Dieterich 1979; Marone 1998; Tullis 1996; Chapter 12). Seismological evidence for healing is provided by duplets or multiple events at the same locations. It is sometimes observed that the stress drop or the magnitude of these repeating earthquakes depends upon the recurrence interval between them. Finally, analysis of fault zone-trapped waves allows imaging of the time variations of elastic properties of fault zones (Chapter 2). Many more of these types of observations are needed.

Processes

Several physical and chemical processes play a role in evolving fault zone strength (Figure 7.2). Testing the relative importance of each of these requires estimating the relative rates and amounts by which they operate and, in turn, necessitates an in-depth, theoretical understanding of the mechanisms. It is difficult to separate the physical from the chemical processes, because they both rely on the presence and behavior of a fluid phase—the physicochemical role of the fluid is of crucial importance. The evolution during the seismic cycle of porosity and permeability of the fault rock in both the fault core and the damage zone are fundamental in controlling the relative rates of these processes. Laboratory friction experiments in the presence and absence of chemically inert and active fluids have started to define the relative time evolutions for restrengthening but need to be complemented by field observations and laboratory porosity—permeability measurements of fault rocks to evaluate the

potential importance of these different processes. The restrengthening shown by the increase of static friction with log of stationary contact time is one manifestation of the evolution of state in the formulation of rate and state friction. As discussed below, several physical and chemical processes offer plausible explanations for this evolution effect; experiments and theoretical modeling need to be undertaken to determine how many of them may play a role.

Physical Processes

One possible explanation for the experimentally observed increase of static friction with log of stationary contact time is that the real area of contacting asperities increases with time. The contact area could grow due to time-dependent creep of the contacts by a number of possible physical processes. Such growth might occur by thermally activated, and therefore time-dependent, dislocation motion. Any dislocation motion might occur either as dislocation creep, in which the strain rate is expected to be proportional to the local differential stress to a power of approximately three, or as dislocation glide in which the strain rate is expected to show an exponential dependence on the local differential stress. Two lines of evidence suggest that, if the evolution effect is due to dislocation-assisted growth of contact area, the process of dislocation glide is the more likely explanation. First, the increase of frictional strength depends on the log of the slip velocity, a functional form that is consistent with dislocation glide, not dislocation creep. Detailed modeling of the growth of contact area should be undertaken to verify this. Second, the stress magnitudes estimated to exist at contacting asperities approach the theoretical strength and, if dislocations are involved at all at such high stress levels, dislocation glide rather than dislocation creep is expected. In fact, the high estimated stress levels suggest that increases in contact area could occur by intracrystalline slip without the propagation of dislocations, that is, deformation at the theoretical strength. Increases in frictional contact area could also occur by small-scale time-dependent microfracturing. The process of subcritical crack growth, typically occurring by chemically assisted stress corrosion (see discussion on chemical processes in Chapter 14), could allow this to occur. It is important to make direct observations of the deformation processes at contacting frictional asperities to determine what role, if any, dislocations and cracking play in increasing asperity contact areas.

Fault strength is partly controlled by the evolution of fluid pressure. Coseismic overpressuring of the slip zone by thermal pressurization may be dissipated by outward fluid flow or by thermal contraction of the heated fluid due to conduction of heat away from the slip zone after frictional heating subsides. These mechanisms of fluid pressure reduction lead to a relatively rapid recovery of postseismic strength. On the other hand, fluid pressure increases may be generated by compaction creep and porosity reduction in parts of fault zones with a low permeability. Generation of an electrochemical potential gradient across the fault zone may also generate an equally large increase in fluid pressure. This could be stimulated by rapid, coseismic introduction of a more saline fluid along the fault from below, or by chemical exchange between the fluid and fresh, intragranular fracture surfaces in the fault rock, possibly with periodic (coseismic?) flushing of the surrounding fracture-damage-zone reservoir by fresh low-salinity fluids. The reader is referred to Chapter 14, Theme 2, for more discussion of these issues.

Chemical Processes

Chemical processes also possibly contribute to interseismic strength evolution, as discussed in detail by Gratier and Gueydan (Chapter 12). Chemical reactions at contacting surfaces assist in the restrengthening seen in the frictional evolution effect, both in situations involving the contact of bare fault surfaces and grain-grain contacts within gouge layers. These processes include pressure-solution compaction hardening of a gouge zone, healing of the fault by cementation, and increase in the asperity contact surface area by diffusive mass transfer and/or grain surface diffusion. Chemical reactions may strengthen contacts, even if no increase in contact area is involved. For example, gradual removal of contaminant water layers separating crystalline phases at a frictional contact can allow stronger bonding between the crystals to develop. Time-dependent polymerization of water-rich gel layers at frictional interfaces is a related way in which such contacts can restrengthen after sliding has ceased. As mentioned above, the relative importance of all of these processes can be established by determining the rate-limiting factor(s) for each process.

Although these processes have been proposed, in some cases with field, microstructural, and borehole evidence, we do not have a unified physical model of fluid effects in which to incorporate the constitutive properties of faults. As discussed by Yardley and Baumgartner (Chapter 11), faulting can affect fluid flow and availability just as fluid presence and pressure can affect faulting. Some of the important factors to consider are the open versus closed nature of the fluid supply and the ultimate source of the fluids involved (Chapter 11). More studies of the fluid from borehole samplings and from fluid inclusions in healed fault rocks in exhumed faults are needed, with attention paid to both the chemistry and the fluid pressure. In particular it would be useful to monitor fluid pressure evolution during the seismic cycle in the fracture damage network around the fault core to assess both the predictive capability of models of fluid pressure evolution during earthquakes and the relative rates of proposed physical processes operating in the fault zone in the interseismic period. In addition, it would be useful to monitor changes through time in the pore-water chemistry in the low-permeability fault core and surrounding fracture zone, coupled with geophysical estimates of porosity changes. This would allow estimation

of the rates of dissolution and precipitation at different times during the interseismic period and the importance of electrochemical potential gradients for generating overpressures in the fault zone.

The Nojima Fault Probe drilling project provides an opportunity to conduct some of these studies. This project is unique in exploring the stress state and physical properties of the faulted crust soon after a major earthquake (M7.2Kobe 1995 earthquake). Borehole injection tests at 1.5-1.8 km depth around the fault zone only two years after the earthquake were followed by migration of fluid pressure-induced microseismicity parallel to the fault at 2-4 km depth. Permeability estimates of 10^{-15} to 10^{-14} m² obtained from velocities of the migrating microseismicity front generally correlate with the highest laboratory core-scale permeability measurements of very heterogeneous fractured damage zones around fault cores. Sealing of the poroelastic fracture damage zone network through the post- and interseismic stages is an important process in controlling fluid communication and pore-pressure distribution. Similar experiments undertaken through time may therefore provide us with valuable information on the rates of sealing. Also of possible help in estimating porosity evolution could be data from remote geophysical experiments, such as Vs/Vpand $Vs \times Vp$, and magnetotelluric studies.

FAULT GEOMETRY AND THE SIGNIFICANCE OF DISTRIBUTED DAMAGE

Faults idealized as in Figure 7.1 accommodate predominantly simple shear, but the boundary conditions applied in the Earth usually deviate from simple shear, being some form of general noncoaxial shear. Hence, no single fault or fault orientation is sufficient to accommodate bulk deformation compatibly. Faults with several orientations must coexist and be simultaneously active. In general, these must meet at bends or junctions. Unlike the triple junctions of oceanic plates, no lithospheric sources (ridges) or sinks (subduction zones) are present, and as a consequence these junctions are unstable. Owing to this, multiple fracturing must occur where faults change strike or faults of different orientations meet (King 1983). The result is the creation of regions of continuously evolving zones of fragmentation or damage (King and Yielding 1984; King 1986; King and Nabelek 1985). The existence of such regions has implications for the mechanics of earthquakes since narrow zones may not adequately represent faults. There is some evidence that earthquakes initiate and terminate in these complex regions. Systematic studies of such regions of deformation should be undertaken and may reveal features important to earthquake mechanics.

Even where boundary conditions are simple, large damage zones appear at the ends of individual faults. Like fault junctions, these zones are associated with multiple fracture directions (Scholz 2002). Information about these regions can be deduced from studies of the cumulative displacement distributions on geological faults (Manighetti et al. 2004, and references therein), which conclude that these process zones are much larger than encountered in engineering materials. Like fragmented regions around fault junctions, these zones evolve as a fault evolves, and because each fault is the product of numerous individual earthquakes, they necessarily play an important role in the earthquake cycle. Geological and seismic studies should examine how slip during individual events combines to create observed geological slip distributions. The damage zones around such evolving faults should also be studied in detail. Together, such data will provide information pertinent to the role of such structures in the dynamics of individual earthquakes.

SUBDUCTION THRUSTS

Convergent continental margins are the Earth's principal locus of important earthquake hazards. Some 90% of global seismicity and nearly all interplate megathrust earthquakes with magnitudes >8 occur in the seismogenic coupling zone between converging plates. Despite the societal, economic, and scientific importance associated with the interplate coupling zone, the processes that shape it and its mechanics are poorly understood. Drilling has not accessed seismogenic plate interfaces, nor are they exposed at the surface onshore.

Seismogenic coupling zones occupy a depth interval of convergent plate interfaces that range from 5 to 10 km depth at the up-dip end and 30 to 60 km at the down-dip end (e.g., Oleskevich et al. 1999, and references therein). The down-dip end tends to align with the coastline in the Circum-Pacific belt of convergent margins. The reason for this loose correlation is poorly understood, although it has significant implications for the potential of large seismic events to generate tsunamis as well as severe ground shaking, depending on the slip distribution on the rupture plane and its relation to the coastline. Features along the seismogenic coupling zone that affect the distribution of slip and ground shaking at the surface include large asperities on the down-going plate (seamounts, transform faults, or oceanic ridges) and lateral variations in fluid pressure. Both of these features affect the areal extent and the degree of seismic coupling and therefore play major roles in the generation of great interplate earthquakes. Because elastic strain prior to earthquake rupture does not accumulate uniformly in time and space, geodetic observations provide a valuable means of deducing the degree of coupling along the seismogenic zone. Questions about material behavior along the coupling surface can be addressed via laboratory experiments. Key questions include: How are rock strength and seismogenic coupling distributed at depth during the various stages of the seismic cycle? What are the constitutive properties of fault rocks along aseismic and seismic fault segments? In particular, what mechanisms are responsible for the upper and lower transitions from seismic to aseismic behavior? And how are all of these factors related to surface hazard? A more reliable assessment of hazard distribution is only possible if we understand why earthquakes with a similar rupture history and magnitude recur along the same fault segment.

The relatively shallow up-dip ends of seismogenic coupling zones are accessible to drilling and, indirectly, to sampling of rock and volatiles from faults that splay upward from the plate interface. Furthermore, the offshore location of these zones allows geophysical imaging at reasonably high resolution. On the other hand, this portion of the subduction interface is difficult to study geodetically. The down-dip transition zone is much less accessible to geophysical imaging and monitoring. Given the intriguing observations of deformation transients and tremor activity, as well as the first hints of temporal changes in geophysically imaged properties, intensified monitoring and imaging of subduction zones should be a high priority for future studies. This should be combined with detailed structural and thermobarometric studies of fault rocks from fossil subduction zones, including eclogitic rocks containing high-pressure and ultrahigh-pressure metamorphic mineral phases. Only then can rock physical properties derived from geophysical studies be related to processes preserved *in situ* in disequilibrium structures of natural fault rocks.

The concomitant occurrence in several subduction zones of seismic tremors and transient slip events which seem to be located along the down-dip extension of the seismogenic zone (e.g., Rogers and Dragert 2003) suggests that they represent aseismic loading of the seismogenic zone. Alternatively, and less likely, they may be part of a long-term nucleation event. As precursors, they could be of value in predicting large underthrust earthquakes, as suggested by a $M_{\rm w} = 6.7$, up-dip thrust event that followed one of the slow slip events observed at Guerrero, Mexico, on April 18, 2002 (Liu and Rice 2005). Understanding the origin of these transient slip events and the tremors sometimes associated with them could help us learn more about the nucleation process. The origin of the tremors could be constrained by looking carefully at the relative onset times of tremor occurrence and the slip transient. Does slip occur first and allow the tremors to occur? Or do the tremors alter the stress state on the plate interface and allow the slip to occur? Dense sampling of the seismic wavefield with arrays of closely spaced instruments would allow us to constrain the depth and propagation of tremors by tracking the phase velocity of locally coherent tremor waveforms across the array. The station spacings within current earthquake monitoring arrays are too broad to allow this. Using the waveforms to determine whether motions in the source region of the tremors involve shearing or dilatancy could help us to understand their origin. Models of how tremors might form should be developed and compared with observations. Magnetotelluric studies might help determine whether the motion of fluids causes tremors. Understanding the evolution of the fluid budget, especially the relative rates of fluid production, percolation, and sealing in the vicinity of faults during the seismic cycle, is key to quantifying the short-term mechanical evolution of the plate interface (Chapter 12). Here again, examination of exhumed parts of fossil seismogenic plate interfaces is required to

understand the interaction between fluid flux, seismic failure, and aseismic slip (Chapter 6). In addition, high sensitivity monitoring to see if tremors exist in a variety of tectonic environments is worthwhile to ensure they represent a fault-related signal, since they have been found by careful re-examination of records in several places where they were not previously recognized.

STRESS ON FAULTS

The discussion above makes clear that although the strength of most fault rocks which are not slipping, or are slipping at rates less than 1 mm s⁻¹, is approximated by Byerlee's Law, faults that slip coseismically may have much lower shear strength, perhaps near zero strength (Figure 7.1, see also Chapter 6, Figure 6.1). This leads to a conceptual model (described below) in which the slip-integrated strength of fault zones in the seismogenic coupling zone is very low compared to their high time-integrated strength and the strength of the surrounding unfaulted crust. Therefore, there may be no heat flow paradox for large faults and no problem with the idea that large plate boundary faults like the San Andreas fault are weak (Zoback et al. 1987). We will return to the issue of the strength of the creeping section of the San Andreas fault below.

If we speculate that faults can be strong most of the time, but weak while slipping in earthquakes, what might be the values of stress on these faults at different times? It is possible that the stress is well below the Coulomb failure value most of the time and that it only reaches that value at unusual places or times. Only in earthquake nucleation zones would the static stress be high enough to overcome the Byerlee's Law static friction coefficient times the local effective normal stress. Elsewhere along a fault experiencing an unusual earthquake event, the stress would only very momentarily overcome the static strength, owing to the dynamic stress concentration at the tip of the propagating rupture (Figure 7.1). This means that the time average of the tectonic stress on the fault could be quite low and that the stress during most of the dynamic slip is even lower, possibly approaching zero. In fact, if the dynamic resistance were quite low as we presently expect (but need to verify), it would be difficult to argue that the ambient or tectonic stress on the fault could be too high. This is because, if it were, the dynamic stress drop (the difference between the initial stress and the dynamic friction) would be too large to fit seismological observations (see Chapter 6). The same would be true of the static stress drop, the difference between the initial and the final stress after slip ceased. Thus, by using seismological observations of stress drops and tying these delta values to a low absolute level equal to low dynamic friction, we obtain a model in which the stress on faults in most places is much lower than predicted by Byerlee's Law. We can have a strong crust that is able to deform seismically at low levels of tectonic stress! How can we test this model?

One way would be to focus on understanding what could cause stress on faults to be so heterogeneous such that it could statically overcome friction in the nucleation zone (Figure 7.1) and simultaneously be much lower elsewhere. This heterogeneity could be in the shear stress on the fault, perhaps due to stress concentrations left from previous slip events, or it could be in the effective normal stress. The nucleation regions could have elevated pore pressures or, perhaps as a result of geometric irregularities, could have reduced normal stress. Evidence that pore-fluid pressure can be involved in the triggering of slip in the nucleation zone and that stress levels must be essentially at the static failure or static friction stress in the nucleation zone comes from fluid injection experiments in deep boreholes, for example, the German deep drilling project, KTB. There, it was shown that pore-pressure perturbations of only 1-100 KPa can induce microseismic events with a magnitude of 2 and sometimes even greater (Shapiro et al. 1997). These data do not tell us how the stress varied over the area that slipped in such microearthquakes, but they do indicate that the preexisting stress at the hypocenter of those small events was high.

What other tests of this model can be devised? Improving our knowledge of dynamic weakening mechanisms is critical, as discussed above. In addition, we need to develop better methods with which to estimate dynamic and static stress drops (Figure 7.1). Measurements of the stress in the SAFOD and Nankai drilling projects will help, although the limited number of stress measurements may be insufficient to capture the heterogeneity of stress state along the faults. The results will have to be interpreted with caution.

Dynamic weakening processes do not operate on creeping faults; thus the stress level along creeping segments of faults (e.g., on San Andreas and North Anatolian faults) are worthy of careful attention. Brune (2002) suggested that the heat flow may be elevated only on the eastern side of the creeping segment of the San Andreas fault; this asymmetrical pattern is expected if rocks there underwent creep at high stresses, whereas rocks on the western side previously slipped dynamically at low stresses when they occupied a position further to the south. As the heat flow data along this creeping section are rare, more data are necessary to test this hypothesis, especially in light of the considerable scatter and attendant averaging of the heat flow values from such measurements.

SUMMARY

Much has been learned about the rheology and mechanics of faults and earthquakes in the past several years, and it is an exciting as well as societally important area of research. We have identified numerous areas where additional research efforts could be focused to fill in some of the critical gaps in our knowledge. This would bring us closer to the goals of creating computer simulations of dynamic rupture of earthquakes using what appear to be realistically low values of dynamic shear resistance. Through such models, we may then be able to predict strong ground motions from first principles. This can help us know where to build structures and how to design them to withstand damage. Studies of the earthquake nucleation process may even bring us closer to the elusive goal of reliable earthquake prediction.

REFERENCES

- Brune, J. 2002. Heat flow on the creeping section of the San Andreas fault: A localized transient perspective. In: 2003 SCEC Annual Meeting Proc. and Abstr., pp. 58–59. Palm Springs, CA: Southern California Earthquake Center.
- d'Alessio, M.A., A.E. Blythe, and R. Bürgmann. 2003. No frictional heat along the San Gabriel fault, California: Evidence from fission track thermochronology. *Geology* 31:541–544.
- Dieterich, J.H. 1979. Modeling of rock friction. 1. Experimental results and constitutive equations. J. Geophys. Res. 84:2161–2168.
- Dieterich, J.H. 1986. A model for the nucleation of earthquake slip. In: Earthquake Source Mechanics, Geophys. Monogr. Ser. 37, pp. 37–47. Washington, D.C.: Am. Geophys. Union.
- Dieterich, J.H. 1992. Earthquake nucleation on faults with rate- and state-dependent strength, *Tectonophysics* **211**:115–134.
- Dieterich, J.H., and B. Kilgore. 1996. Implications of fault constitutive properties for earthquake prediction. *PNAS* **93**:3787–3794.
- Di Toro, G., D.L. Goldsby, and T.E. Tullis. 2004. Friction falls toward zero in quartz rock as slip velocity approaches seismic rates. *Nature* **427**:436–439.
- Dodge, D.A., G.C. Beroza, and W.L. Ellsworth. 1996. Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *J. Geophys. Res.* 101:22,371–22,392.
- Goldsby, D., and T.E. Tullis. 1998. Experimental observations of frictional weakening during large and rapid slip. *EOS Trans. Am. Geophys. Union* **97**:F610.
- Goldsby, D., and T.E. Tullis. 2002. Low frictional strength of quartz rocks at subseismic slip rates. *Geophys. Res. Lett.* **29**:1844, doi:10.1029/2002GL015240.
- Goldsby, D.L., and T. Tullis. 2003. Flash heating/melting phenomena for crustal rocks at (nearly) seismic slip rates. In: 2003 SCEC Annual Meeting Proc. and Abstr., pp. 98–90. Palm Springs, CA: Southern California Earthquake Center.
- Hirose, T., and T. Shimamoto. 2005. Growth of a molten zone as a mechanism of slip weakening of simulated faults in gabbro during frictional melting. *J. Geophys. Res* 110:B05202, doi:10.1029/2004JB003207.
- Johnston, M.J.S., and A.T. Linde. 2002. Implications of crustal strain during conventional, slow, and silent earthquakes. In: International Handbook of Earthquake and Engineering Seismology, ed. W. Lee et al., pp. 589–605. San Diego: Academic.
- Junger, J.A., and T.E. Tullis. 2003. Fault roughness and matedness suggest significant fault-interface dilatancy with slip. *EOS Trans. Am. Geophys. Union* **84 Suppl.**: Abstr. S51B-03.
- Kato, N., and T.E. Tullis. 2001. A composite rate- and state-dependent law for rock friction. *Geophys. Res. Lett.* 28:1103–1106.

- King, G.C.P. 1983. The accommodation of strain in the upper lithosphere of the Earth by self-similar fault systems: The geometrical origin of b-value. *Pure & Appl. Geophys.* 121:761–815.
- King, G.C.P. 1986. Speculations on the geometry of the initiation and termination of earthquake rupture and the evolution of morphology and geological structures. *Pure* & *Appl. Geophys.* **124**:567–585.
- King, G.C.P., and J. Nabelek. 1985. The role of bends in faults in the initiation and termination of earthquake rupture: Implications for earthquake prediction. *Science* 228:986–987.
- King, G.C.P., and G. Yielding. 1984. The evolution of a thrust fault system: Processes of rupture initiation, propagation and termination in the 1980 El Asnam (Algeria) earthquake. *Geophys. J. Roy. Astron. Soc.* 77:915–933.
- Liu, Y., and J.R. Rice. 2005. Aseismic slip transients emerge spontaneously in 3D rate and state modeling of subduction earthquake sequences. J. Geophys. Res. 110:B08307, doi:10.1029/2004JB003424.
- Manighetti, I., G.C.P. King, and C. Sammis. 2004. The role of off-fault damage in the evolution of normal faults and the back-arc extension in the Aegean. *Earth Planet. Sci. Lett.* 217:339–408.
- Marone, C.J. 1998. Laboratory-derived friction constitutive laws and their application to seismic faulting. Ann. Rev. Earth Planet. Sci. 26:643–696.
- McGuire, J., and Y. Ben-Zion. 2005. High-resolution imaging of the Bear Valley section of the San Andreas fault at seismogenic depths with fault zone head waves and relocated seismicity. *Geophys. J. Intl.* **163**:152–164.
- Oleskevich, D.A., R.D. Hyndman, and K. Wang. 1999. The updip and downdip limits to great subduction earthquakes: Thermal and structural models of Cascadia, South Alaska, SW Japan, and Chile. J. Geophys. Res. 104:14,965–14,991.
- Rogers, G., and H. Dragert. 2003. Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* 300:1942–1943; ScienceExpress, 10.1126.
- Roig Silva, C., D.L. Goldsby, G. Di Toro, and T.E. Tullis. 2004. The role of silica content in dynamic fault weakening due to gel lubrication. *EOS Trans. Am. Geophys. Union* Suppl. 85:T21D-07.
- Ruina, A.L. 1983. Slip instability and state variable friction laws. J. Geophys. Res. 88:10,359–10,370.
- Scholz, C.H. 2002. The Mechanics of Earthquakes and Faulting. Cambridge: Cambridge Univ. Press.
- Segall, P., and J.R. Rice. 2004. Dilatancy, compaction, and slip instability of a fluidinfiltrated fault. EOS Trans. Am. Geophys. Union 75 Suppl.:425.
- Shapiro, S.A., E. Huenges, and G. Borm. 1997. Estimating the crust permeability from the fluid-injection induced seismic emission at the KTB site. *Geophys. J. Intl.* 131:F15–F18.
- Tsutsumi, A., and T. Shimamoto. 1997. High-velocity frictional properties of gabbro. Geophys. Res. Lett. 24:699–702.
- Tullis, T.E. 1996. Rock friction and its implications for earthquake prediction examined via models of Parkfield earthquakes. In: Earthquake Prediction: The Scientific Challenge, ed. L. Knopoff. *PNAS* 93:3803–3810.
- Wibberley, C.A.J., and T. Shimamoto. 2003. Internal structure and permeability of major strike-slip fault zones: The Median Tectonic Line in Mid Prefecture, Southwest Japan. J. Struct. Geol. 25:59–78.
- Zoback, M.D., M.L. Zoback, J. Mount et al. 1987. New evidence on the state of stress of the San Andreas fault system. *Science* 238:1105–1111.